ANGELO ANTONIO MANZATTO

SEGMENTAÇÃO ÓSSEA DO CORPO HUMANO EM TOMOGRAFIAS COMPUTADORIZADAS USANDO REDES *DEEP LEARNING*

Curitiba - PR

Brasil 2021

ANGELO ANTONIO MANZATTO

SEGMENTAÇÃO ÓSSEA DO CORPO HUMANO EM TOMOGRAFIAS COMPUTADORIZADAS USANDO REDES *DEEP LEARNING*

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de mestre em Informática.

Orientador: PROF. DR. EDSON EMÍLIO SCALABRIN

Curitiba - PR

Brasil 2021

Agradecimentos

Inicio meus agradecimentos à Pontifícia Universidade Católica do Paraná e seu excepcional corpo docente por tornar possível este projeto em parceria com a CAPES pelo suporte financeiro.

Ao Prof. Dr. Alceu de Souza Britto Jr. por apresentar o programa de pós-graduação da PUC-PR mostrando o vasto mundo de possibilidades no ramo da inteligência artificial.

Aos meus grandes orientadores, Prof. Dr. Edson J. R. Justino e Prof. Dr. Edson Scalabrin, que me guiaram por essa difícil trilha do mestrado dando tanto o apoio e suporte em minhas decisões quanto a oportunidade de correções nos eventuais desvios que ocorreram durante o percurso.

Ao Prof. Dr. Manoel Moreira, que sempre me incentivou a refletir sobre as "ür-questões", ou seja, aquelas questões essenciais na busca por conhecimento.

Ao Prof. Dr. Oge Marques pelos conselhos sobre o mestrado: ser um caminho de paciência e perseverança.

Aos doutores e colegas de mestrado e doutorado, Flávio de Almeida e Silva, Diogo Olsen e William John Pereira Brobouski, que ajudaram tanto a dar foco e direção ao trabalho como também sempre estiveram disponíveis durante horas para boas conversas e trocas no laboratório.

Ao meu amigo de décadas, Antônio Carlos Foltran, sempre trazendo sabedoria e palavras de coragem fazendo jus à palavra "amizade".

A minha amada esposa Caren, que sempre esteve ao meu lado nos momentos difíceis dessa trajetória, dando-me apoio, suporte e sendo um dos grandes pilares da família bem como companheira inseparável.

Aos meus queridos pais, Antônio e Márcia, que acreditaram no meu potencial para vencer mais esse desafio da vida.

Ao meu irmão do coração, Guilherme, que além de irmão sempre foi um dos melhores amigos que tive com quem sempre pude contar em qualquer situação.

E, finalmente, a você meu filho, Davi, que é uma das grandes razões da vida do papai, que veio ao mundo lutando como um leão e, agora, pintando e bordando, enche de cores e sabores essa minha jornada pela vida.

Resumo

A segmentação dos órgãos do corpo humano em imagiologia médica é um processo muito utilizado tanto na medicina para detecção e diagnóstico de doenças como na educação auxiliando estudantes no aprendizado da anatomia humana. Apesar de sua significativa importância, esse processo é demorado e custoso uma vez que requer especialistas na área, tempo e ferramentas adequadas para fazê-lo. Seguindo os avanços da inteligência artificial, este trabalho tem como objetivo comparar a segmentação dos ossos do corpo humano entre diferentes arquiteturas de redes *deep learning* em imagens de Tomografia Computadorizada (TC) de corte axial nas bases de corpo completo disponibilizadas pelo *Visible Human Project* (VHP). Cada arquitetura de rede *deep learning* faz a segmentação classificando cada *pixel* da tomografia entre 19 classes (18 tipos de ossos + fundo/*background*) sendo estas: crânio, mandíbula, clavícula, escápula, úmero, rádio, ulna, mãos, costelas, esterno, vértebras, sacro, bacia, fêmur, patela, tíbia, fíbula e pés. A metodologia proposta consiste em utilizar a base tomográfica do corpo masculino do *VHP* contendo 1.865 imagens de tamanho 512 x 512 *pixels* para treinar redes *deep learning* de diversas arquiteturas em modo supervisionado para testar suas generalizações na segmentação óssea da base do corpo feminino, contendo 1.732 imagens. A avaliação quantitativa do desempenho entre as diversas topologias de rede, ao final do processo, foi feita usando o coeficiente *Dice* como métrica e análises estatísticas. Nesta dissertação, é demonstrada a superioridade da rede U-Net, frente às demais arquiteturas concorrentes, com um Dice global de 0.6854 comprovando que existe diferença significativa nos resultados, dependendo da topologia de rede *deep learning* selecionada.

Palavras-chave: *Deep learning*. Segmentação médica. Tomografia computadorizada. Ossos.

Abstract

Organ segmentation of the human body in medical imaging is a widely used process not only in medicine for detection and diagnosis of diseases, as well as in education helping students in learning the human anatomy. Despite its importance, this process is very time-consuming and costly as it requires experts in the field, time and adequate tools to do so. With the advances of artificial intelligence, the goal of this work is to compare the bone segmentation among different deep learning networks in computed tomography (CT) images on the full body databases provided by the Visible Human Project (VHP). The segmentation is done by each model classifying each pixel from the tomography image into one of the following 19 classes (18 bones + background) being: cranium, jaw, clavicle, scapula, humerus, radius, ulna, hands, ribs, sternum, vertebras, sacrum, pelvis, femur, patella, tibia, fibula and feet. The proposed method consists of using the male body tomographic base containing about 1,865 images of size 512 x 512 to train various deep learning network architectures in supervised mode including data augmentation techniques and external databases aiming to test its generalization for the bone segmentation on the female body base containing close to 1,730 images. The quantitative evaluation between the various network topologies at the end of the process will be made using the Dice coefficient as metric and statistical analysis. This project demonstrates the superiority of the U-Net network when compared to other competing architectures reaching a global Dice score of 0.6854 proving that there is a significant difference in results depending on the selected deep learning topology.

Keywords: Deep learning. Medical segmentation. Computed tomography. Bones.

Lista de ilustrações

Lista de Tabelas

Lista de quadros

Lista de abreviaturas e siglas

ANOVA *Analisys of Variance*

ANN *Artificial Neural Network*

ASPP *Atrous Spatial Pyramid Pooling*

c.f. conforme

cGAN *Conditional Generative Adversarial Network*

CNN *Convolutional Neural Network*

CRF *Conditional Random Filled*

CRN *Convolutional Residual Networks*

DAC Diagnóstico Assistido por Computador

DICOM *Digital Imaging and Communications in Medicine*

2D *2 Dimensional Pixel*

3D *3 Dimensional Pixel*

FCN *Fully Convolutional Network*

GAN *Generative Adversarial Network*

GPU *Graphics Processing Unit*

HU *Hounsfield Units*

i.e. ou seja, isto é

IoU *Intersection over Union*

IRCAD Instituto de Pesquisa contra o Câncer do Aparelho Digestivo (na França)

NDN *Nested Dilation Networks*

PET *Positron Emission Tomography*

ReLU *Rectified Linear Unit*

RM Ressonância Magnética

RMS *Root Mean Square*

RNN *Recurrent Neural Networks*

TC Tomografia Computadorizada

VHP *Visible Human Project*

Sumário

1 Introdução

Segmentação em imagiologia médica é uma área crucial tanto dentro da medicina, auxiliando na identificação e tratamento de doenças, como na educação, ensinando alunos sobre a anatomia do corpo humano. As ferramentas de Diagnóstico Assistido por Computador (DAC) fazem uso cada vez mais frequente dessa técnica para entregar informações críticas aos profissionais da área médica como na localização de tumores e lesões em órgãos do corpo, auxiliando no combate ao câncer, por exemplo.

Imagiologia médica são imagens geradas por dispositivos que utilizam radiação eletromagnética por meio das máquinas de ressonância e tomografia para visualizar as estruturas internas do corpo humano. As imagens geradas pelo scanner tomográfico, denominadas Tomografias Computadorizadas (TC), consistem de matrizes de *pixels*, em que cada *pixel* representa um pequeno elemento de volume, ou *voxel*, de um "corte" ou "fatia" da imagem da parte examinada (Fleckenstein, Tranum-Jensen, 2004, p. 22) [1] conforme pode ser observado na Figura 1. Os tamanhos das matrizes variam geralmente de 128 x 128 *pixels* a 1024 x 1024 *pixels*, sendo o tamanho de 512 x 512 o mais utilizado.

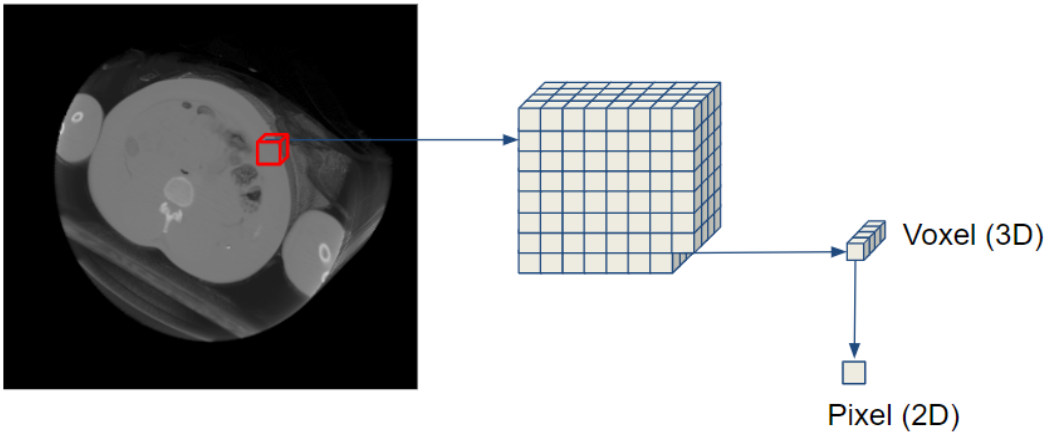


Figura 1 – Definição de um voxel como um pixel em três dimensões.

Fonte: Compilação do autor

A segmentação de uma imagem encerra um processo que consiste em particioná-la em regiões semanticamente interpretáveis (H. Barrow, J. Tennenbaum, 1978) [2] Em imagens computadorizadas, isso consiste em classificar cada *pixel* como pertencente a uma certa classe (Figura 2). Na imagiologia médica, essa técnica é utilizada para definir as fronteiras entre os objetos de estudo, como órgãos e lesões das partes não interessadas ou fundo, podendo ser aplicada tanto em imagens 2D como em imagens 3D. A título de

ilustração, pode-se dizer que se quer segmentar a imagem do pulmão de uma TC a fim de mostrar quais *pixels* da imagem pertencem ao pulmão e quais não pertencem ao pulmão.

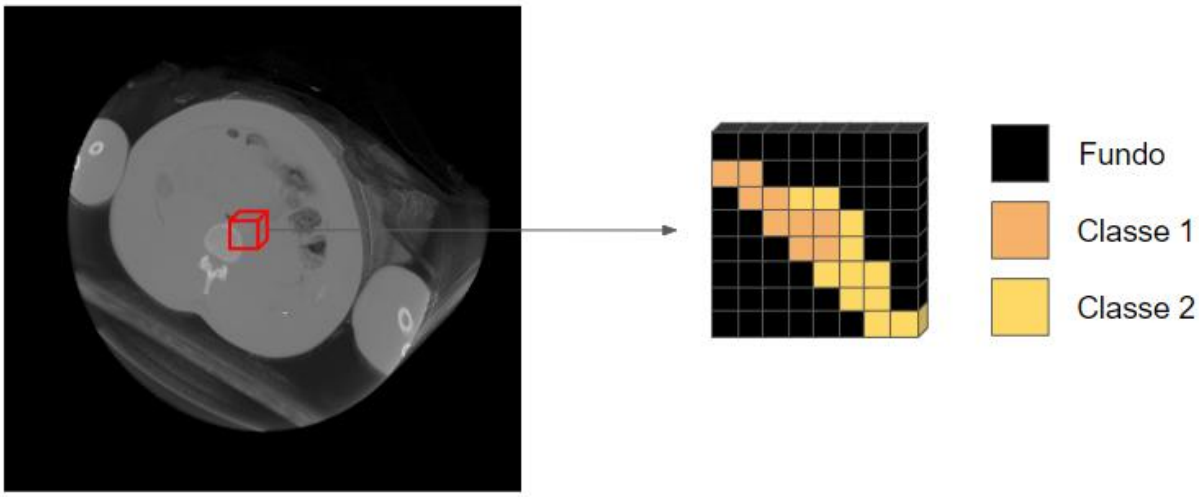


Figura 2 - Classificação de um pixel em uma determinada classe.

Fonte: Compilação do autor.

Em sistemas tecnológicos tradicionais, o processo de segmentação automatizado ou semiautomatizado é feito de forma rudimentar, aplicando-se filtros de detecção de borda e métodos matemáticos, como por exemplo, morfologia matemática. Com os avanços das aplicações de inteligência artificial, em particular, envolvendo aprendizagem de máquina, algoritmos mais complexos foram desenvolvidos para compor extratores de características mais elaboradas colocando em prática uma segmentação de imagem mais precisa com baixo esforço/intervenção humana.

Em meados de 2000, com a chegada de computadores de alto desempenho, contendo placas também de alto desempenho de aceleração gráfica, permitiu-se consolidar a implementação de redes neurais de aprendizado profundo, ou redes *deep learning*. Tais redes são capazes de aprender a identificar o melhor conjunto de filtros para segmentar imagens médicas sem a necessidade de intervenção humana explicitamente no processo. As redes *deep learning* fazem uso do aprendizado estruturado em hierarquias, similar ao funcionamento do neocórtex cerebral, no qual cada camada é responsável pelo aprendizado de uma certa característica inerente à tarefa executada.

No processo de segmentação de imagens, enquanto as camadas mais externas da rede *deep learning* aprendem detalhes como contorno e cores dos órgãos, as camadas mais internas aprendem minúsculas formas e outros detalhes mais sutis. Na medicina, muitas pesquisas envolvendo redes *deep learning* na segmentação de órgãos têm sido feitas citando desde segmentação dos pulmões, buscando por nódulos pulmonares [4, 5, 6, 7, 8, 9], segmentação do pâncreas para encontrar lesões [10, 11], localização de pólipos do cólon e do reto para tratamento de câncer colorretal [12], dentre muitas outras existentes.

Em estudos mais recentes como, por exemplo, o de Hesamian et al., (2019) [3], mostra-se que o desempenho alcançado pelas redes *deep learning* na tarefa de segmentação de imagens é superior a quaisquer outros métodos já utilizados o que tem levado a um aumento expressivo nos esforços de pesquisa com vista ao desenvolvimento de novas e mais promissoras arquiteturas nessa área.

Apesar da evolução das redes *deep learning*, há poucos estudos que testaram exaustivamente os limites da possibilidade de uma única rede segmentar todos os órgãos do corpo, na direção de produzir um atlas médico. Em [13] foi proposto um modelo/arquitetura de rede *deep learning* 3D treinada de forma semissupervisionada para segmentar 16 órgãos abdominais e, em [14] foi proposta a utilização da rede *deep learning* do tipo U-Net para a segmentação da imagem de seis tipos de ossos diferentes. Tanto a falta de bases de imagens tomográficas de corpo completo, como a falta de especialistas nas áreas disponíveis para criar as máscaras de *ground truth* de imagens segmentadas para cada órgão tem sido um dos grandes impeditivos dos avanços nessa questão.

Unindo o desafio da multissegmentação com o avanço das aplicações de inteligência artificial, em particular, aprendizagem de máquina, uma lacuna importante que este trabalho de pesquisa busca preencher encerra a produção de um corpo de conhecimentos *vis-à-vis* aos diferentes modelos de redes *deep learning* na tarefa de segmentar imagens médicas do tipo tomografia computadorizada, visando investigar se há diferença significativa nos resultados obtidos de acordo com a topologia de rede *deep learning* utilizada.

Para buscar a resposta ao problema,serão selecionadas várias topologias de redes *deep learning*, treinadas e testadas na segmentação de imagens entre 18 classes de ossos distintos. Para tal, será usada a base de imagens anatômicas *Visible Human Project* (VHP), que é uma das bases de imagens mais completas disponíveis. A avaliação do desempenho individual de cada rede *deep learning*, após a segmentação, será feita por meio do cálculo de métricas sobre os resultados obtidos. As comparações entre elas serão feitas usando métodos estatísticos.

Os principais desafios à solução do problema são: a rotulação do *ground truth* na base de treinamento que possui muitas classes de ossos; o equipamento adequado para treinamento das redes; a escolha de topologias de redes *deep learning* que sejam representativas na tarefa de segmentação de órgãos; metodologia não enviesada de treinamento e testes até a escolha de análises estatísticas adequadas que permitam chegar a uma conclusão assertiva na comparação do desempenho individual e em grupo das redes.

1.1 Objetivos

Este trabalho tem como objetivo realizar a segmentação óssea do corpo humano em tomografias computadorizadas utilizando redes *deep learning* (Figura 3). Para a consecução desse objetivo foram definidos os seguintes objetivos específicos:

- Criar a segmentação de *ground truth* nas bases de imagens tomográficas selecionadas com vistas ao treinamento supervisionado das redes *deep learning*;
- Selecionar, com base na literatura, topologias de redes *deep learning* para a segmentação de ossos do corpo humano;
- Realizar experimentos com as redes *deep learning* selecionadas, contemplando treinamento e testes de segmentação de imagens;
- Avaliar os resultados em termos de segmentação de imagens das tomografias da base de testes e verificar se alguma topologia de rede *deep learning* apresenta desempenho significativamente superior.

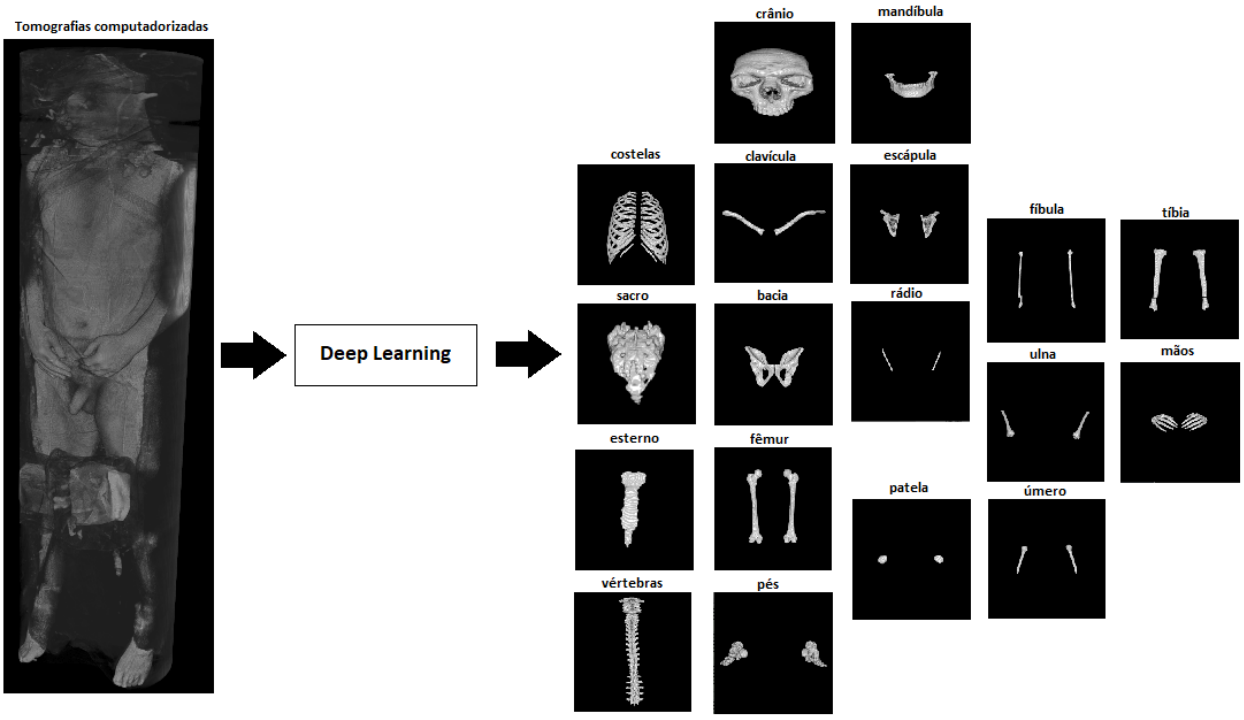


Figura 3 – Segmentação de uma base tomográfica em 18 classes de ossos.

Fonte: Compilação do autor.

1.2 Hipóteses

A forma de mensurar os resultados será por meio do cálculo dos coeficientes Dice entre cada tomografia segmentada e seu *ground truth* correspondente, levando-nos a duas hipóteses de pesquisa:

H_0: Todas as redes *deep learning* possuem coeficientes Dice médios iguais entre si;

H_1: Uma rede *deep learning* pelo menos possui um coeficiente Dice médio diferente.

Caso a primeira hipótese seja verdadeira, pode-se afirmar que o resultado da segmentação é independente da topologia de rede *deep learning* selecionada. Por outro lado, caso seja falsa, então, serão conduzidos testes estatísticos entre pares para encontrar as diferenças entre elas..

1.3 Contribuições

Este projeto de dissertação apresenta as seguintes contribuições científicas:

- Um *framework* computacional para comparação de diferentes topologias de redes *deep learning* objetivando encontrar qual rede possui desempenho superior na tarefa de segmentação óssea em tomografias de corte axial sob condições

desafiadoras para treinamento com poucos exemplos e muitas classes. Até o presente momento, o máximo de classes de ossos segmentados eram seis em La Rosa [14], sendo que a metodologia proposta aqui são 18.

- Uma base de imagens tomográficas devidamente rotuladas com *ground truth* de 18 classes de ossos, auxiliando em pesquisas de algoritmos de segmentação na área médica.
- Dados detalhados da segmentação contendo quais topologias foram melhores para cada classe, erros mais frequentes (*matriz de confusão*), gráficos de desempenho por amostra e por arquitetura de rede (ver **Apêndice A**).

1.4 Organização

A organização deste documento de dissertação está estruturada em seis Capítulos. O primeiro Capítulo apresenta uma visão geral sobre os objetivos, desafios e contribuições deste projeto. O segundo traz a fundamentação teórica das redes *deep learning* e principais arquiteturas encontradas na literatura. O terceiro Capítulo apresenta trabalhos relacionados com o uso de redes *deep learning* na segmentação de imagens médicas, apresentando as topologias e técnicas estudadas pelos autores e os resultados alcançados. No quarto Capítulo, é abordada a metodologia de condução do trabalho, trazendo detalhes que vão desde as ferramentas utilizadas, passando pela confecção das máscaras ósseas, terminando com a definição dos procedimentos de testes e avaliação das redes *deep learning*. No quinto Capítulo, são apresentados os resultados dos experimentos e as análises estatísticas, colocando em evidência a solução para o problema de pesquisa e a verificação, ou seja, se existe diferença de desempenho entre as redes *deep learning* face aos resultados da segmentação óssea, em termos da topologia utilizada. Por fim, o sexto e último Capítulo apresenta as conclusões deste trabalho e sugestões para trabalhos futuros.

2 Fundamentação Teórica

Este capítulo traz o embasamento teórico desta dissertação estando dividido em duas partes: a primeira parte introduz os conceitos de redes *deep learning* e a segunda apresenta as principais topologias de tais redes utilizadas com vista na segmentação de imagens médicas.

2.1 *Deep Learning*

As redes *deep learning* são o mais novo pilar de aplicação da inteligência artificial. Elas dotam os sistemas de capacidades de aprendizagem automática a partir de dados e tais capacidades vão transformar as vidas cotidianas das pessoas, seja ajudando a decidir o próximo compromisso, seja ajudando a escolher a roupa que melhor se adapta ao corpo, seja conduzindo um automóvel de forma autônoma para levar uma pessoa ao seu destino; nesse contexto, a pessoa poderá, enquanto é conduzida ao seu destino, assistir a um filme também recomendado por um desses algoritmos. Segundo a consultoria de negócios Fortune Business Insights [19], estima-se que o investimento no mercado de inteligência artificial saltará de 20,8 bilhões de dólares, estimados em 2018, para 202,6 bilhões de dólares em 2026, o que torna esse assunto altamente relevante no atual contexto tecnológico.

Para se entender melhor o conceito de redes *deep learning*, precisa-se, primeiramente, entender dois conceitos mais abrangentes sob os quais esse termo está situado, que são *inteligência artificial* e *machine learning*, em que o segundo está incluso no primeiro (Figura 4).

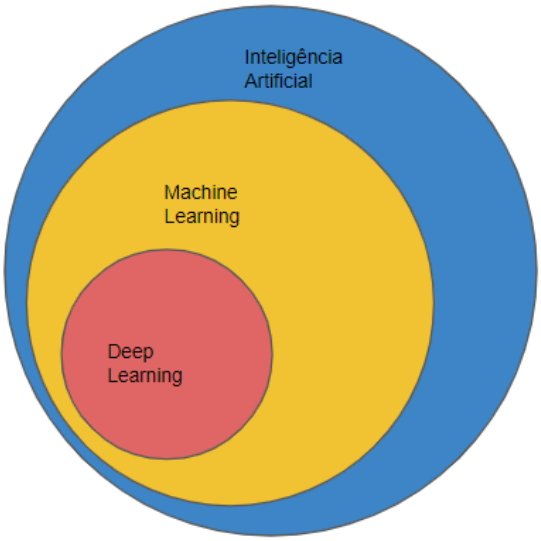


Figura 4 - As redes *deep learning* são essencialmente especializações de *machine learning*.

Fonte: Compilação do autor

Inteligência artificial é uma área da Ciência da Computação que tem por objetivo criar máquinas capazes de realizarem tarefas normalmente reservadas ao ser humano. O cunho do termo nasceu em meados da década de 1950 quando Alan Turing propôs questionar se as máquinas poderiam pensar e, desde então, o número de pesquisas nessa área evoluiu de sistemas baseados em regras do tipo *if-then* para sistemas baseados em decisões estatísticas e modelos matemáticos complexos.

Em 1959, Arthur Samuel cunhou o termo *machine learning*, ou aprendizado de máquina, ao qual se referiu como a capacidade de uma máquina de aprender sem ser explicitamente programada para tal. Sendo mais específico, *machine learning* envolve a construção de algoritmos que adaptam seus modelos de acordo com os dados a eles expostos de forma a fazer previsões cada vez mais corretas.

Um sistema de *streaming* musical, por exemplo, que deseja personalizar a experiência de um ouvinte, pode fazê-lo com o uso de técnicas de *machine learning*, em que tal sistema aprende a recomendar novas músicas, associando o gosto deste usuário com a de outros cujas preferências são similares. Filtros de *spam*, detecção de fraudes em seguradoras, identificação de padrões em imagens, análise do mercado financeiro e reconhecimento de voz são outros exemplos de aplicação de *machine learning*.

No final das contas, *machine learning*, é uma junção da inteligência artificial com modelos matemáticos complexos que aprendem, a partir de dados, a minimizar o erro ou maximizar a probabilidade em relação ao objetivo definido. Por exemplo, se por um lado,

para um classificador de imagens de animais buscamos a máxima probabilidade de classificar corretamente entre gatos e cachorros, por outro lado, para um sistema de logística buscamos manter o mínimo de estoque necessário no armazém.

São três as principais modalidades possíveis de treinamento dos algoritmos de *machine learning* para a realização das tarefas desejadas: *aprendizado supervisionado, aprendizado não supervisionado* e *aprendizado por reforço*.

No *aprendizado supervisionado*, dados de entrada (X) e saídas esperadas (Y) são fornecidos ao algoritmo de forma que ele aprenda a gerar uma função de mapeamento entre a entrada e a saída do tipo Y = f(X). Em contraste, no *aprendizado não supervisionado* são fornecidos apenas dados de entrada (X) e nenhuma resposta de saída. Aqui, o objetivo é obter um modelo que consiste em encontrar padrões ou extrair informações das intrínsecas às estruturas internas. Detecção de anomalias em transações bancárias ou segregação de pessoas em grupos de consumo com vistas a comandar campanhas de propagandas dirigidas são exemplos de aplicações que usam esse tipo de treinamento. Já no *treinamento por reforço,* o algoritmo aprende pela sua interação com um cenário dinâmico por meio de ações que recebem como resposta punições ou recompensas, de forma a modelar um comportamento ideal. Carros autônomos assim como agentes em jogos ou até mesmo uma máquina autônoma de aspirar pó são aplicações ideais nessa área.

Existem vários tipos de modelos de *machine learning* pesquisados e desenvolvidos com vistas a solucionar os mais variados problemas, como exemplo: redes neurais artificiais, máquina de suporte de vetores, árvore de decisão, algoritmo genético, *clustering*, redes bayesianas, entre outros. Todos esses algoritmos podem ser classificados como planos, pois possuem apenas um nível, ou camada, de abstração dos dados.

Dentre os algoritmos citados acima, as redes neurais artificiais eram construídas, inicialmente, com apenas uma camada de abstração, ou camada oculta, pois se acreditava que mais camadas poderiam torná-las, além de ineficientes, impossíveis de serem treinadas. No entanto, o avanço das técnicas de otimização estudadas por [20] permitiu a construção de redes neurais contendo múltiplas camadas de processamento (Figura 5), em que cada camada é capaz de aprender um nível de abstração diferente dos dados, similar ao funcionamento do cérebro humano [18]. A essa nova forma de aprendizado, contendo múltiplos níveis, deu-se o nome de aprendizado profundo ou *deep learning*.

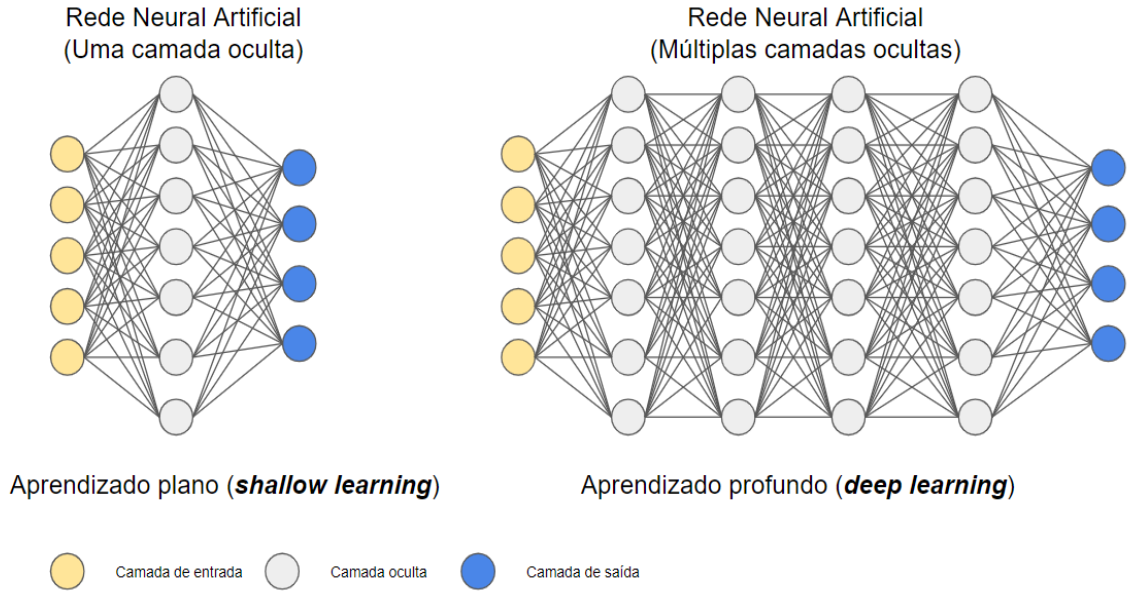


Figura 5 – As redes *deep learning* são redes neurais com múltiplas camadas ocultas.

Fonte: Compilação do autor.

Redes *deep learning* são um avanço das redes neurais artificiais que, por sua vez, é um tipo de algoritmo de *machine learning* inserido em um contexto mais abrangente da inteligência artificial. Sua popularização recente deu-se por dois fatores cruciais: o primeiro foi o aumento da capacidade dos computadores, e o segundo, a quantidade de dados úteis disponíveis para treinamento.

Interessa-nos destacar um tipo de rede neural conhecida como rede neural convolutiva formada por múltiplas camadas, contendo filtros capazes de aprender características relacionadas a imagens que têm sua aplicação em tarefas, que vão desde classificação até localização e segmentação de objetos. A segmentação em imagens médicas é feita por meio do uso desse tipo de rede para extrair os órgãos do corpo humano, sendo que discutiremos essa questão com um pouco mais de profundidade na próxima seção.

2.2 Redes Neurais Artificiais

Redes neurais artificiais são um tipo de algoritmo de *machine learning* inspirado no funcionamento do cérebro humano. O Dr. Robert Hecht-Nielsen, um dos inventores do primeiro neurocomputador, definiu as redes neurais artificiais, em tradução livre, como “um

sistema de computação feito por um número de simples, altamente conectados elementos de processamento, que processam informações pelas respostas de seus estados dinâmicos as entradas externas".

As redes neurais são tipicamente organizadas em camadas sequenciais que consistem de uma camada de entrada, várias camadas ocultas e uma camada de saída (Figura 6). Cada camada é composta de múltiplos nós que se conectam à camada seguinte por meio de uma rede de "pesos" responsáveis pelo processamento da rede.

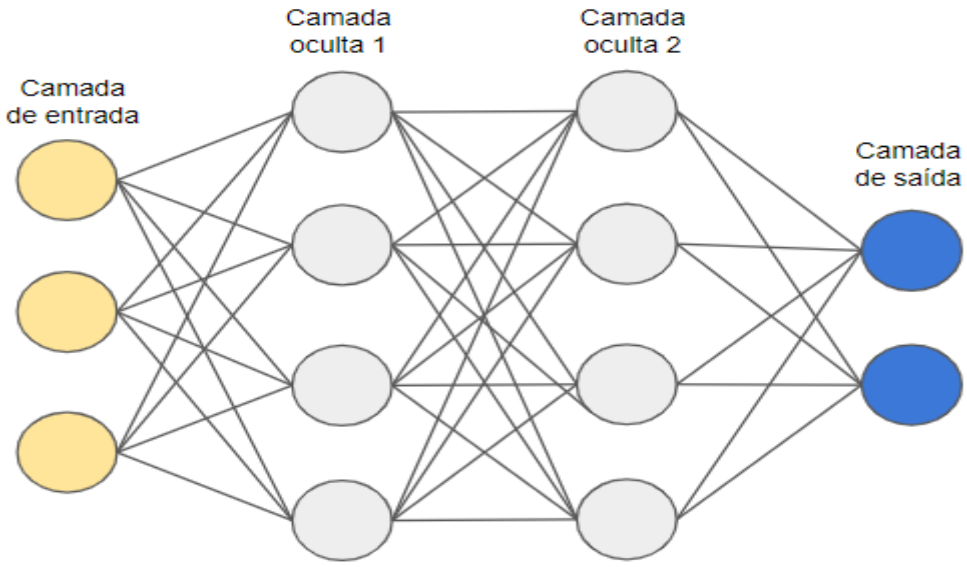


Figura 6 - Rede neural multicamada.

Fonte: Compilação do autor

Os nós possuem funcionamento semelhante a um neurônio biológico e, por isso, o nome "redes neurais", sendo a unidade básica de processamento de cada camada. Todo nó recebe, na sua entrada, sinais provenientes de outros nós ou de uma fonte externa, faz o processamento e envia um sinal de saída. O processamento é realizado fazendo-se a soma da média ponderada entre cada entrada (x) associada a seu respectivo peso (w) e aplicando-se ao valor calculado uma função não linear chamada "função de ativação" (Figura 7).

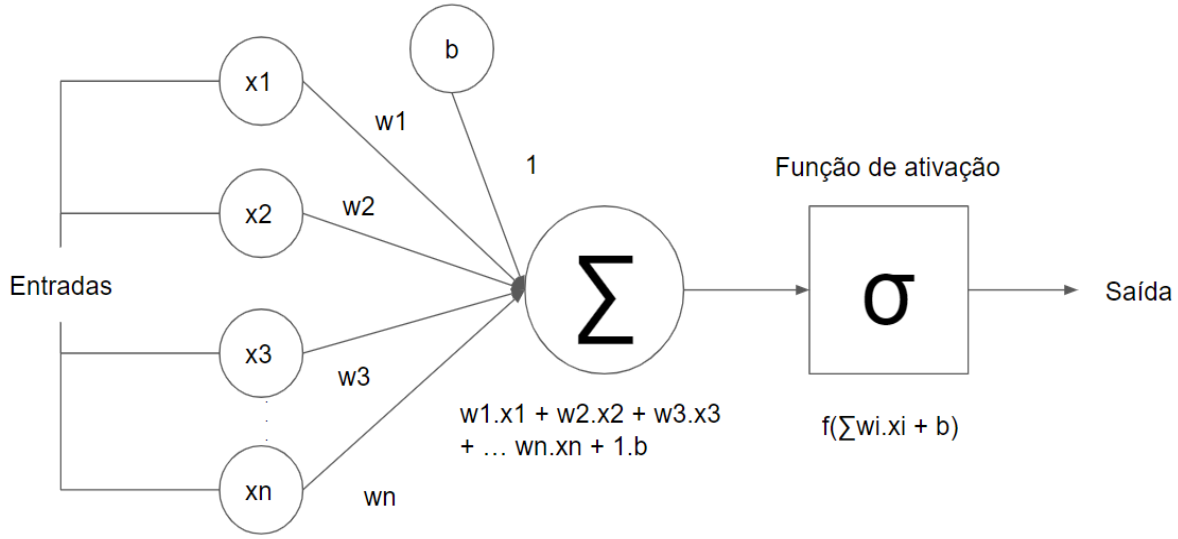


Figura 7 - Representação da operação realizada em um nó da rede.

Fonte: Compilação do autor

Matematicamente, pode-se escrever que a saída de cada nó é definida pela Equação 1, na qual *n* representa o número de entradas que estão conectadas ao neurônio. O termo *b* vem de *bias* e é usado para fazer o deslocamento da função de ativação similar ao coeficiente linear de uma equação de primeiro grau.

$$Z = f\left(\sum_{i=1}^{n} wi\,.xi + b\right) \qquad \text{(Equação 1)}$$

$$Z \in d_{1\,x\,1},\ x \in d_{1\,x\,n,}\ w \in d_{n\,x\,1,}\ b \in d_{1\,x\,1}$$

A função de ativação tem o papel de introduzir características não lineares ao modelo, possibilitando o aprendizado da rede. Essa condição é necessária; caso contrário, a rede ficaria resumida a uma única função linear formada de múltiplas adições e multiplicações entre matrizes e vetores. Além disso, ela não teria a capacidade de aproximar as relações não lineares entre as várias dimensões dos dados de entrada com os dados que se quer prever na saída. Idealmente ela deve ser monotônica para convergência rápida e obrigatoriamente diferenciável para possibilitar o aprendizado da rede durante o processo de *backpropagation*.

Outra característica importante da função de ativação é definir os limites dos valores que cada nó pode assumir. Em um problema de classificação, em que se deseja prever uma

probabilidade, é recomendável que os valores da camada de saída estejam limitados a apenas uma faixa entre 0 e 1.

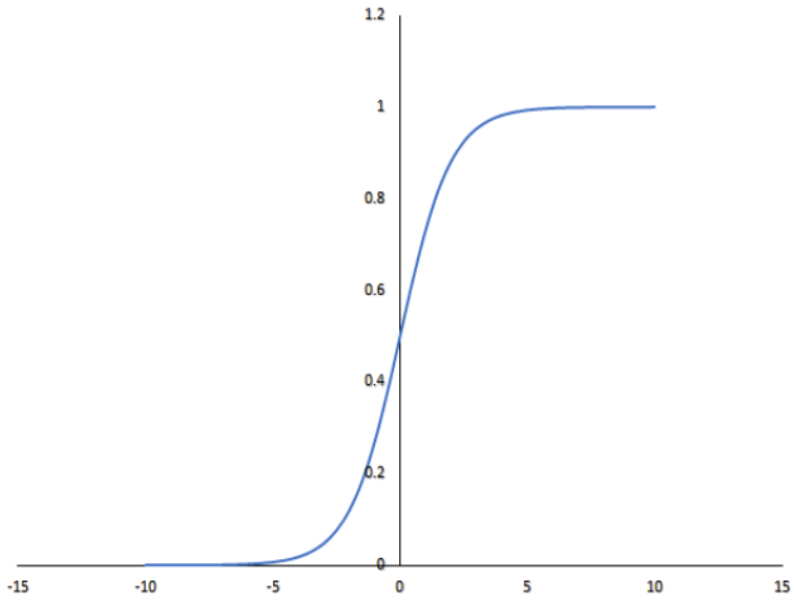
Existem vários tipos de função de ativação. As mais populares são:

- Sigmoidal
- Tangente Hiperbólica
- *Rectified Linear Unit* (ReLU)
- *Softmax*

A seguir serão comentadas brevemente essas quatro funções de ativação.

2.2.1 Função Sigmoidal

O nome da função vem do fato de que ela se parece com um “S” (Figura 8). Como ela é diferenciável e contínua, pode assumir valores na faixa que vão de 0 a 1. É ideal para os casos em que se quer obter uma probabilidade entre classes independentes.



$$f(x) = \frac{1}{1+e^{-x}}$$

$$f'(x) = f(x)(1 - f(x))$$

Figura 8 - Função de ativação sigmoidal.

Fonte: Compilação do autor

A contrapartida dessa função está na faixa de operação da curva que é muito estreita, tendo como efeito a saturação da rede para casos em que os valores de entrada sejam demasiadamente positivos ou negativos. Como a derivada nesses pontos assume valores infinitesimais, o gradiente da função custo utilizada para atualizar os pesos assume valores próximos de zero, prevenindo ou até mesmo impossibilitando o aprendizado.

Outra desvantagem está no fato de a saída não ser centrada em zero e, nesse caso, sempre positiva. Isto faz com que o cálculo do gradiente, durante o processo de otimização, sempre assuma valores totalmente positivos ou totalmente negativos, indiferentemente do sinal de cada peso, movendo todas as atualizações apenas em uma única direção, o que torna o aprendizado mais demorado.

2.2.2 Função Tangente Hiperbólica

A função tangente hiperbólica é similar à função sigmoidal no que tange ao seu funcionamento; porém, com uma faixa de valores que vão de -1 a 1 (Figura 9).

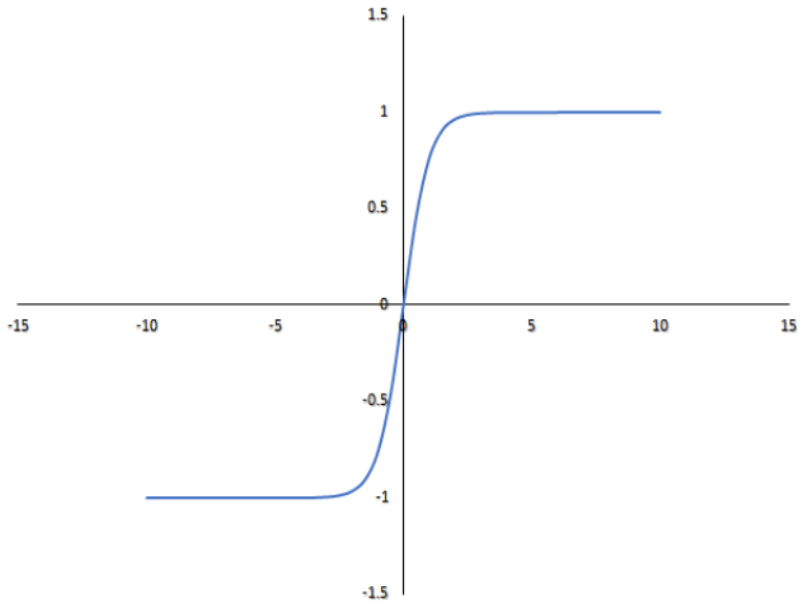


Figura 9 – Função de ativação tangente hiperbólica.

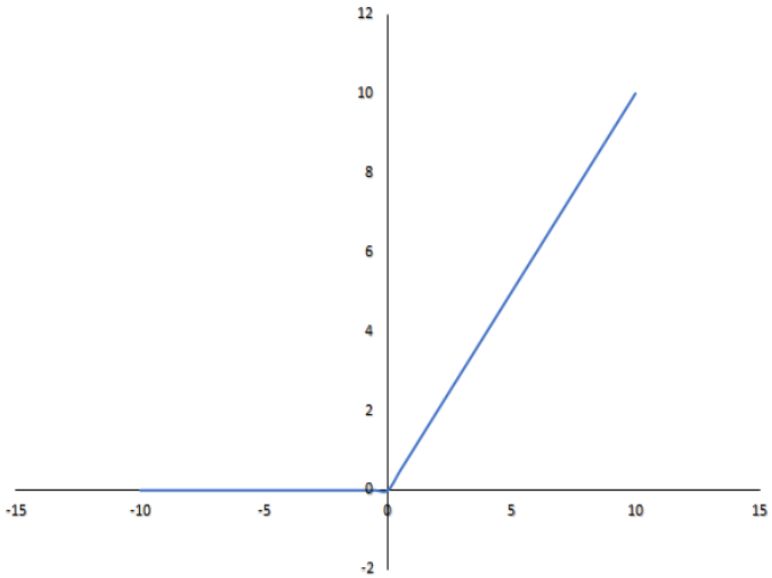
Fonte: Compilação do autor

Ao contrário da função sigmoidal, a função tangente hiperbólica é centrada em zero, o que acelera o processo de aprendizado, mas ainda sofre com valores extremos, dado que o gradiente sai da faixa de operação da função.

2.2.3 Função *Rectified Linear Unit*

***Re**ctified **L**inear **U**nit*, ou simplesmente ReLU, é uma das funções mais utilizadas, principalmente nas camadas ocultas, por ser simples, não linear, e computacionalmente

eficiente (Figura 10). Basicamente, a função ReLU zera qualquer valor abaixo ou igual a zero conservando os demais valores intactos.



$$f(x) = \begin{cases} 0\ para\ \ x \leq 0 \\ x\ para\ \ x > 0 \end{cases}$$

$$f'(x) = \begin{cases} 0\ para\ \ x \leq 0 \\ 1\ para\ \ x > 0 \end{cases}$$

Figura 10 - Função de ativação *Rectified Linear Unit* ou ReLU.

Fonte: Compilação do autor

O grande problema no uso dessa função está em um efeito chamado *dying neurons*: esse efeito ocorre quando apenas entradas negativas são apresentadas durante a etapa de treinamento, condenando a saída a ficar estacionada em zero permanentemente, o que leva a um gradiente nulo impedindo a otimização dos pesos. Na prática, diz-se que este nó ou "neurônio" está morto.

Para contornar tal efeito, existe uma função parecida com essa, chamada *Leaky* ReLU ou *Leaky Rectified Linear Unit* (Figura 11), na qual se introduz um pequeno "vazamento" caso os valores de entrada sejam menores que zero, impedindo gradientes estacionários.

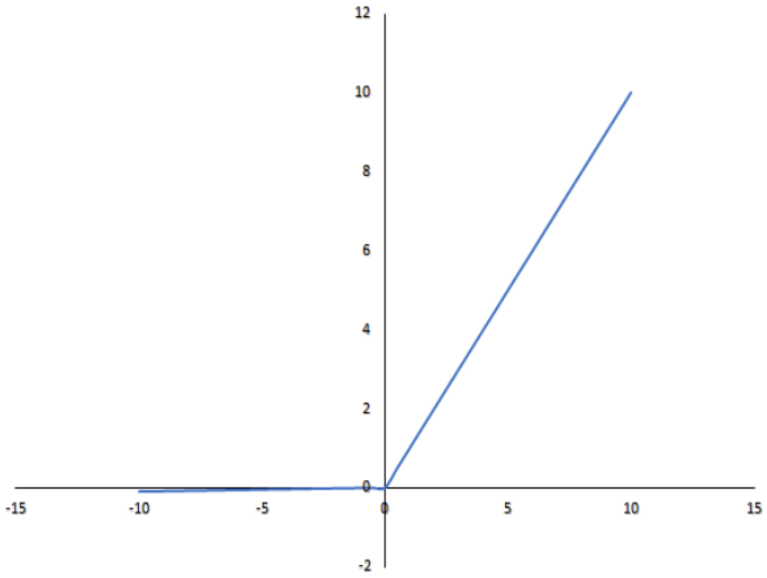


Figura 11 - Função de ativação *Leaky Rectified Linear Unit.*

Fonte: Compilação do autor

2.2.4 Função *Softmax*

A função *Softmax* é completamente diferente das demais, na medida em que ela opera sobre um vetor ao invés de um escalar na entrada. Seu objetivo é normalizar todos os nós de uma camada de forma que possam assumir somente valores entre 0 e 1.

O detalhe importante é que a somatória de todos os nós, após a normalização, é obrigatoriamente 1, o que é fundamental na criação de classes mutuamente exclusivas. Em uma rede treinada para classificar uma imagem entre classes gatos, cachorros e pássaros, contendo três nós na camada de saída, um para cada animal, não se pode ter, por exemplo, a probabilidade de ser gato como sendo 2.0, a probabilidade de ser cachorro 1.0 e a probabilidade de ser pássaro de 0.1. Na Figura 12 pode-se observar o efeito da aplicação da função *Softmax* ao problema descrito acima, tendo como saída a normalização correta das probabilidades para 0.7, 0.2 e 0.1.

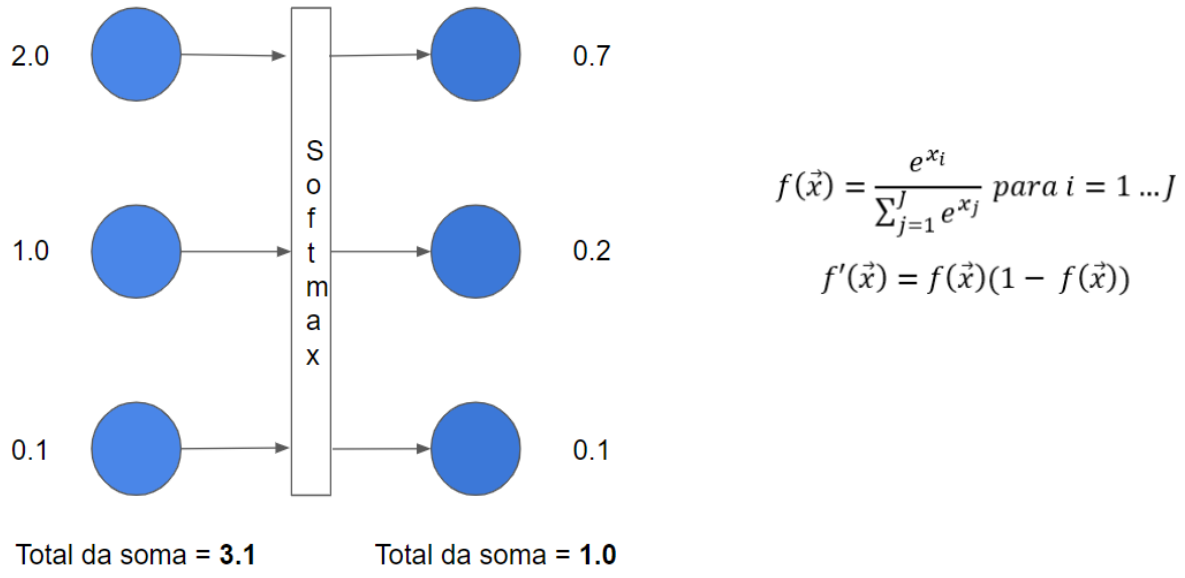


Figura 12 – Função de ativação *Softmax*.

Fonte: Compilação do autor

A função *Softmax* é empregada quase que exclusivamente na camada de saída, sendo muito utilizada em problemas não apenas de classificação, mas também de segmentação de imagens, onde se quer estimar a probabilidade que cada *pixel* tem de pertencer a uma determinada classe.

2.3 Camadas Densamente Conectadas

Apesar de mencionado anteriormente que cada camada é composta de um conjunto independente de nós, que se conecta à camada seguinte por meio de uma rede de pesos, matematicamente e computacionalmente é melhor considerar dois conjuntos de nós interligados como sendo apenas uma camada. As camadas formadas por nós nas quais cada nó liga-se a todos os outros nós da camada seguinte dá-se o nome de camadas densamente conectadas (Figura 13).

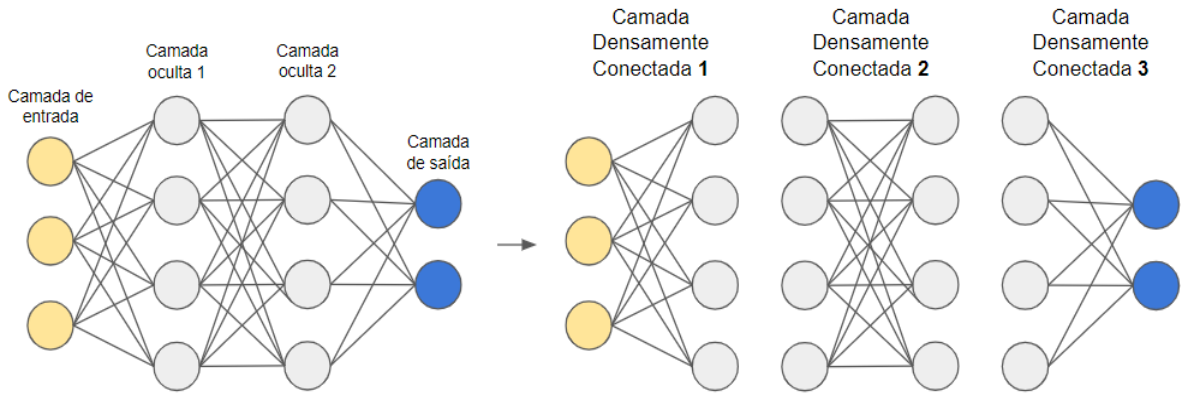


Figura 13 – Representação de uma rede neural densamente conectada.

Fonte: Compilação do autor

Pode-se expressar a operação de uma camada densamente conectada como a aplicação de uma função de ativação **σ** em uma multiplicação de matrizes de pesos **W** pela entrada **X** somados a um vetor de *bias* **b**. A Figura 14 mostra a conversão de uma camada densamente conectada para notação matricial.

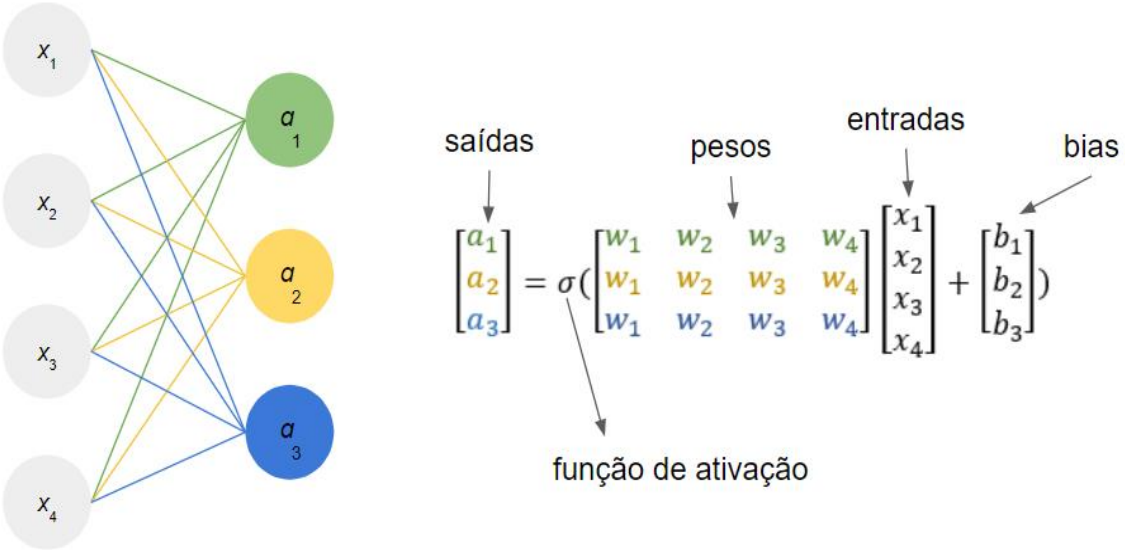


Figura 14 – Representação de uma camada densamente conectada como uma multiplicação matricial.

Fonte: Compilação do autor

A matriz de pesos $\mathbf{W}^{n,m \in \mathbb{Z}}$ tem sua dimensão determinada como a quantidade de nós de entrada *n* pela quantidade de nós na saída *m*. O vetor bias $b^{m \in \mathbb{Z}}$ tem sua dimensão igual ao

número de nós na saída *m*. A equação Y = f(x) descreve o comportamento dessa camada, que pode ser dada pela Equação 2 escrita em notação vetorial.

$$Y = \boldsymbol{\sigma}(X.W + b) \qquad \text{(Equação 2)}$$

$$X \in \mathbb{Z}_{1 \times n},\ W \in \mathbb{Z}_{n \times m},\ b \in \mathbb{Z}_{1 \times m},\ Y \in \mathbb{Z}_{1 \times m}$$

O aprendizado dessas camadas é feito pelo ajuste da matriz de pesos e vetor de *bias* durante a etapa de treinamento. Nesse processo, uma função custo é utilizada para calcular a distância entre o valor previsto de saída e o valor real esperado de forma a mensurar a precisão total da rede. Esse erro então é utilizado para calcular o gradiente de ajuste dos pesos e *bias* para cada camada, começando pela saída retroagindo até a entrada num processo conhecido como *backpropagation*.

2.4 *Feedforward* e *Backpropagation*

Existem duas formas de como as informações trafegam pela rede. Na primeira forma, os dados entram pela camada de entrada seguindo sequencialmente até a camada de saída num processo chamado *feedforward* (Figura 15). Esse processo é utilizado tanto no modo de teste quando se quer apenas prever o comportamento da rede, submetida a uma determinada entrada, como no modo de treinamento quando se quer calcular o erro total da rede, ajustando os pesos.

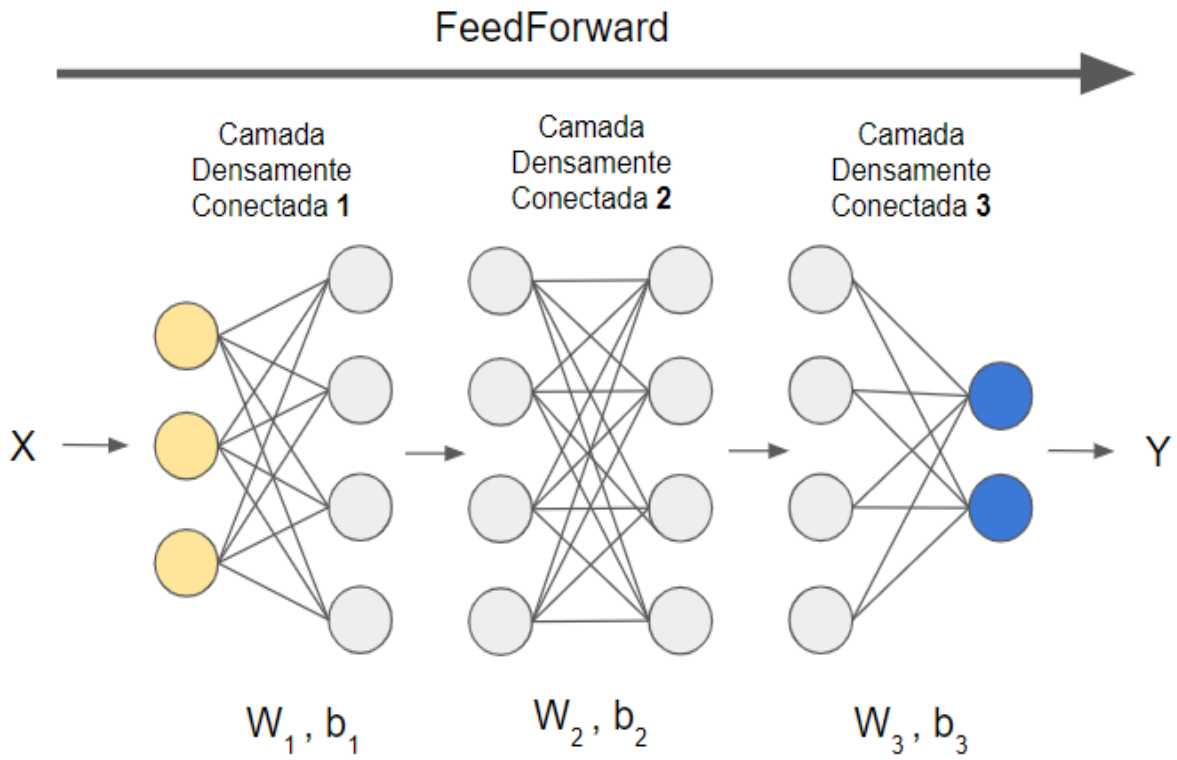


Figura 15 – Operação *Feedforward*: dados fluem da camada de entrada até a saída.

Fonte: Compilação do autor

Em uma rede formada por uma sequência de camadas densamente conectadas, o resultado do processo de *feedforward* pode ser escrito como uma série de operações matemáticas consecutivas, na qual o resultado da operação seguinte depende da saída da camada anterior. Para o exemplo da **Figura 21**, formado por três camadas densamente conectadas, é possível definir a saída Y conforme a Equação 3:

$$Y = \sigma(\sigma(\sigma(\mathrm{X}.\mathrm{W1} + \mathrm{b1}).\mathrm{W2} + \mathrm{b2}).\mathrm{W3} + \mathrm{b3}) \quad \text{(Equação 3)}$$

A segunda forma de veiculação das informações faz-se de forma inversa à primeira, utilizada apenas no modo de treinamento quando se propaga o gradiente de erro da última camada até a primeira, fazendo-se o ajuste dos pesos. Esse processo chama-se *backpropagation* (Figura 16) e sempre ocorre depois do processo de *feedforward*.

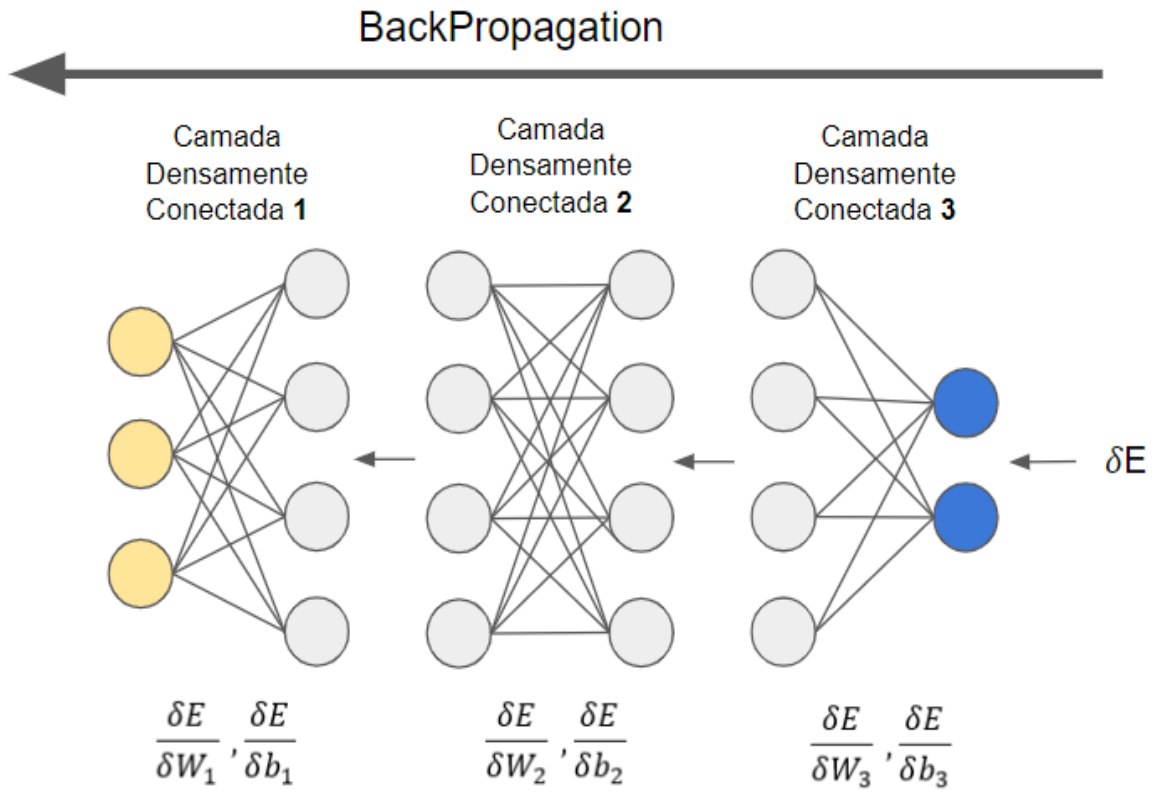


Figura 16 – Operação de backpropagation: o erro retroage até o início da rede com ajustes de pesos.

Fonte: Compilação do autor.

O processo de *backpropagation* é iniciado calculando-se a derivada δE da função custo em relação ao erro entre o resultado previsto e o resultado esperado, retroagindo pelas camadas intermediárias até a camada de entrada para ajustar os parâmetros da rede a cada iteração. Isso permite diminuir o erro global de forma a maximizar a assertividade nos resultados previstos.

O treinamento da rede ocorre de forma progressiva realizando-se várias vezes o processo de *feedforward* seguido de *backpropagation* numa base de dados de forma que, a cada iteração, ocorra o ajuste gradativo e incremental dos parâmetros da rede de forma a minimizar o erro até um patamar aceitável.

2.5 Redes Neurais Convolutivas

As redes neurais convolutivas, *ConvNets* ou redes invariantes ao deslocamento (*shift invariant network*), são um dos grandes avanços das redes neurais voltadas ao reconhecimento de padrões em imagens aplicadas em inúmeros casos, que vão desde a

análise de imagens médicas a sistemas de recomendações, sequenciamento de genomas, processamento natural de linguagem e reconhecimento de imagens em vídeos etc.

Enquanto o modelo tradicional é composto apenas por camadas densamente conectadas, em que todos os nós estão fortemente ligados entre si por meio de uma rede de pesos, as redes neurais convolutivas são compostas não só por estas, mas também por camadas cujos pesos dividem seus parâmetros com múltiplos nós sem estarem fortemente ligado a algum em particular.

Isso permite que se variem as posições das dimensões de entrada sem que haja alterações significativas no resultado de saída. Tomando como exemplo uma tarefa de classificação de uma imagem entre gato ou não gato. Treinando uma rede tradicional, a classificação pode ser correta caso o gato esteja sempre na mesma posição, mas completamente errada caso ele seja colocado em outro lugar uma vez que os pesos relacionados àqueles nós não receberam estímulos suficientes para detectar um gato nessa nova posição. Isso não ocorre na rede neural convolutiva, na medida em que os pesos não aprendem características fortemente ligadas a uma posição fixa, mas em qualquer posição, pois aprendem deslocando-se pela imagem toda.

Popularizado por [22] em seu trabalho voltado à classificação de dígitos manualmente escritos (Figura 17), essas redes são compostas de mais dois tipos de camadas voltadas ao aprendizado espacial, chamadas de camada convolutiva e camada de *pooling*.

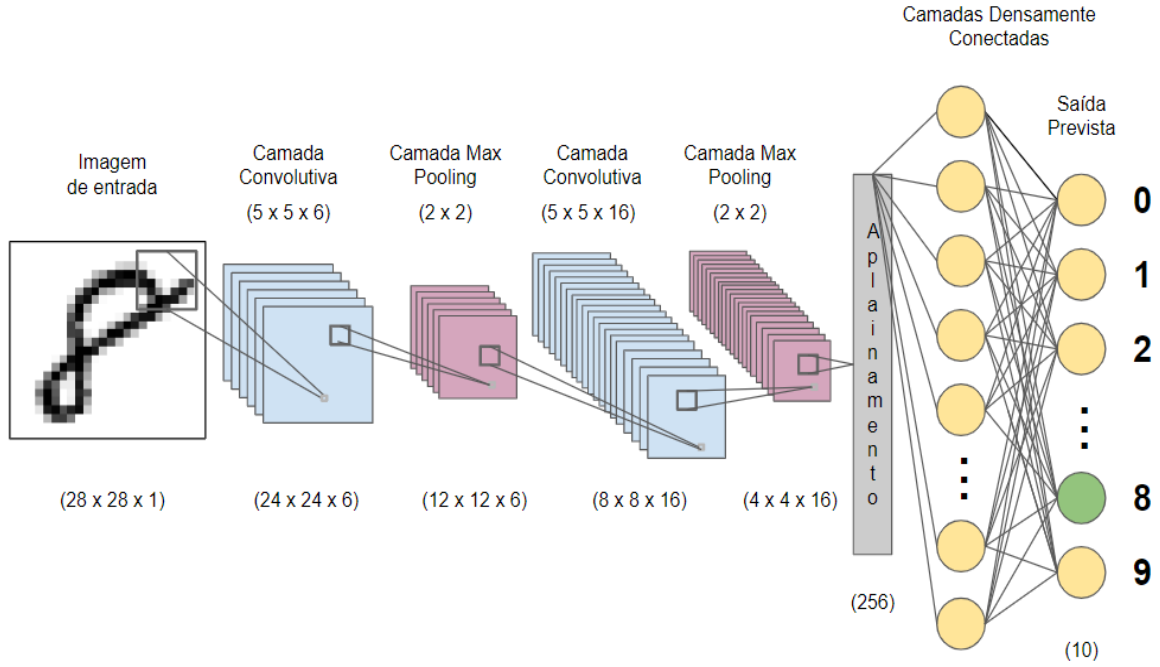


Figura 17 – Rede Neural Convolutiva composta por camadas convolutivas e de *pooling* com a tarefa de classificar uma imagem de um algarismo qualquer em um número.

Fonte: Compilação do autor.

Enquanto o primeiro tipo de camada tem a finalidade de entender as estruturas que compõem uma imagem buscando aprender características que ajudem a resolver uma determinada tarefa, o segundo tipo tem o objetivo de reduzir a quantidade de dados melhorando a performance computacional ao mesmo tempo que reduz o espaço de busca e elimina excesso de informações.

A seguir, serão apresentadas com mais detalhes as operações referentes às camadas novas citadas.

2.5.1 Camadas convolutivas

As camadas convolutivas são o coração das redes neurais convolutivas e, em geral, das redes *deep learning*. Seu principal objetivo é extrair as características mais importantes da imagem de entrada auxiliando no objetivo geral da rede, seja essa classificação, localização ou até mesmo segmentação de objetos.

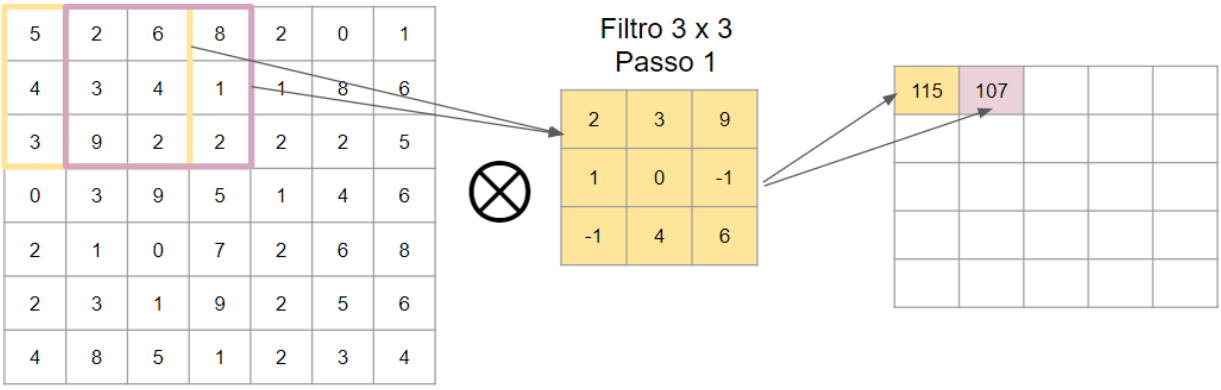
Ao contrário das camadas densamente conectadas na qual todos os nós possuem uma forte conexão entre si, as camadas convolutivas possuem filtros que respondem de forma diferente a determinados campos receptivos na imagem. Isso significa que nem toda região

terá tratamento igual, sendo que aquelas partes contendo os objetos de interesse terão um sinal de resposta maior do que as outras partes.

Cada camada é constituída por conjuntos de filtros que possuem os pesos responsáveis por aprender as intrincadas estruturas que compõem cada imagem, gerando um mapa de características (*feature map*) na saída. Os filtros têm geralmente dimensões pequenas em termos de comprimento e altura. Porém, sempre a mesma profundidade que a imagem de entrada. Por exemplo, para uma imagem colorida de dimensões W_1 x H_1 x 3 no formato RGB, cada filtro possui um tamanho de W_2 x H_2 x 3.

Esses filtros aprendem a extrair esses detalhes importantes de forma automatizada durante o modo de treinamento da rede. Isso é feito sem que haja necessidade de intervenção humana para configurá-los. Esse avanço extingue a necessidade do uso de *feature engineering* para determinar o tipo de característica que se busca detectar na imagem como tradicionalmente era feito, em que havia a necessidade de reconfigurá-los caso algum detalhe da aplicação fosse alterado.

Durante o processo de *feedforward*, essas camadas executam a operação de convolução entre a imagem, ou tensor de entrada, e o conjunto de filtros pertencentes a tal camada. Essa operação é executada (**Figura 18**) deslizando-se cada filtro sobre a imagem e aplicando uma multiplicação matricial ponto a ponto, ou produto escalar, sobre cada região delimitada pelo filtro, produzindo um único valor de saída.



(5*2) + (2*3) + (6*9) + (4*1) + (3*0)+(4*-1) + (3*-1) + (9*4) + (2*6) = 115

(2*2) + (6*3) + (8*9) + (3*1) + (4*0)+(1*-1) + (9*-1) + (2*4) + (2*6) = 107

Figura 18 – Operação de convolução aplicada a uma imagem de um canal utilizando um filtro de tamanho 3 x 3 a passo 1.

Fonte: Compilação do autor

As dimensões do tensor de saída dependem de uma série de parâmetros. Aqui o fundamental é saber o número de conjunto de filtros, altura e largura tanto do tensor de entrada quanto de cada filtro e passo de deslocamento dos filtros (*stride*). Existem parâmetros adicionais como *padding* e *dilation*, referentes à borda e expansão dos filtros que podem ser incluídos quando se quer manter o tamanho da imagem de saída igual ao tamanho de entrada, ou capturar informações de forma mais esparsa.

No exemplo da Figura 19, para uma imagem W^1 x H^1 x C^1 de tamanho 7 x 7 x 3, um conjunto de filtro W^2 x H^2 x C^2 de tamanho 3 x 3 x 3 deslizando a um passo P de 1 x 1, temos uma imagem de saída W^3 x H^3 x C^3 de 5 x 5 x 1, que pode ser calculada pelo seguinte conjunto de fórmulas (Equações 4, 5 e 6):

$$W^3 = \frac{(W^1 - W^2)}{P} + 1 \qquad \text{(Equação 4)}$$

$$H^3 = \frac{(H^1 - H^2)}{P} + 1 \qquad \text{(Equação 5)}$$

$$C^3 = Número\ de\ conjunto\ de\ filtros \qquad \text{(Equação 6)}$$

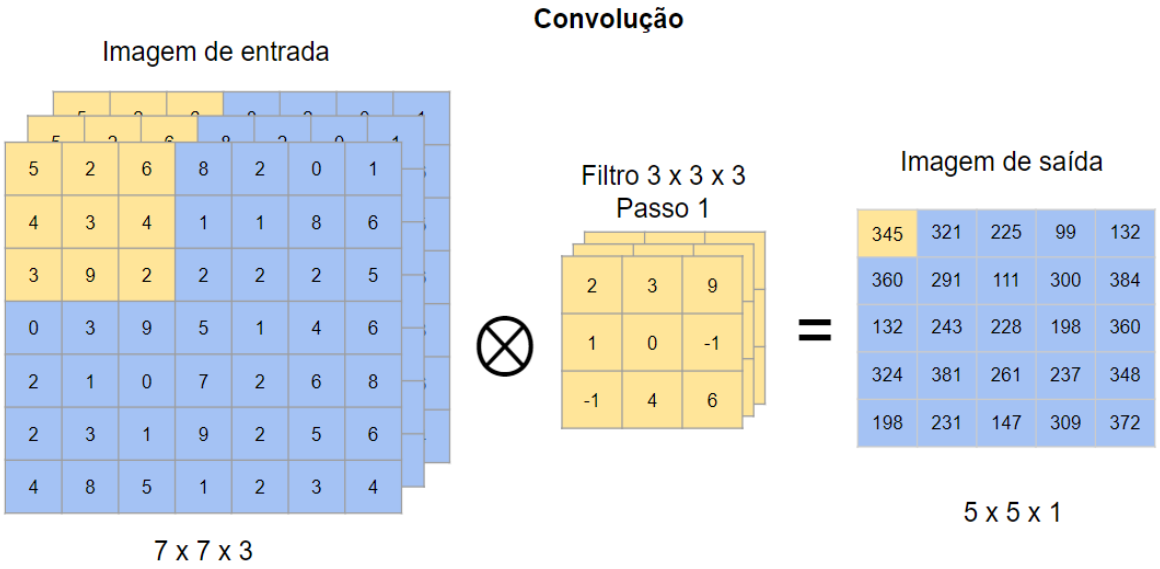


Figura 19 – Operação de convolução: a quantidade de canais de cada filtro é igual à quantidade de canais da imagem de entrada.

Fonte: Compilação do autor.

É importante ressaltar que a operação de convolução sempre resulta em perda de informação. Entretanto, é um efeito desejável uma vez que se quer abstrair apenas as características mais relevantes a serem aprendidas nas próximas camadas. Geralmente, o tamanho da imagem original é reduzido drasticamente ao passo que o número de canais adicionados contendo características particulares é amplamente ampliado.

O resultado da aplicação de múltiplas camadas de convolução é um mapa de características que melhor descreve o tensor original. Ele é comumente utilizado como vetor de entrada para camadas densamente conectadas quando se deseja fazer uma classificação. Em tarefas de segmentação, esse mapa de características é decodificado por camadas de deconvolução ou de *upsampling* para uma dimensão próxima da original, em que cada *pixel* representa a que classe cada *pixel* da imagem original pertence.

2.5.2 Camadas de Pooling

As camadas de *pooling* geralmente são inseridas após as camadas convolutivas. Sua função é reduzir o espaço dimensional das imagens de entrada sem perder a profundidade com vistas a obter os seguintes benefícios:

- Aumento de performance computacional devido à redução do volume de informação transmitida para as camadas seguintes;
- Menor chance de ocorrer sobreajuste uma vez que se elimina o excesso de informação não relevante ao modelo;
- Oferecer algum tipo de tolerância relacionada à variação espacial, pois mesmo que ocorra algum tipo de translação ou rotação, as características mais marcantes são selecionadas.

A operação de *pooling* mais conhecida é a *max pooling* (Figura 20), na qual um filtro bidimensional desliza sobre cada canal da imagem selecionando os valores máximos. O deslocamento do filtro é determinado pelo tamanho do passo, ou *stride*. Aqui, o uso mais empregado é um filtro de tamanho 2 x 2 com passo 2, o que reduz a imagem de entrada pela metade.

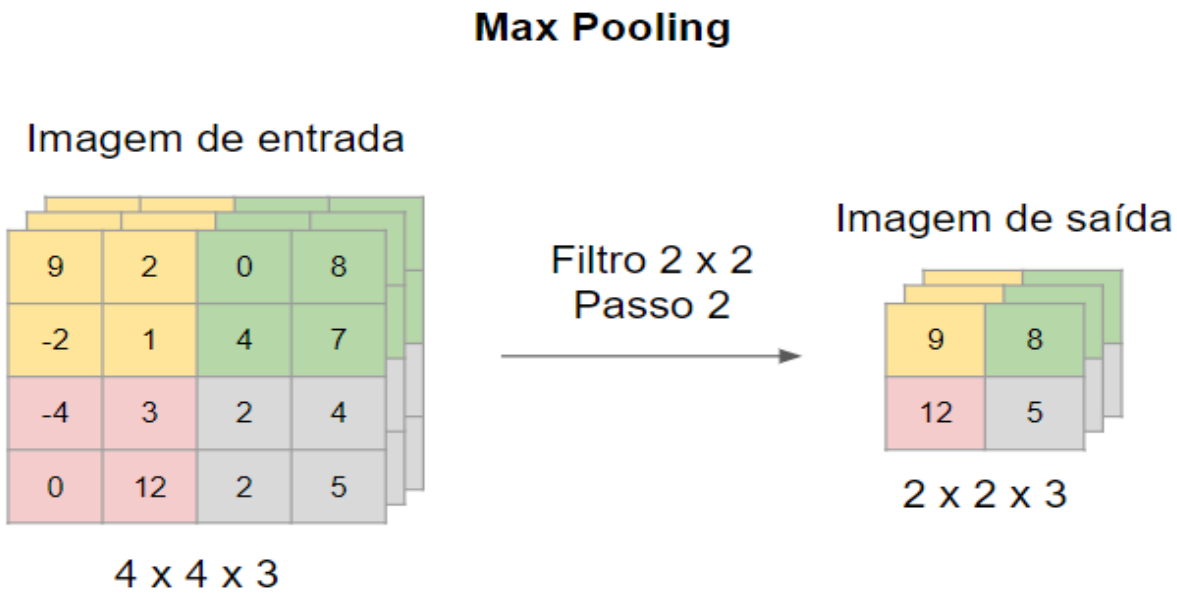


Figura 20 – A operação de pooling reduz o espaço dimensional da imagem de entrada às informações mais relevantes. O deslizante é selecionado.

Fonte: Compilação do autor.

Ao contrário das outras camadas, esta não possui parâmetros de aprendizado que facilitam o processo de *backpropagation* durante o treinamento.

Dada uma imagem de entrada com dimensões $W^1 \times H^1 \times C^1$ pode-se calcular o tamanho da imagem de saída $W^2 \times H^2 \times C^2$ para uma operação de *pooling* com filtro F e passo P, utilizando-se das Equações 7, 8 e 9:

$$W^2 = \frac{(W^1 - F)}{P} + 1 \quad \text{(Equação 7)}$$

$$H^2 = \frac{(H^1 - F)}{P} + 1 \quad \text{(Equação 8)}$$

$$C^2 = C^1 \quad \text{(Equação 9)}$$

Existem outros tipos de operação de *pooling* como *average pooling*, em que é calculada a média de todos os valores delimitados pela área do filtro, ou *stochastic pooling*, na qual a escolha é feita de forma aleatória. Casos mais extremos, como por exemplo, *global average pooling*, reduzem a largura e altura da imagem de entrada em um único valor, conservando apenas o número de canais, sendo mostrada sua eficiência tanto em tarefas de classificação como na localização.

Apesar dos benefícios oferecidos pelas camadas de *pooling*, há trabalhos como [22] que evidenciam bons resultados em tarefas de classificação de pequenas imagens, substituindo tais camadas de *pooling* por camadas convolutivas com passo (*stride*) maior.

2.6 Redes *Deep Learning* para Segmentação

As redes neurais convolutivas, também denominadas redes *deep learning* por conterem múltiplas camadas ocultas, são utilizadas para diversos tipos de tarefa como classificação, localização e segmentação de imagens, sendo este último tipo o principal tema de interesse deste projeto de pesquisa.

A segmentação de imagem envolve particionar uma imagem em múltiplos segmentos ou objetos. Isso é amplamente utilizado em aplicações que vão da análise em imagens médicas para detecção de tumores e doenças até carros autônomos. Existe uma família de arquiteturas de rede *deep learning* chamadas *Encoder-Decoder*, que são especializadas nesse tipo de problema e serão apresentadas detalhadamente mais à frente.

Em um *survey* realizado por [23] sobre aplicação das redes *deep learning* para segmentação de imagens, definiu-se que as redes do tipo *Encoder-Decoder* (Figura 21) são uma família de modelos em duas estruturas que aprendem a mapear os pontos dos dados do domínio de entrada para um domínio de saída.

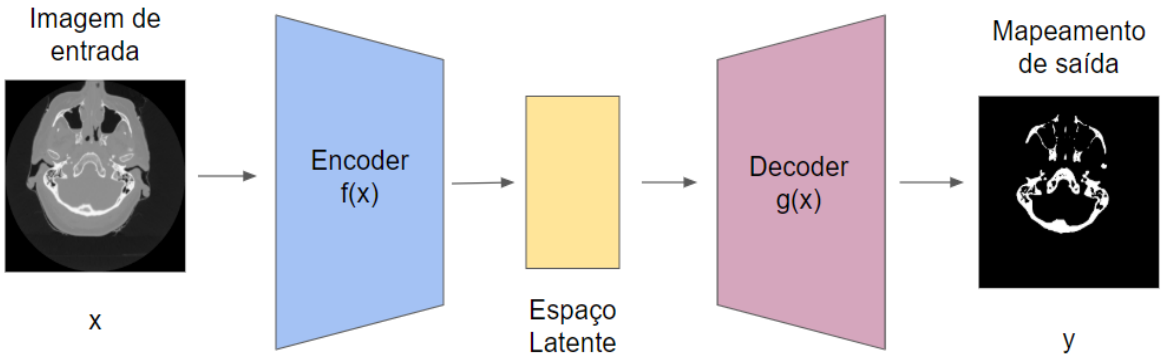


Figura 21 – Arquitetura de rede *deep learning* do tipo *Encoder-Decoder* para segmentação.

Fonte: Compilação do autor.

A parte do *encoder* aprende a reduzir o espaço dimensional aplicando uma função de codificação f(x) a um vetor de características essenciais chamado espaço latente. Nas redes *deep learning*, essa função geralmente é executada por uma sequência de camadas convolutivas e *pooling* combinadas de diferentes formas conforme a topologia escolhida. O *decoder*, também formado, geralmente, por camadas deconvolutivas ou de *upsampling*, tenta reconstruir a saída a ser prevista aplicando uma função de transformação g(x) ao espaço latente de características.

A arquitetura *Encoder-Decoder* não é a única família de arquiteturas utilizadas na segmentação. Existe uma nova família de redes generativas chamadas GAN (*Generative Adversarial Networks*). Essa família é composta por dois modelos de redes, um chamado gerador (*G)* e o outro discriminador (*D)*, que competem entre si na fase de treinamento. Eles procuram gerar saídas próximas do alvo a ser previsto.

Nesta próxima seção serão apresentadas as principais arquiteturas de redes *deep learning* utilizadas na segmentação de imagens.

2.6.1 U-Net

A rede U-Net é uma das arquiteturas de redes *deep learning* mais utilizadas na segmentação de imagens médicas, principalmente em tomografias computadorizadas. Na literatura é possível encontrar várias aplicações na medicina, que vão desde a segmentação de órgãos do corpo como fígado, rins e pulmões [27, 28, 7], até a segmentação de tumores e lesões como os apresentados por [15].

Para aplicações referentes à segmentação óssea do corpo humano, [24] usou esse modelo na segmentação da espinha dorsal com a finalidade de auxiliar médicos em processos pré-cirúrgicos de alto risco. Lee et al. [25] usaram uma combinação dessas redes, a qual chamaram de MGB-Net, para segmentação do osso orbital para cirurgias de reconstrução da parede orbital do crânio maxilo-facial. Já La Rosa [14] fez uso dessa rede para segmentação dos ossos da região abdominal sem um propósito específico.

Desenvolvida em 2015, por Ronneberger et. al [26], a U-Net teve como objetivo segmentar estruturas neuronais capturadas por microscópio eletrônico no desafio proposto pela ISBS (International Symposium on Biomedical Imaging) contendo vários *data sets* com poucas amostras para treinamento. Essa topologia de rede *deep learning* alcançou um "índice de interseção sobre união" (Intersection over Union - IoU) de 92% no *data set* PhC-U373 e 77 % no *data set* DIC-HeLa ficando muito acima do segundo lugar com 83 % e 46 % respectivamente.

A arquitetura da rede consiste em uma parte de contração e outra de expansão. Cada caminho da parte contractiva contém dois blocos formados por uma camada de convolução 3x3 sem *padding*, seguida de uma camada de ativação do tipo *rectifier linear unit* (ReLU), contraindo com uma camada de *max pooling* 2x2. O número de filtros é sempre dobrado após cada bloco de contração seguindo a sequência de 64, 128, 256, 512 e 1024 filtros até o nível mais profundo, reduzindo na volta o número de filtros pela metade a cada bloco de expansão.

Cada bloco de expansão é feito usando-se camadas de *upsampling* ou convolução transposta 2x2 concatenando-se a saída com as *features* da camada de contração apropriada. Uma camada de convolução 1x1 é aplicada ao final para reduzir de um espaço multidimensional de 64 para uma imagem monocromática.

Neste trabalho foram introduzidas camadas de regularização como *batch normalization* e *dropout* para reduzir o *overfitting* durante o treinamento da rede uma vez que a imensa maioria dos *pixels* pertence à classe de fundo nas máscaras. Outra alteração da estrutura original, cf. a Figura 22, foi manter o *padding* nas camadas de convolução de

contração, possibilitando que a máscara de saída tenha as mesmas dimensões em termos de altura e largura da imagem de entrada, evitando a necessidade de aplicar técnicas de ampliação de imagem ao resultado final causando distorções indesejadas.

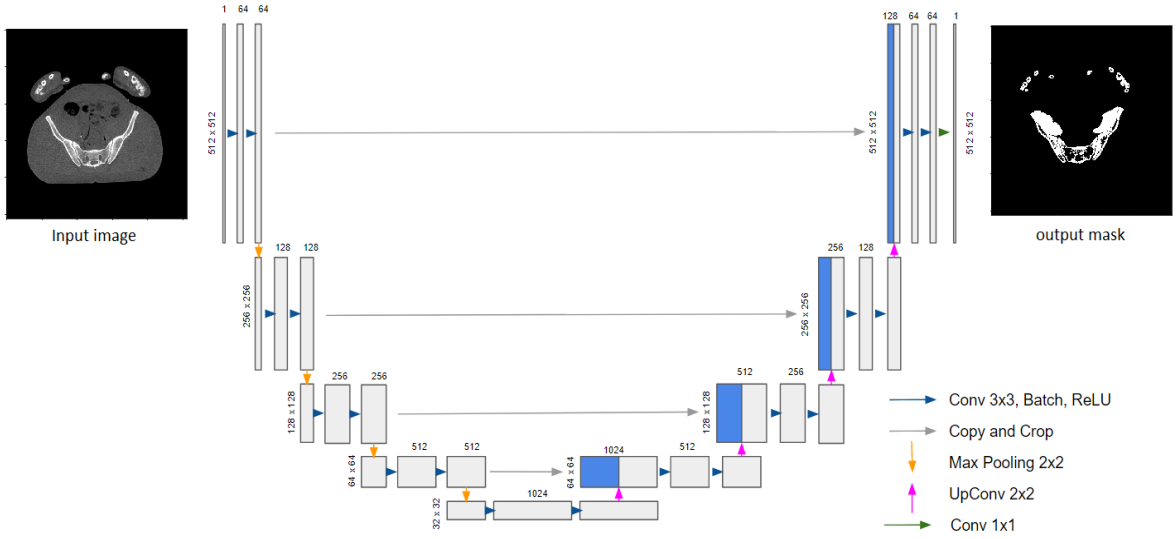


Figura 22 – Topologia U-Net utilizada na segmentação.

Fonte: Compilação do autor.

2.6.2 SegNet

Com artigo publicado em 2016 por [29], SegNet é uma rede *deep learning* no estilo *Encoder-Decoder* com o objetivo de fazer a segmentação *pixel* a *pixel* de uma imagem com foco na tarefa de *scene understanding*, ou entendimento de cena, no qual o algoritmo consegue fazer a separação semântica de cada item que compõe a imagem. Essa aplicação é comumente utilizada em carros autônomos, por exemplo, quando se deseja distinguir carros de pedestres e rodovias para que o controle possa tomar as ações necessárias com relação a cada item detectado.

Em trabalhos relacionados com segmentação médica, tal rede foi utilizada por [9] na segmentação de fissuras e lóbulos em tomografias para tratamento de doenças pulmonares. Já [12] utilizaram essa rede para a localização de pólipos retais em processos de colonoscopia, visando tratamento e prevenção de câncer de cólon.

A arquitetura dessa rede, como pode ser visto na Figura 23, é formada por uma primeira parte composta por 13 camadas, como uma arquitetura VGG16 tradicional, tirando as camadas f*ully connected* no final, comumente usadas na tarefa de classificação, e por uma

parte composta de camadas de *upsampling* e convolução traduzindo o mapa de características de um espaço multidimensional para um mapa de segmentação.

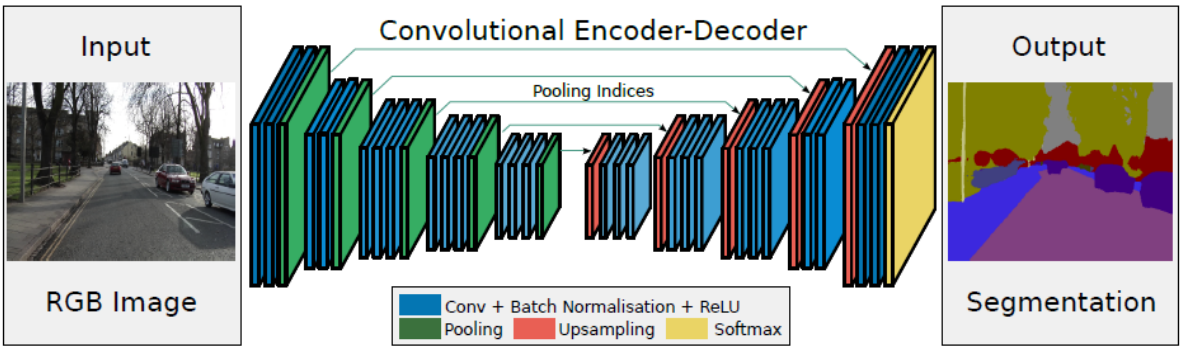


Figura 23 – Topologia SegNet utilizada na segmentação

Fonte: [29]

A principal vantagem dessa rede está na utilização dos índices das camadas de *pooling* como função não linear para as camadas de *upsampling*, o que permite um melhor delineamento do contorno dos objetos. Essa é uma característica muito desejada em tarefas de segmentação médica que requerem precisão uma vez que uma das dificuldades está na precisa separação de órgãos e tecidos. Na segmentação dos ossos da região abdominal, La Rosa [14] sugere que alguns erros de classificação entre o osso sacro e a espinha podem ter ocorrido devido ao problema de interseção. Outra grande vantagem dessa rede é o reduzido número de parâmetros a serem aprendidos quando comparada com outras arquiteturas similares.

2.6.3 DenseNet

As redes neurais convolutivas densamente conectadas, ou DenseNet, foram apresentadas por [30] em 2018 mostrando ganhos de performance nos *benchmarks* de classificação de imagens na competição ImageNet e CIFAR-10, em relação a algoritmos de *deep learning* similares.

Na área médica, essas redes foram modificadas por [31] para comportar imagens tomográficas 3D e utilizadas na segmentação de oito órgãos da região abdominal para auxiliar em procedimentos de endoscopia gastrointestinal. Essa arquitetura também foi utilizada por [24] para segmentar a espinha dorsal auxiliando em processos cirúrgicos de alto risco.

A arquitetura dessa rede *deep learning* consiste em vários blocos com camadas convolutivas, todas conectadas entre si, de forma que a saída de uma camada concatena-se à seguinte sempre formando um bloco maior. Após a execução do bloco, a redução do espaço dimensional é feita por camadas de *pooling*. Uma visão geral da arquitetura é dada na Figura 24. Em tarefas de segmentação, essa rede é modelada como uma rede U-Net em que a informação das camadas de *encoder* é concatenada com as camadas de *decoder*.

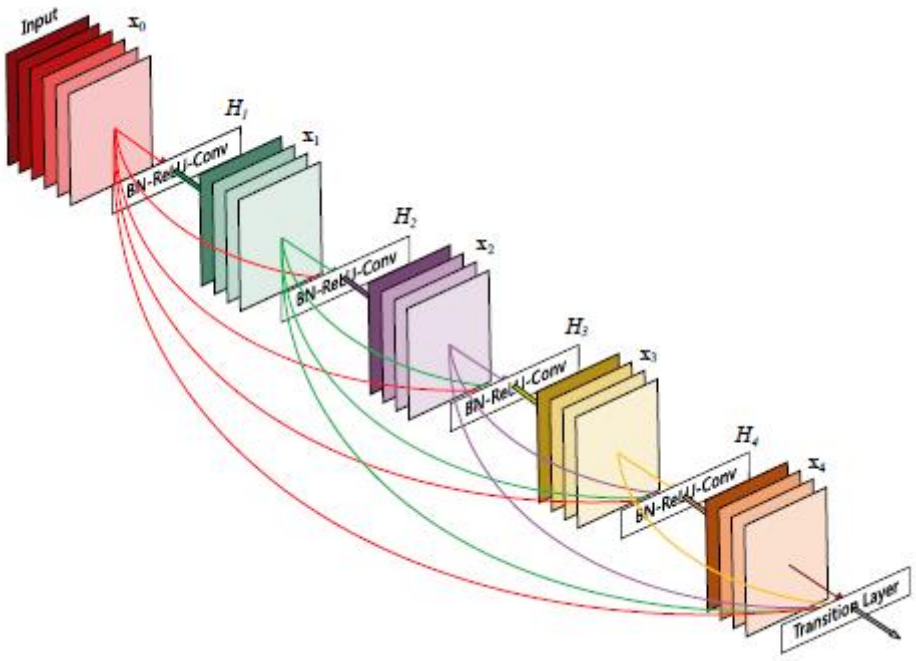


Figura 24 – Topologia DenseNet utilizada na segmentação de imagens.

Fonte: [30].

Existem várias vantagens trazidas por essa arquitetura, sendo que as mais citadas são: necessidade de um número bem reduzido de parâmetros necessários a ser aprendidos quando comparados com arquiteturas tradicionais; robustez quanto a problemas de *vanishing gradient* durante o processo de aprendizado; e acurácia muito superior uma vez que as informações de alta resolução são transferidas para as camadas mais profundas da rede pelo processo de concatenação.

2.6.4 ResNet

A arquitetura de rede residual, ou ResNet, passou a ser amplamente usada desde 2016 quando possibilitou que as redes *deep learning* pudessem conter um número quase que

ilimitado de camadas com capacidade de aprendizado. Sua principal contribuição foi resolver um problema conhecido como *vanishing/exploding gradient* no qual os pesos da rede são praticamente zerados ou tendem a valores infinitos quando a rede possui muitas camadas introduzindo o conceito de *shortcut connection*.

Essa técnica de blocos residuais foi utilizada por [32] numa nova arquitetura chamada *Feature-fusion Encoder-Decoder Network* para segmentação e localização de lesões no fígado. Em uma visão mais abrangente [33] utilizaram redes residuais em combinação com convoluções dilatadas para múltiplas tarefas de segmentação na área médica, que vai desde segmentação do pâncreas para detecção de patologias até a segmentação do cólon para detecção de pólipos numa arquitetura à qual chamaram de *Nested Dilation Networks* (NDN).

A arquitetura das redes *deep learning* residuais são formadas por blocos residuais que podem conter dois ou três conjuntos de camadas convolutivas, seguidas de *batch normalisation* e ReLU. O que define um bloco residual é a *shortcut connection*, como pode ser observado na Figura 25, na qual a entrada do bloco é conectada à sua saída com uma operação de soma executando uma função matemática conhecida como identidade.

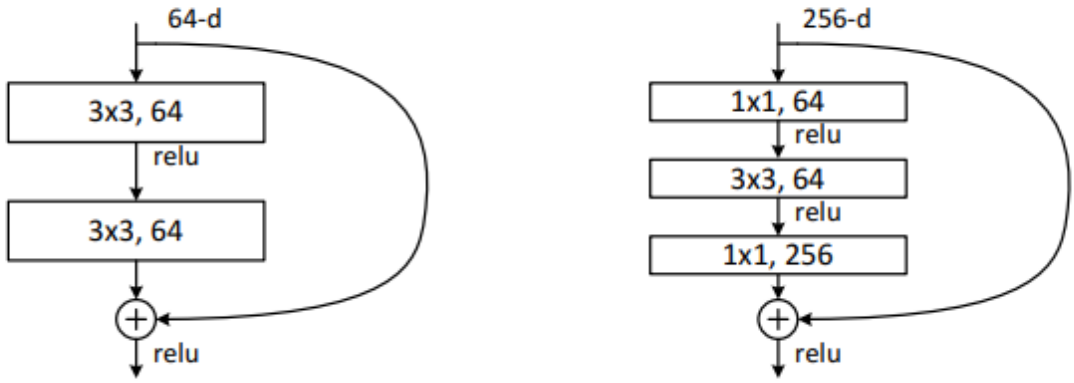


Figura 25 – Blocos residuais contendo *shortcut connection*.

Fonte: [44].

No final de 2016, [34] apresentaram uma revisão dessa arquitetura na qual foi removido o efeito de *down sampling* de algumas camadas para aplicação dessas redes em segmentação de imagens. Esse efeito reduz o espaço dimensional das imagens e, para eliminá-lo nas camadas convolutivas, o passo (*stride*) foi colocado com valor *um* ou aplicado efeito de dilatação maior que um. Já nas camadas de *pooling*, assim como nas camadas convolutivas, o passo foi deixado com valor unitário.

2.6.5 DeepLab

As redes *DeepLab*, apresentadas por [35] em 2017, contribuíram com a tarefa de segmentação semântica resolvendo três problemas distintos:

1. Uso de camadas convolutivas com *upsampling*, chamadas de *atrous convolution*, para resolver o problema de perda de resolução no uso de múltiplas camadas com *downsampling* como convolução ou *pooling* com *stride* maior que um.

2. Aplicação de camadas totalmente conectadas contendo o algoritmo *Conditional Random Field* (CRF) no refinamento dos mínimos detalhes.

3. Aplicação de um processo denominado *Atrous Spatial Pyramid Pooling* (ASPP) no qual ocorre a fusão de múltiplas *atrous convolution* com diferentes *sampling rates* aplicadas à imagem de entrada para capturar diferentes escalas dos objetos que se deseja segmentar.

A combinação dessa rede com a rede Pix2Pix foi utilizada por [36] na segmentação do fígado para prevenção e tratamento de câncer. Inspirados pelo mesmo tipo de arquitetura de rede, [37] desenvolveram uma arquitetura própria com ênfase em detecção de bordas a qual chamaram CDED-net. Ela foi usada para segmentação de pólipos na prevenção de câncer de cólon.

A arquitetura em si já apresenta mais de uma versão, estando atualmente na chamada V3+, apresentada na Figura 26, que utiliza não só a chamada *atrous convolution* como também a *depthwise separable convolution* diferenciando-se da convolução normal no montante em que cada filtro é aplicado a apenas um canal da imagem de entrada, reduzindo assim o número de parâmetros necessários ao aprendizado.

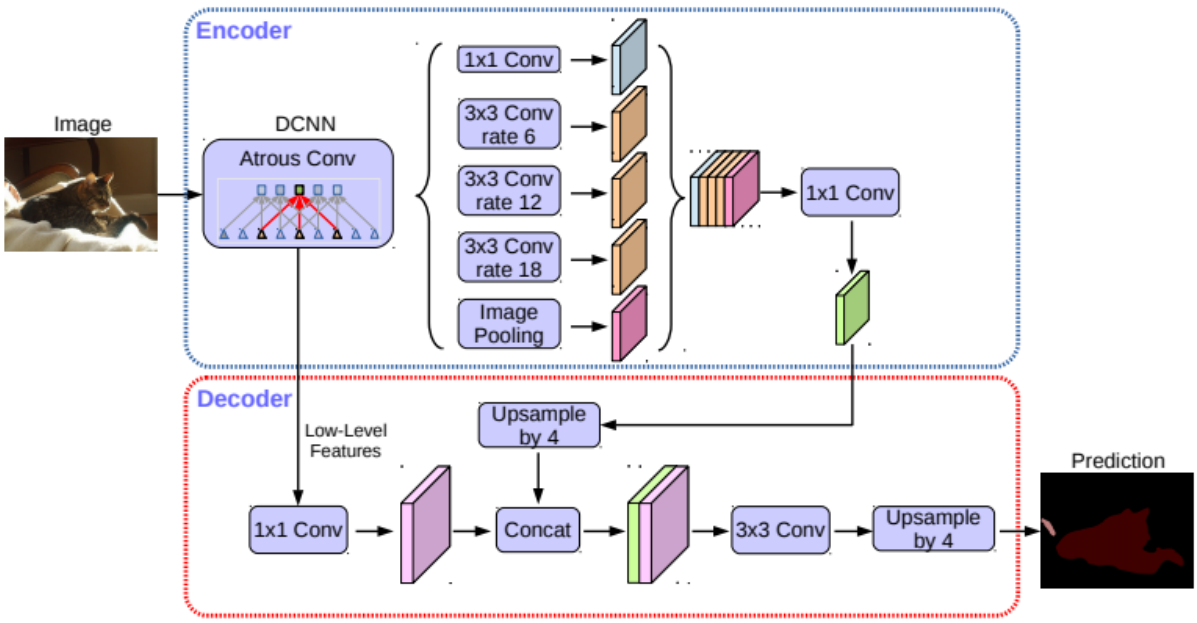


Figura 26 – Topologia DeepLab V3+ utilizada para segmentação.

Fonte: [35].

2.6.6 *Fully Convolutional Network* (FCN)

Em topologias utilizadas para classificação de imagens, tem-se a combinação sequencial das camadas de convolução seguidas das camadas de *pooling* e ativação para extrair características de vários níveis, sendo que o vetor resultante é inserido numa sequência de camadas densamente conectadas para inferir a probabilidade de a entrada pertencer a uma determinada classe.

Aproveitando-se da capacidade de aprendizado das camadas de convolução, [46] publicaram uma topologia de rede completamente convolutiva substituindo as camadas densamente conectadas das topologias de classificação por camadas de *upsampling* ou deconvolução, conforme a Figura 27, restaurando assim a estrutura espacial do tensor de saída ao tamanho da imagem de entrada. A camada de ativação final faz a tradução entre as imagens de entrada e saída inferindo a probabilidade de cada *pixel* pertencer a uma determinada classe.

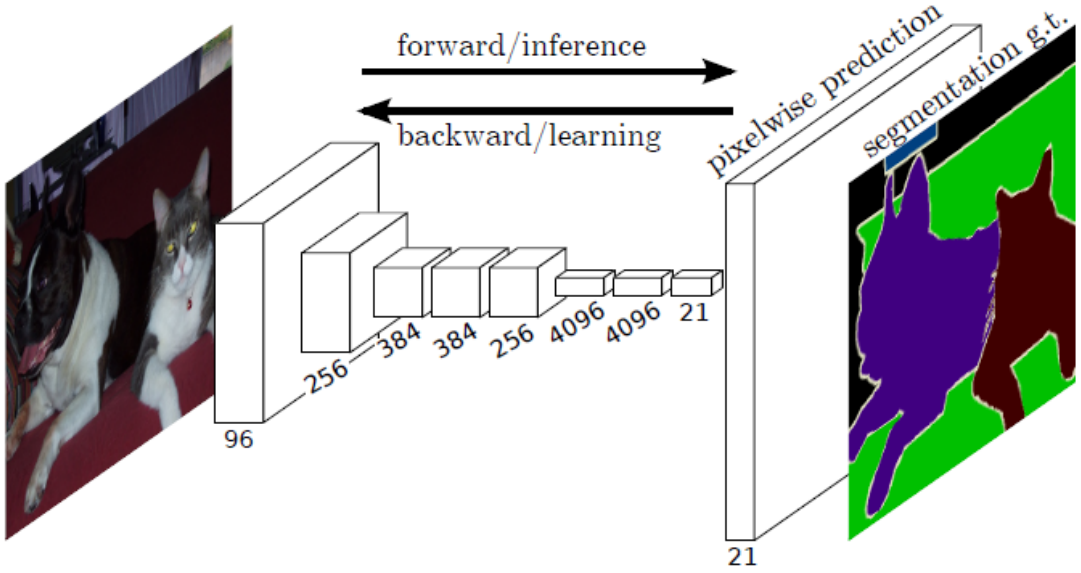


Figura 27 – Topologia FCN utilizada para segmentação.

Fonte: [46].

A reconstrução espacial ao tamanho original é possível ser feita usando-se apenas uma camada de *upsampling*. Porém, parte significativa da informação da imagem original é perdida durante a compactação das características pelas camadas de *pooling*, reduzindo assim a acurácia da segmentação final. A estratégia adotada nessa topologia, para atenuar essa perda, é combinar várias operações de *upsampling* com diferentes *strides* aplicadas às saídas das camadas intermediárias de *pooling* (Figura 28), de forma a combinar informações grosseiras de alto nível das camadas mais externas com informações de baixo nível das camadas internas melhorando assim o resultado.

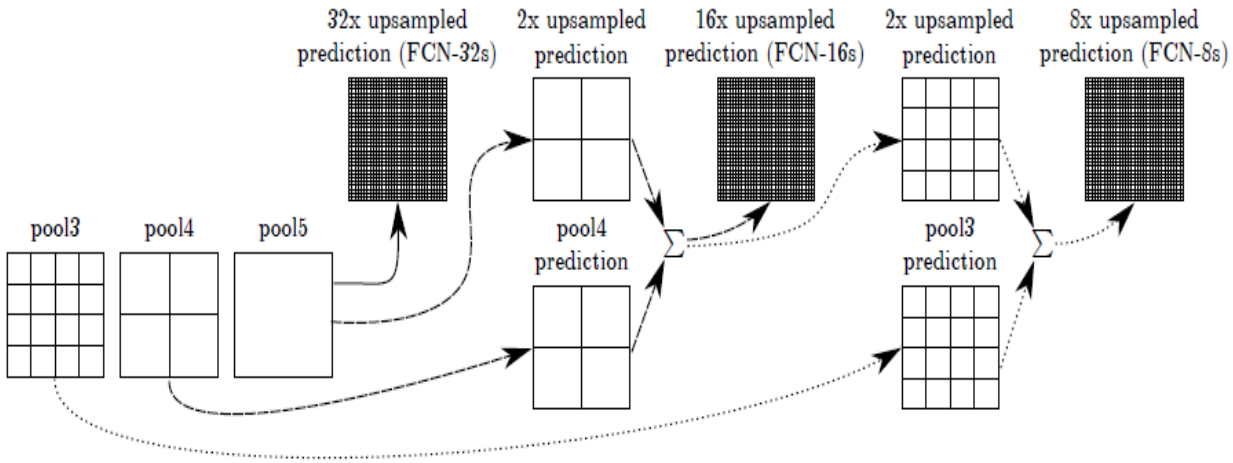


Figura 28 – Utilização de *upsampling* aplicado às camadas *de pooling* para combinar características de alto nível com baixo nível na segmentação final.

Fonte: [46].

2.6.7 *Generative Adversarial Networks*

Os avanços das redes *deep learning* chamadas *Generative Adversarial Networks* ou GAN [38], no campo da visão computacional, estão possibilitando colocar em prática a aplicação da arquitetura *deep learning* em segmentação de imagens. Essa rede, baseada na teoria dos jogos, é formada por dois modelos distintos chamados *generator* e *discriminator*, que competem entre si, criando um mútuo fortalecimento para gerar na saída a resposta correta dada uma imagem real e um sinal de ruído na entrada.

O modelo *generator* possui a finalidade de gerar uma entrada a partir de um modelo de ruído. Sua função é aprender a gerar imagens o mais próximo do real quanto for possível, a ponto de enganar o modelo *discriminator*. O modelo *discriminator*, ou crítico, tem como objetivo estimar a probabilidade de que uma certa entrada seja real ou falsa. Quanto melhor for seu treinamento, mais exigente ele será com os dados apresentados e, consequentemente, o resultado da saída da rede será melhor e mais realístico. A composição dessa arquitetura pode ser vista na Figura 29.

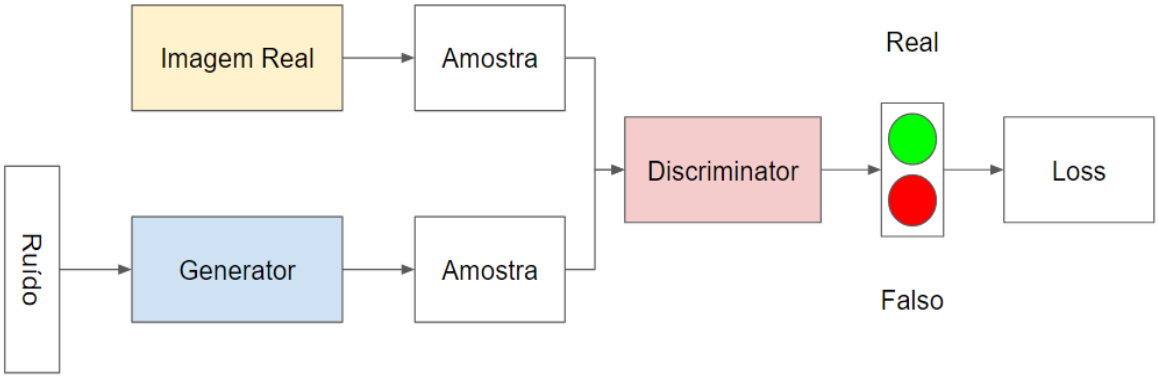


Figura 29 – Composição da arquitetura *Generative Adversarial Networks* (GAN).

Fonte: Compilação do autor.

Dentre os diversos modelos de GANs existentes, a arquitetura mais empregada na tarefa de *image-to-image translation* usada na segmentação de imagens chama-se cGAN, *Conditional* GAN ou Pix2Pix [39]. Essas redes foram empregadas tanto por [36] combinadas com DeepLab na segmentação dos pulmões como, por exemplo, em [40], obtendo resultados impressionantes no desafio de segmentação cardíaca.

A aplicação das redes adversárias não se limita apenas à segmentação de órgão na área médica. Tem outras inúmeras finalidades, como gerar imagens sintéticas auxiliando no treinamento de outras redes [41] ou até mesmo para o uso indiscriminado de imagens médicas de múltiplos domínios, como Tomografias Computadorizadas (TCs), Ressonâncias Magnéticas (RMs) e Tomografias por Emissão de Pósitrons (TEPs) para treinamento e uso em uma única rede [42].

2.7 Considerações finais

Neste capítulo introduziram-se, inicialmente, os conceitos de redes *deep learning* relatando sua origem, evolução e aplicações ao longo dos anos. Apresentou-se a base teórica demonstrando como tais redes *deep learning* são construídas, quais os principais componentes que as compõem, de que forma aprendem, finalizando com a ilustração de várias topologias utilizadas na tarefa de segmentação de imagens médicas.

No próximo capítulo serão apresentados trabalhos relacionados ao uso de redes *deep learning* na segmentação de imagens médicas de órgãos do corpo humano, trazendo suas principais aplicações. Esse detalhamento será útil na etapa comparação de desempenho das

redes *deep learning* na segmentação de imagens médicas vis-à-vis às principais topologias utilizadas.

3 Trabalhos Relacionados

Segmentação em imagens médicas utilizando redes *deep learning* tem se destacado na área acadêmica. Este capítulo apresenta uma revisão da literatura, abrangendo 37 artigos sobre o tema. Essa revisão aborda as principais aplicações e topologias de redes *deep learning*, bem como as métricas de avaliação de desempenho. No final, serão apresentadas pesquisas direcionadas ao uso das redes *deep learning* aplicadas à segmentação óssea, bem como um quadro-resumo contendo informações relevantes de cada artigo.

Todos os artigos aqui apresentados focam na segmentação de imagens médicas do tipo Tomografias Computadorizadas (TC). Outras modalidades, como Ressonância Magnética (RM) e Tomografias por Emissão de Pósitrons (PETs), podem também estar contidas na análise, mas não fazem parte do objeto principal de estudo.

3.1 Aplicações de redes *deep learning* à segmentação médica

Em meados de 2000, com a chegada de computadores mais rápidos, i.e., contendo placas de aceleração gráficas poderosas consolidaram-se as aplicações de redes neurais de aprendizado profundo, ou redes *deep learning*. Tais redes são capazes de aprender a identificar o melhor conjunto de filtros para segmentar imagens médicas sem a necessidade de serem explicitamente definidos por uma pessoa.

Desde então, muitas pesquisas foram realizadas nessa área focando principalmente na detecção, prevenção e combate a doenças graves. Os principais destaques estão nas pesquisas ligadas ao uso de redes *deep learning* para segmentar tomografias computadorizadas dos pulmões [4, 5, 12, 6, 7, 8, 9], com a finalidade de detectar nódulos pulmonares visando combater o câncer pulmonar, bem como de segmentar imagens tomográficas computadorizadas do fígado [47, 27, 49, 48, 50, 32, 15, 51, 36, 52, 13] tanto para o rastreamento de tumores e lesões como para o auxílio em atividades pré-operatórias.

Outras pesquisas aplicadas diretamente à área médica estão ligadas ao uso de redes *deep learning* para segmentar a superfície cerebral em atividades pós-cirúrgicas de pacientes com epilepsia [19], segmentar o esôfago para tratamento de câncer [54, 55, 56] e segmentar a espinha dorsal [43, 57, 24] para auxiliar em atividades pré-cirúrgicas.

É importante mencionar que nem todos os artigos têm como objetivo direto o tratamento de alguma enfermidade. Em [40] propõe-se uma topologia de rede adversarial do tipo GAN (*Generative Adversarial Network*), em que o foco principal é o aprendizado

multimodal que possibilita tanto o uso de tomografias computadorizadas (TC) quanto ressonâncias magnéticas (RM) de forma indiscriminada na segmentação de imagem cardíaca.

A **Tabela 1** traz uma visão geral de quantas vezes cada órgão é utilizado na segmentação médica nos artigos pesquisados nesta dissertação; deve-se notar que um mesmo artigo pode usar vários órgãos simultaneamente.

Tabela 1 – Frequência de órgãos segmentados dentre 37 artigos analisados.

Órgão segmentado	***Frequência***
fígado	11
pulmão	7
pâncreas	3
espinha dorsal	3
rim	3
esôfago	3
cérebro	2
coração	2
baço	1
próstata	1
bexiga	1
reto	1
aorta	1
traqueia	1
outros	3

3.2 Topologias de redes *deep learning* usadas na segmentação

Como a forma de montar uma rede *deep learning* é completamente modular, i.e., várias camadas podem ser combinadas de diferentes formas para produzir uma topologia única, o que torna difícil escolher, dentre todas as possibilidades, aquelas que melhor representam o uso na tarefa de segmentação de imagens médicas. Em [3] é feita uma revisão sistemática da literatura abordando com detalhes os modelos U-Net, *Fully Convolutional Network* (FCN), V-Net, *Convolutional Residual Networks* (CRNs) e *Recurrent Neural Networks* (RNNs) e seu uso na segmentação de tomografias computadorizadas.

Complementarmente, o estudo incluiu também um levantamento, nos 37 artigos selecionados com foco restrito à segmentação de imagens de órgãos em TCs, o uso dos

seguintes modelos: AlexNet, DenseNet, SegNet, DeepLab, entre outros. A frequência de utilização de cada topologia está discriminada na **Tabela 2**.

Tabela 2 – Frequência de utilização por topologia dentre 37 artigos analisados.

Redes	***Frequência***
U-Net	12
FCN	9
DeepLab	3
AlexNet	2
DenseNet	2
ResNet	2
GAN	2
SegNet	2
DeepBeliefNet	1
SharpMask	1
Outros	4

É importante ressaltar que um modelo-base pode originar vários outros modelos, encerrando pequenas alterações como, por exemplo, o modelo U-Net. Esse modelo, primariamente, foi concebido para segmentar imagens de tomografias 2D e, posteriormente, foi adaptado para segmentar em 3D [15] imagens de tumores do fígado.

No levantamento, cf. mostra a **Tabela 2**, pode-se concluir que o modelo U-Net é o mais utilizado na literatura estudada, seguida pela topologia FCN. As redes AlexNet, SegNet, DenseNet, ResNet e DeepLab são mencionadas em mais de um artigo.

Visto que todo o processo de treinamento, teste e comparação entre as redes *deep learning* é demorado, foram selecionadas as seis topologias mais frequentemente citadas cf. **Tabela 2**, construiu-se um protótipo de avaliação e realizou-se um teste preliminar para conferir se há convergência de forma incremental no aprendizado da rede a cada iteração de treino.

Os resultados mostraram que todas as redes tiveram um bom desempenho de aprendizado após 100 épocas de treinamento, exceto a rede SegNet e AlexNet que estagnaram nas primeiras iterações. Sendo assim, as topologias de redes *deep learning* selecionadas para compor o estudo comparativo nesta dissertação foram: U-Net, FCN, DenseNet, ResNet e DeepLab.

3.3 Métricas de avaliação

Medir a eficiência de qualquer rede *deep learning* é parte importante do processo de avaliação e comparação dos resultados entre diferentes topologias. A partir do estudo da literatura, foram levantadas todas as métricas de aferição de performance das redes *deep learning* citadas nos 37 artigos analisados. Pode-se afirmar, cf. mostrado na **Tabela 3**, que a principal métrica utilizada é o coeficiente Dice.

Tabela 3 – Frequência de utilização por métrica dentre 37 artigos analisados.

Métrica	***Frequência***
Dice	32
Accuracy	5
Jaccard	4
Sensitivity	4
IoU	3
Average Surface Distance	3
Positive Predicted Value	2
Specificity	2
Volume Overlap Error	2
Hausdorff Distance	2
Relative Absolute Volume Difference	1
Maximum Surface Distance	1
Relative Error	1
Contour Mean Distance	1
Recall	1
Precision	1
Modified Hausdorff Distance	1
Mean Absolute Distance	1
Absolute	1
Mean Volume Difference	1

O coeficiente Dice, ou *Sørensen–Dice*, é uma métrica utilizada para avaliar o grau de similaridade entre duas amostras. Ele é dado pela Equação 10, onde a amostra X é comparada com uma amostra Y.

$$Dice = \frac{2|X \cap Y|}{|X| + |Y|} \qquad \text{(Equação 10)}$$

No caso de segmentação médica, a métrica é usada para calcular o grau de sobreposição entre duas imagens booleanas, em que cada pixel é avaliado como pertencente ou não pertencente a uma determinada classe. Para esse caso específico, pode-se usar a Equação 11, que permite calcular o coeficiente usando os seguintes valores: VP para Verdadeiro Positivo (quando o *pixel* previsto pertence à classe correta), FP para Falso Positivo (quando o modelo classifica um *pixel* em uma das classes de osso e o correto é como pertencente ao fundo) e FN para Falso Negativo (quando erra a classe à qual o *pixel* pertence).

$$Dice = \frac{2VP}{2VP + FP + FN} \quad \text{(Equação 11)}$$

Os *pixels* classificados como VN (Verdadeiro Negativo) não têm qualquer contribuição na equação, dado que no caso de tomografias computadorizadas eles representam o fundo ou áreas sem qualquer informação relevante, sendo assim descartados na avaliação final pela equação.

Esse coeficiente é amplamente utilizado na literatura (cf. mostrado na **Tabela 3**) para avaliação do desempenho da segmentação em imagens médicas. Sendo assim, ele foi o escolhido como a única métrica de comparação entre as diferentes redes *deep learning* selecionadas na seção anterior, a saber: U-Net, FCN, DenseNet, ResNet e DeepLab.

3.4 Segmentação óssea

Os ossos do corpo humano foram selecionados como itens a serem segmentados a partir de imagens de tomografias computadorizadas, utilizando-se redes *deep learning.* É importante considerar que o objetivo do estudo é comparar o desempenho de diferentes topologias de redes *deep learning* para a segmentação de imagens de tomografias computadorizadas,. Deve-se notar, contudo, que outros estudos já fizeram aplicação de redes neurais nesse tipo de tarefa.

Três estudos [14, 24, 43] com aplicação de redes *deep learning* à segmentação óssea foram escolhidos para comparar algumas características em itens comuns a todas as pesquisas e a esta dissertação. A **Error! Reference source not found.**

Quadro *1* mostra os tamanhos das bases de dados para treinamento, número de classes de ossos a serem segmentados e métricas de mensuração de desempenho das redes *deep learning*.

Quadro 1 - Redes deep learning especificamente para segmentação óssea.

Referência	Qtd. classes	Classes de óssos	Qtd. Base de Treino	Rede	Métrica
[14] La Rosa, F. (2017)	6	vértebras, sacrum,bacia, costela, fêmur e esterno	15.653	U-Net	Dice
[24] Kuok et al. (2018)	1	vértebras	200	DenseNet	Dice
[43] Fang et al. (2018)	1	vértebras	4.500	FCN	Acurácia
Esta dissertação.	**18**	**crânio, mandíbula, clavícula, escápula, úmero, rádio, ulna, mãos, costelas, esterno, vértebras, sacrum, bacia, fêmur, patela, tíbia, fíbula e pés**	**4.688 (Base masculina VHP + 20 bases IRCAD)**	**U-Net,** DenseNet, ResNet, DeepLab e FCN	**Dice**

Enquanto La Rosa utiliza uma base contendo 15.653 amostras para treinamento de uma única rede U-Net para a segmentação de seis classes de ossos da região abdominal, nesta dissertação foram utilizadas várias bases, contendo 4.688 (1.865 masculino *VHP* + 2.823 IRCAD) amostras para segmentar entre 18 classes de ossos do corpo inteiro (1.730 feminino VHP) sobre cinco topologias de redes distintas, o que representou um desafio significativo em termos de volume de experimentos. A única classe comum a todos os trabalhos é a *vértebra* e, como constatado anteriormente, a métrica Dice é a predominante para a mensuração de desempenho.

É importante destacar que, enquanto os três trabalhos [14, 24, 43] procuram alcançar a melhor segmentação treinando uma topologia de rede *deep learning*, aqui se propõe investigar se há diferença significativa nos resultados das segmentações para cinco topologias diferentes de redes *deep learning*.

As próximas seções trazem de forma mais detalhada os três artigos selecionados.

3.4.1 Uma Abordagem Usando Redes *Deep Learning* Para Segmentação Óssea em Tomografias Computadorizadas.

La Rosa faz uso de redes *deep learning* do tipo U-Net em imagens de tomografias computadorizadas para segmentar os ossos da região abdominal em seis classes distintas: *coluna vertebral, quadril, esterno, costelas, osso sacro e fêmur*. A arquitetura U-Net utilizada recebe como entrada fatias em 2D com cinco níveis de profundidade e sete canais de saída, sendo uma para cada classe mais um canal para representar os *pixels* pertencentes ao fundo.

A base de imagens de tomografias computadorizadas utilizada consistiu em 21 *scans* abdominais, perfazendo 15.653 fatias 2D de tamanho 512x512 *pixels*. As máscaras *ground truth* utilizadas no processo de treinamento foram criadas por meio de um processo semi-automático, aplicando-se a técnica *Otsu threshold* seguida por algumas operações de convolução para suavizar cada imagem e remover ruídos. As máscaras utilizadas na base de teste foram feitas por especialistas na área em quatro *scans* de abdome completos totalizando 3.370 fatias.

O treinamento da rede foi feito utilizando-se um lote de oito imagens de tomografias computadorizadas em 150 épocas com o algoritmo *Adam* para otimização dos parâmetros [58]. A taxa de aprendizado escolhida foi de 0.0001. Para reduzir problemas de *overfitting*, uma camada de *dropout* com probabilidade de 0.4 foi aplicada entre as duas camadas de convolução para cada caminho de contração e expansão da rede, assim como uma camada de *batch normalization*. As principais técnicas de *data augmentation* utilizadas foram translação horizontal e vertical, rotação, *zoom* e espelhamento vertical.

Os resultados finais obtidos de La Rosa foram avaliados utilizando-se três métricas distintas: coeficiente Dice, *Recall* e *Precision*. O coeficiente Dice médio obtido na base de validação, levando-se em conta todas as seis classes, foi de aproximadamente 0.93. Os resultados de *Recall* e *Precision* foram utilizados para plotar uma curva e verificar o melhor valor de corte a ser utilizado para distinguir cada classe com os resultados de saída da rede *deep learning*.

Os principais desafios apontados na segmentação do modelo tinham a ver com as extremidades dos ossos das costelas, que são muito fragmentados e se confundem facilmente com as cartilagens; também com a intersecção da espinha dorsal com o osso sacro e algumas partes das clavículas que são classificadas erroneamente como costelas.

Outro teste realizado foi quanto à capacidade do modelo operar com a introdução de ruído gaussiano nas imagens de tomografia computadorizada. Foram aplicados diferentes

valores de ruído e verificou-se que o coeficiente Dice reduziu de 0.93 para 0.90, o que reflete uma boa resistência a esse tipo de problema comumente encontrado durante a aquisição das tomografias pelos aparelhos de tomografia computadorizada, o qual cai drasticamente conforme o ruído é ampliado.

O artigo finaliza propondo uma melhoria para os resultados aplicando o modelo U-Net de 2D para 3D por meio da composição de três modelos em projeção ortogonal e voto majoritário para decidir a que classe cada voxel pertence, ou utilizando convoluções em 3D em conjunto com redes do tipo GAN no pós-processamento para remover possíveis ruídos. Contudo, também ressalta as dificuldades do treinamento dessas redes *deep learning* devido a limitações das GPUs.

3.4.2 Uma abordagem eficaz utilizando CNN na segmentação de vértebras em TCs 3D.

Segmentação óssea é um importante passo nas atividades preparatórias para cirurgias de alto risco como, por exemplo, dos ossos das vértebras. O artigo em questão propõe o uso de redes *deep learning* do tipo FC-DenseNet para segmentar as vértebras em tomografias computadorizadas, cf. mostra a Figura 30.

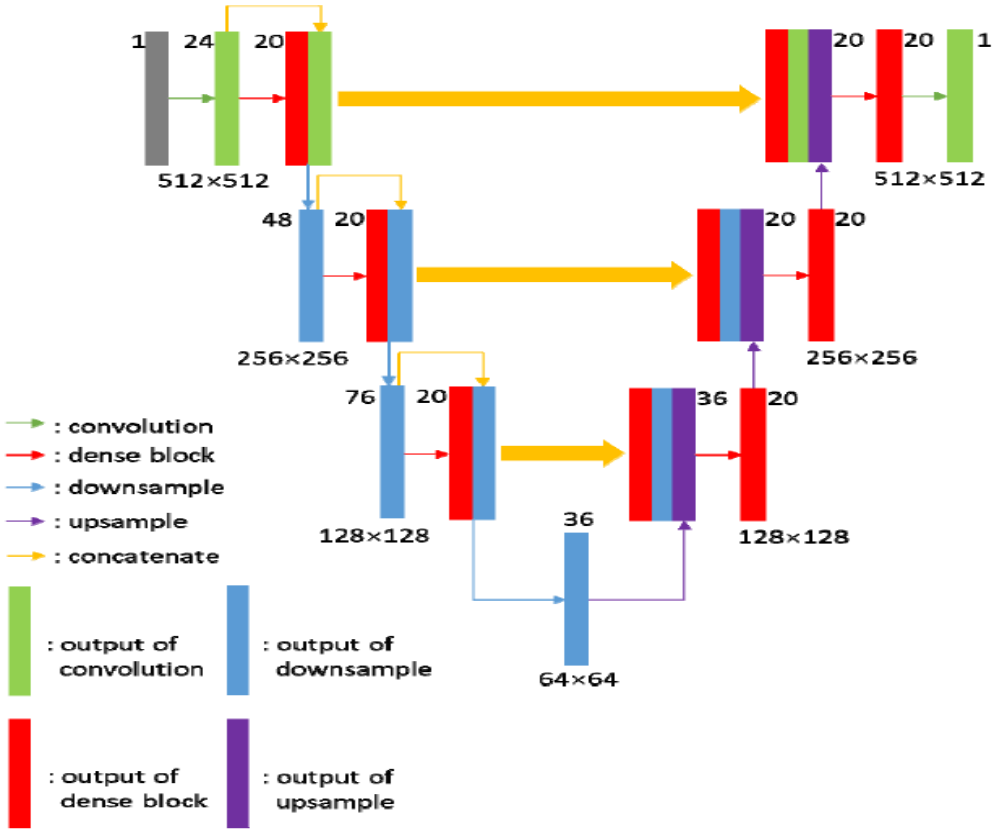


Figura 30 – Topologia FC-DenseNet para segmentação das vértebras.

Fonte: [24].

Cinco casos médicos, não disponibilizados publicamente, fornecidos pela National Cheng Kung University Hospital e Kaohsiung Veterans General Hospital de Taiwan foram utilizados tanto para o treinamento quanto para os testes das redes *deep learning*. As imagens tomográficas têm tamanhos padrões de 512x512 contendo uma distância de aproximadamente 0.3 ~ 0.5 (mm) entre uma fatia e outra. As máscaras *ground truth* foram geradas para 50 tomografias nos primeiros quatro casos e para 120 tomografias no último caso por um especialista na área.

O treinamento da rede foi realizado em 220 épocas com lotes de quatro imagens e a rede foi otimizada com a utilização do algoritmo *Root Mean Square Propagation* (RMS Propagation). A taxa de treinamento utilizada foi com ajuste dinâmico, sendo que as primeiras 40 épocas foram realizadas com um valor de 0.001 e as subsequentes com 0.0001.

O artigo em questão faz um comparativo entre a utilização de redes *deep learning* U-Net e DenseNet na segmentação das vértebras e mostra que o método proposto, utilizando um número consideravelmente menor de parâmetros, atinge um resultado para o coeficiente Dice muito melhor para todos os cinco casos médicos analisados. Foram propostos como futuros

trabalhos o uso de redes *deep learning* CNN 3D e ampliação do número de casos como forma de aprimorar os resultados já obtidos.

3.4.3 Reconstrução da Espinha e Segmentação Tridimensional Baseada em *Fully Convolution Network* (FCN) *e Marching Cubes* em Tomografias Volumétricas.

Em processos de tratamento radioterápicos, um passo importante comumente executado é a segmentação da espinha como uma forma de guia para marcar regiões em risco. Nesse artigo, os autores utilizam técnicas de redes *deep learning* combinadas com a técnica de *Marching Cubes* para fazer a segmentação da espinha em volumes 3D. A justificativa para o uso dessa técnica, em detrimento das técnicas tradicionais já utilizadas, é a necessidade de identificação precisa das bordas dos ossos em imagens de baixa resolução, como é o caso das imagens de tomografias computadorizadas.

A base de dados utilizada para o treinamento da rede *deep learning* foi feita a partir de 40 *scans* tomográficos, sendo 25 masculinos e 15 femininos, fornecidos pelo hospital de câncer afiliado à Universidade de Zhengzhou, não disponíveis publicamente. Para a formação do *ground truth*, cada uma das tomografias foi devidamente segmentada utilizando-se técnicas simples com *threshold* e, logo após, transformada em imagens contendo máscaras binárias separando o fundo dos ossos em cada canal.

A arquitetura de rede *deep learning* utilizada assemelha-se muito ao formato *Encoder-Decoder*. A parte do *encoder* é formada por seis blocos de convolução seguida de *pooling* em que, a cada bloco, o número de filtros é dobrado, iniciando com 64 e terminando com 512. Ao final, a última camada de *pooling* é ligada a duas camadas *fully connected* antes de iniciar o processo de *decoder*.

A parte de restauração do espaço multidimensional para uma imagem que compõe a máscara de segmentação foi feita utilizando-se duas técnicas. A primeira considera apenas uma camada de deconvolução enquanto a segunda considera três, seguidas de um *cropping*. Ambas as arquiteturas, referidas como FCN *Original* e *Improved* podem ser observadas na Figura 31.

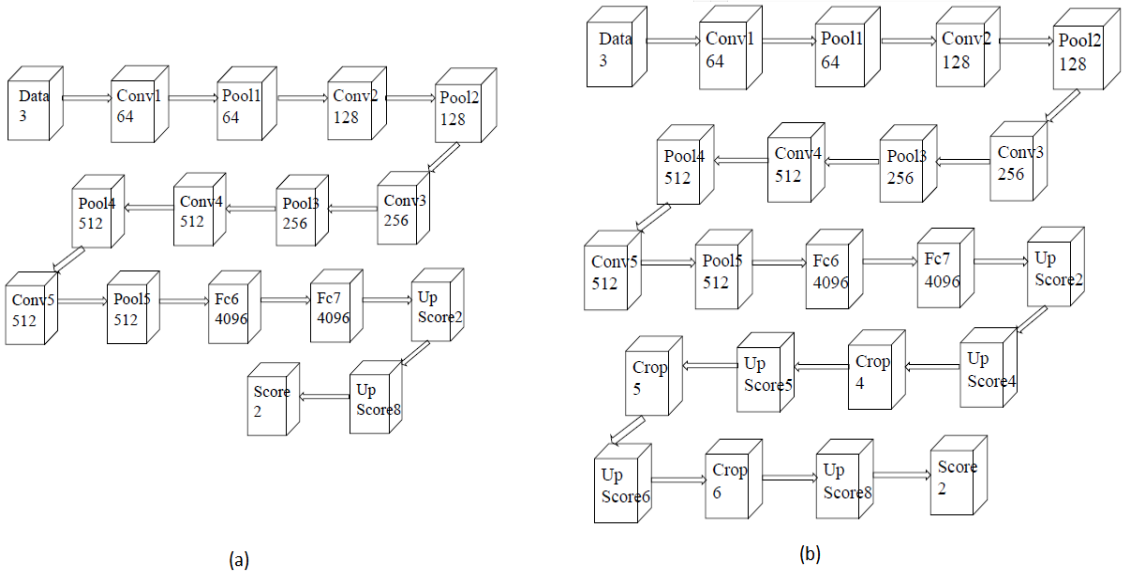


Figura 31 – Topologia FCN *Original* (a) e *Improved* (b) utilizada na segmentação da espinha.

Fonte: [43].

Normalmente as imagens tomográficas tem tamanho 512x512. Porém, para acelerar o processo de treinamento da rede *deep learning*, as imagens foram recortadas e redimensionadas para o tamanho de 256x256. Todo o processo de treino e teste foi realizado utilizando-se o Caffe Framework, com pesos pré-treinados e em mais de 100.000 iterações, sem especificar o tamanho do lote de imagens processadas simultaneamente.

Para medir o desempenho da rede foi utilizada a métrica de acurácia, sendo que a rede *deep learning* mais simples obteve uma taxa de acerto na segmentação de 84% enquanto a rede *deep learning* melhorada, com mais camadas de deconvolução, alcançou um índice de 93% nos testes.

Destaca-se a utilização de três camadas de deconvolução para melhorar o desempenho da parte *decoder* responsável pelo processo de *upsampling* na topologia FCN, reconstruindo o vetor de características resultantes da parte *encoder* na imagem segmentada final sob a justificativa de que as camadas de convolução, durante o *downsampling*, removem muitos detalhes. Outro fator a se colocar em discussão é a utilização da acurácia como métrica de avaliação de desempenho em vez do coeficiente Dice, amplamente utilizado na segmentação de imagens médicas.

3.5 Quadro-resumo

A dissertação embasou-se em 37 artigos com foco em segmentação de órgãos em imageologia médica usando redes *deep learning* para a escolha das topologias a serem comparadas. Um compilado dos principais dados extraídos de cada artigo, tais como: tipo de imageologia médica a ser segmentada, resolução da segmentação, topologia base, topologias propostas, base de dados, métrica de avaliação, além de uma pequena descrição, pode ser consultado no

Quadro 3, tendo o **Quadro 2** o título do artigo relacionado com seu número de referência.

Quadro 2 - Lista dos 37 artigos utilizados como principal base de pesquisa para esta dissertação.

Referência	Título
1	*A Bottom-Up Approach for Pancreas Segmentation Using Cascaded Superpixels and (Deep) Image Patch Labeling*
2	*AHCNet: An Application of Attention Mechanism and Hybrid Connection for Liver Tumor Segmentation in CT Volumes*
3	*An Augmentation Strategy for Medical Image Processing Based on Statistical Shape Model and 3D Thin Plate Spline for Deep Learning*
4	*An Effective CNN Approach for Vertebrae Segmentation from 3D CT Images*
5	*Automatic Detection and Segmentation of Lung Nodule on CT Images*
6	*Automatic Liver Segmentation Using Multi-plane Integrated Fully Convolutional Neural Networks*
7	*Automatic Lung Segmentation on Thoracic CT Scans Using U-Net Convolutional Network*
8	*Automatic Multi-Organ Segmentation on Abdominal CT With Dense V-Networks*
9	*Automatic Segmentation and 3D Reconstruction of Spine Based on FCN and Marching Cubes in CT Volumes*
10	*Automatic Segmentation of Shoulder Joint in MRI Using Patch-Based and Fully Convolutional Networks*
11	*Automatic tissue segmentation by deep learning: from colorectal polyps in colonoscopy to abdominal organs in CT exam*
12	*Automatic Tumor Segmentation Using Machine Learning Classifiers*
13	*Co-Learning Feature Fusion Maps from PET-CT Images of Lung Cancer*
14	*Deep Belief Network Modeling for Automatic Liver Segmentation*
15	*Deep multi-task and task-specific feature learning network for robust shape preserved organ segmentation*
16	*Deep Q Learning Driven CT Pancreas Segmentation With Geometry-Aware U-Net*
17	*Densely connected deep U-Net for abdominal multi-organ segmentation*
18	*Extracting Lungs from CT Images Using Fully Convolutional Networks*
19	*Feature Fusion Encoder Decoder Network for Automatic Liver Lesion Segmentation*
20	*Fully automated esophagus segmentation with a hierarchical deep learning approach*
21	*H-DenseU-Net: Hybrid Densely Connected U-Net for Liver and Tumor Segmentation From CT Volumes*
22	*Learning Deep Spatial Lung Features by 3D Convolutional Neural Network for Early Cancer Detection*
23	*Liver Segmentation in CT Images Using Three Dimensional to Two Dimensional Fully Convolutional Network*
24	*Liver Semantic Segmentation Algorithm Based on Improved Deep Adversarial Networks in Combination of Weighted Loss Function on Abdominal CT Images*
25	*Medical Images Sequence Normalization and Augmentation: Improve Liver Tumor Segmentation from Small Data Set*
26	*MGB-NET: Orbital Bone Segmentation from Head and Neck CT Images Using Multi-Graylevel-Bone Convolutional Networks*
27	*Multiclass Brain Tissue Segmentation in 4D CT Using Convolutional Neural Networks*
28	*Nested Dilation Network (NDN) for Multi-Task Medical Image Segmentation*
29	*Pelvic Organ Segmentation Using Distinctive Curve Guided Fully Convolutional Networks*
30	*PnP-AdaNet: Plug-and-Play Adversarial Domain Adaptation Network at Unpaired Cross-Modality Cardiac Segmentation*
31	*Pulmonary Lobe Segmentation Using A Sequence of Convolutional Neural Networks For Marginal Learning*
32	*Robust Boundary Segmentation in Medical Images Using a Consecutive Deep Encoder-Decoder Network*
33	*Segmentation of Organs at Risk in thoracic CT images using a SharpMask architecture and Conditional Random Fields*
34	*Segmenting the Brain Surface From CT Images With Artifacts Using Locally Oriented Appearance and Dictionary Learning*
35	*Semi-Supervised 3D Abdominal Multi-Organ Segmentation Via Deep Multi-Planar Co-Training*
36	*U-Net Plus: Deep Semantic Segmentation for Esophagus and Esophageal Cancer in Computed Tomography Images*
37	*UNSUPERVISED SEMANTIC SEGMENTATION OF KIDNEYS USING RADIAL TRANSFORM SAMPLING ON LIMITED IMAGES*
38	SEGMENTAÇÃO ÓSSEA DO CORPO HUMANO EM TOMOGRAFIAS COMPUTADORIZADAS USANDO REDES DEEP LEARNING

Referência	Tipo de imagem	Resolução	Modelo-base	Modelo proposto	Base de dados	Métrica	Orgãos	Descrição
1	CT	2D	AlexNet	AlexNet	Pâncreas-CT	Dice, Jaccard	Pâncreas	Segmentação do pâncreas utilizando redes deep lerning em conjunto com algoritmos de Random Forest e SuperPixel.
2	CT	3D	U-Net	U-Net + Attention	LiTS, 3D-IRCADb, 117 Casos clínicos	Dice	Fígado	Segmentação do fígado utilizando redes *deep learning* que combinam mecanismo de atenção frouxos e rígidos juntamente com *skip connections* longas e curtas.
3	CT, MRI	3D	U-Net	U-Net, multi-scale 3D CNN	NCI-ISBI 2013, MICCAI18 MSD	Dice	Próstata, Fígado	Desenvolvimento de técnicas de *data augmentation* em três dimensões para aumentar a performance da segmentação das redes *deep learning*.
4	CT	2D	DenseNet	DenseNet	cinco casos de estudo fornecidos por hospitais	Dice	Espinha dorsal	Segmentação da espinha dorsal para aplicação em processos pré-cirúrgicos.
5	CT	2D	FCN	FCN	LIDC	Dice	Pulmão	Segmentação em vários estágios do pulmão para detecção de nódulos pulmonares.
6	CT	3D	-	MPNet	LiTS	Dice	Fígado	Segmentação do fígado utilizando redes *deep learning* com camadas de convolução dilatada.
7	CT	2D	U-Net	U-Net ConvNet	Própria	Accuracy	Pulmão	Segmentação do pulmão usando redes deep learning para auxílio em diagnóstico de câncer.
8	CT	3D	DenseNet	DenseVNet	Pâncreas-CT, BTCV	Dice	Esôfago, Estômago, Duodeno, Fígado, Baço, Vesícula biliar, Pâncreas, Rim esquerdo	Segmentação de oito órgãos da região abdominal utilizando redes *deep learing*.
9	CT	3D	FCN	FCN + Marching Cubes	40 casos de estudo fornecidos por um hospital	Accuracy	Espinha dorsal	Segmentação da espinha dorsal utilizando redes *deep learning* para processos de radioterapia.
10	MRI	2D	U-Net, AlexNet	AANet (3 U-Net + AlexNet)	oito casos de estudo fornecidos por um hospital	Dice, Positive Predicted Value, Sensitivity	Juntas dos ombros	Segmentação das juntas dos ombros utilizando três combinações de redes U-Net juntamente com uma rede AlexNet.

Quadro 3 – Compilado de informações-chaves extraído dos 37 artigos.

Referência	Tipo de imagem	Resolução	Modelo-base	Modelo proposto	Base de dados	Métrica	Orgãos	Descrição
11	CT	2D	SegNet, Deep Lab	SegNet + DeepLab + LSTM	CVC-ColonDB, CVC-ClinicDB	IoU	Cólon, Pulmão	Segmentação do cólon utilizando redes *deep learning* para detecção de pólipos.
12	CT	2D	-	Gabor features + Random Forest + DNN	3D-IRCADb	Dice, Accuracy, Sensitivity, Specificity, Volume Overlap Error, Relative Absolute Volume Difference, Average Symmetric Surface Distance, Maximum Surface Distance	Figado	Segmentação do fígado e tumores da região abdominal utilizando redes *deep learning* para cura de doenças.
13	PET, CT	2D	-	MC CNN - CT encoder + Pet encoder + Reconstruction	50 casos de estudo fornecidos por um hospital	Dice	Pulmão	Segmentação do pulmão utilizando redes *deep learning* que combinam características de vários tipos de imagem médicas (PET e CT).
14	CT	3D	DBN	DBN-DNN	MICCAI-SLiver07, 3D-IRCADb	Dice, Relative Error	Fígado	Segmentação do fígado utilizando redes *deep learning* com treinamento supervisionado e não supervisionado.
15	CT, MRI	2D	U-Net	U-Net multi-task	3 bases de dados sintéticas	Dice	Rim, Osso femural	Segmentação de órgãos utilizando redes *deep learning* multitarefas na qual uma parte é responsável por detectar bordas e outra regiões.
16	CT	3D	U-Net	Deep Q+ UNet	NIH Pâcreas-CT	Dice	Pâncreas	Segmentação do pâncreas utilizando redes *deep learning* do tipo U-Net com aprendizado não supervisionado para localização mais precisa do órgão.
17	CT	2D	U-Net	Dense U-Net	3D-IRCADb, LiTS	Dice	Figado, Baço	Segmentação de órgãos utilizando redes *deep learning* combinada com uma função custo que auxilia no aumento da precisão dos resultados.
18	CT	2D	FCN	FCN + CRF	HUG-ILD, VESSEL12	Dice	Pulmão	Segmentação dos pulmões utilizando redes *deep learning* do tipo FCN em combinação con CRF *(Conditional Random Fields)*.
19	CT	2D	ResNet	FED-Net	LiTS	Dice	Fígado	Segmentação do fígado utilizando redes *deep learning* do tipo ResNet combinadas com fusão de características baseadas num mecanismo de atenção.
20	CT	3D	SharpMask	SharpMask	30 casos de estudo fornecidos por um hospital	Dice	Esôfago	Segmentação do esôfago utilizando redes *deep learning* to tipo SharpMask em dois estágios.

Referência	Tipo de imagem	Resolução	Modelo-base	Modelo proposto	Base de dados	Métrica	Orgãos	Descrição
21	CT	3D	U-Net	3D U-Dense U-Net	LiTS, 3D-IRCADb	Dice	Fígado	Segmentação do fígado utilizando uma combinação de redes do tipo U-Net densamente conectadas.
22	CT	3D	Encoder-Decoder CNN	Encoder-Decoder CNN	Kaggle Data Science Bowl 2017	Accuracy	Pulmão	Segmentação do pulmão utilizando redes *deep learning* para detecção precoce de câncer.
23	CT	3D	FCN	FCN 3D + CRF	MICCAI 2015	Dice	Fígado	Segmentação do fígado utilizando redes *deep learning* 3D para 2D em conjunto com *Conditional Random Fields* para melhorar detecção de bordas.
24	CT	2D	DeepLab, GAN	DeepLab + GAN	LiTS	Dice, Jaccard, Volume Overlap Error, Relative Volume Difference, Average Symmetric Surface Distance, Maximum Symmetric Surface Distance	Fígado	Segmentação de órgãos utilizando uma combinação de modelos tradicionais de *deep learning* com redes generativas utilizando características de nível profundo com multiescala.
25	CT	3D	FCN	Cascade FCN	3D-IRCADb	Dice	Fígado	Segmentação do fígado utilizando redes *deep learning* com foco em detecção de tumores.
26	CT	2D	U-Net	MGB-Net	83 casos de estudo fornecidos por um hospital	Dice, Accuracy, Sensitivity, Specificity, IoU	Osso orbital	Segmentação do osso orbital do crânio utilizando redes *deep learning* em paralelo do tipo U-Net .
27	CT	3D	U-Net	U-Net 3D	157 casos de estudo fornecidos por um hospital	Dice, Contour Mean Distance, Absolute Volume Difference, Mean Volume Difference	Cérebro	Segmentação do cérebro em pacientes com AVC cerebral utilizando redes *deep learning* 3D.
28	CT, MRI, Imagens endoscópicas	2D	ResNet	Nested Dilation Networks (NDN)	GIANA2018	Dice, Jaccard	Fígado, Pâncreas	Segmentação de órgãos utilizando diversos tipos de imageologia médica usando redes *deep learning* residuais.
29	CT	2D	FCN	FCN + Distinct Curve	313 CTs de 313 pacientes	Dice, Positive Predicted Value, Sensitivity, Average Surface Distance	Próstata, Bexiga e Reto	Segmentação de órgãos da região pélvica em dois estágios utilizando redes *deep learning* do tipo FCN.

Referência	Tipo de imagem	Resolução	Modelo-base	Modelo proposto	Base de dados	Métrica	Órgãos	Descrição
30	CT, MRI	2D	GAN	PnP-AdaNet	MICCAI 2017 Multi-Modality Whole Heart Segmentation (MM-WHS)	Dice, Average Surface Distance	Coração	Uso de redes *deep learning* para adaptação de segmentação de imagens em domínios distintos.
31	CT	3D	SegNet	Seg3DNet	COPDGene clinical trial	Dice	Pulmão	Uso de redes *deep learning* para segmentação do pulmão buscando detectar nódulos pulmonares.
32	CT	2D	DeepLab	CDED-net	CVC-ColonDB, CVC-ClinicDB, ETIS-Larib PolypDB, Pedro Hispano Hospital (PH2), ISBI 2016 skin lesion	Dice, Accuracy, Recall, Precision, IoU	Orgãos abdominais	Segmentação de órgãos da região abdominal utilizando redes *deep learning* multimodais.
33	CT	3D	FCN	CRFasRNN (FCN + SharpMask + CRF)	30 bases de CT scans	Dice	Coração, esôfago, aorta, traqueia	Segmentação de órgãos utilizando redes *deep learning* FCN e SharpMask em conjunto com *Conditional Random Field* (CRF) com foco em tratamento de radioterapia.
34	CT	3D	U-Net	Deep U-Net	Base de dados de pacientes com epilepsia	Dice, Hausdorff Distance, Modified Hausdorff Distance, Mean Absolute Distance	Cérebro	Segmentação do cérebro utilizando redes *deep learning* para monitoramento de implantação de eletrodos e acompanhamento de recuperação.
35	CT	3D	FCN	Deep Multi-Planar Co-Training (DMPCT)	210 casos de estudos de pacientes	Dice	Aorta, Glândula adrenal, Celíaca AA, Cólon, Duodeno, Vesícula Biliar, Veia Cava Interior (IVC), Rim (esquerda, direita), Fígado, Pâncreas, Mesentérico superior Artéria (SMA), Intestino delgado, Baço, Estômago, Veias	Segmentação de multiplos órgãos através de um conjunto de redes *deep learning* em um *framework* de fusão multiplanar.
36	CT	2D	U-Net	U-Net Plus	16 casos de estudo fornecidos por um hospital	Dice, Hausdorff Distance.	Esôfago	Segmentação do esôfago utilizando redes *deep learning* para detecção de câncer.
37	CT	2D	FCN	FCN	20 bases de CT scans	Dice	Rins	Segmentação dos rins através de redes *deep learning* utilizando técnicas de treinamento não supervisionado.
38	**CT**	**2D**	**U-Net, DenseNet, ResNet, DeepLab, FCN**	**U-Net, DenseNet, ResNet, DeepLab, FCN**	**Visible Human Project, 3DIRCADb**	**Dice**	**Crânio, mandíbula, clavícula, escápula, úmero, rádio, ulna, mãos, costelas, esterno, vértebras, sacrum, bacia, fêmur, patela, tíbia, fíbula e pés**	**Comparação entre múltiplas redes *deep learning* verificando se há diferença significativa nos resultados de acordo com a topologia utilizada.**

3.6 Considerações finais.

A revisão da literatura exibida neste capítulo ajudou a selecionar cinco topologias de redes *deep learning* que bem representam a tarefa de segmentação de órgãos em imagens de tomografia computadorizada, a saber: U-Net, FCN, SegNet, DenseNet, ResNet e DeepLab. Observou-se também que o coeficiente Dice é a principal métrica usada para mensurar a eficiência individual de cada rede *deep learning*. Esse coeficiente foi escolhido para comparar o desempenho da segmentação final entre as topologias selecionadas para base de testes do corpo feminino do VHP.

No próximo capítulo serão descritos detalhadamente: a confecção do *ground truth* das bases de tomografias computadorizadas, o procedimento de treinamento e testes das redes *deep learning* e a metodologia de avaliação estatística de desempenho de cada rede *deep*

learning selecionada. Essa avaliação visa responder à questão proposta que é verificar se há diferença significativa no desempenho entre as diferentes topologias utilizadas na segmentação dos ossos do corpo humano e qual topologia se destaca.

4 Metodologia

Este capítulo apresenta a metodologia empregada no desenvolvimento desta dissertação contemplando a seleção de ferramentas, linguagens de programação de software, seleção e criação de bases de dados de treinamento e base de *ground truth*, topologias de redes *deep learning*, assim como método de coleta e armazenamento de resultados.

A **Figura 32** mostra o esquema geral dos experimentos. Cada uma das cinco redes passa por um processo de treinamento sendo, logo após, usadas na segmentação da base de testes, coleta de dados e análise comparativa usando ferramentas estatísticas.

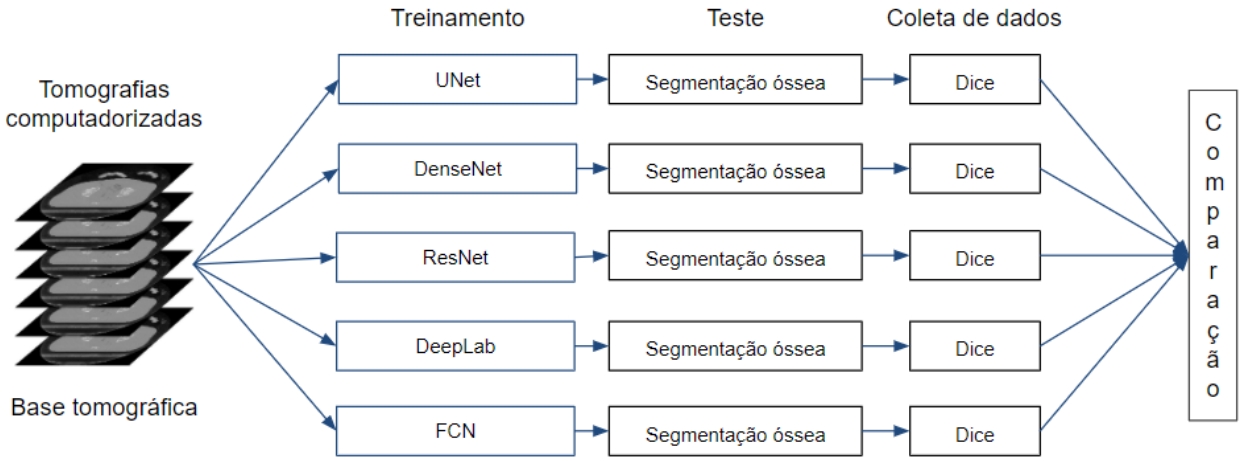


Figura 32 – Esquema básico de avaliação de redes *deep learning*.

Fonte: Compilação do autor.

O detalhamento metodológico seguirá a seguinte organização e ordem:

- Ferramentas de desenvolvimento colocando em evidência os hardwares e softwares utilizados na construção, treinamento, testes de modelos de segmentação de imagens realizadas por meio de redes *deep learning* e análises de seus resultados.
- Bases de dados de treinamento e testes das redes *deep learning*, descrevendo o processo de seleção das bases contendo tomografias de corte axial do corpo humano e construção das máscaras de *ground truth* úteis ao treinamento supervisionado.
- Seleção de topologias de redes *deep learning*.
- Metodologia de treinamento de redes e coleta de dados de experimentos.

- Metodologia de avaliação estatística dos resultados.

4.1 Ferramentas de desenvolvimento

Para o desenvolvimento deste projeto, dois itens foram considerados fundamentais: o primeiro foi a escolha de uma linguagem de programação fácil de usar e que fornecesse bibliotecas de manipulação de imagem e redes neurais; o segundo foi a escolha de um computador com alto poder de computação necessário para o treinamento de redes *deep learning*. Deve-se destacar, nesse caso, que o poder de computação necessário pode advir do hardware de aceleração gráfica conectado a um computador.

Visando atender ao requisito do primeiro item, destaca-se aqui a adoção da linguagem de programação Python, que fornece um ambiente de programação o qual permite coabitar vários paradigmas – orientado a objetos, funcional e imperativo. Além disso, a linguagem é de alto nível, interpretada dinâmicamente e dispõe de uma vasta gama de bibliotecas de desenvolvimento. As principais bibliotecas usadas foram: OpenCV e Matplotlib para o pré-processamento das imagens e Keras para a prototipagem, treinamento e teste de redes *deep learning*.

O dispositivo de hardware usado nos testes preliminares foi um notebook HP ENVY TS™ com processador Intel™ Core i7-4700MQ com 2.40GHz, 16GB de RAM, placa de vídeo NVIDIA™ GeForce 840M com 2GB de RAM, 1TB de armazenamento. O sistema operacional usado foi Windows 10 Home.

Para atender ao segundo requisito, foi utilizado o ambiente de desenvolvimento em nuvem *Colaboratory*, disponibilizado pela empresa Google. Esse ambiente provê um espaço onde é possível executar *programas* escritos na linguagem Python sobre uma máquina virtual que disponibiliza uma GPU com memória variável de acordo com a demanda do servidor central. Esses valores de memória variaram de 12GB a 16GB, o que é suficiente para processar um lote (*batch*) contendo imagens de tamanho 512 x 512 *pixels* concatenados simultaneamente por uma rede *deep learning* num espaço razoável de tempo, i.e., de um a dois dias para completar um treinamento. Em alguns momentos, foi possível contar com a ajuda de uma máquina virtual com processador Intel Xeon™ de 2.10GHz, 16GB de RAM, placa de vídeo NVIDIA Tesla™ com de 8GB de RAM para conduzir treinamentos em paralelo de forma a reduzir o tempo dos experimentos.

Ao final de cada etapa de treinamento e teste de cada rede *deep learning* tanto o armazenamento dos parâmetros bem como os relatórios contendo os dados estatísticos resultantes da segmentação completa de imagens do corpo humano feminino do VHP foram coletados e armazenados no ambiente em nuvem.

Para a criação das máscaras foi utilizado o software Photoshop. Essa ferramenta mostrou-se essencial nessa tarefa devido à vasta quantidade de técnicas de seleção e manipulação de imagem que ela traz. A segmentação manual de todas as classes de ossos foi feita com a ajuda de um atlas médico, sendo que, para alguns itens que não ficaram claros, um especialista foi consultado.

Na etapa de análise estatística para validar ou não as hipóteses formuladas, foi utilizado o software SPSS™ da IBM. Esse software contém uma ampla gama de testes paramétricos e não paramétricos de avaliação de médias, além de trazer análises detalhadas dos resultados.

4.2 Bases de dados

Um dos maiores desafios para treinar redes *deep learning* em segmentação de imagens médicas é encontrar bases de dados devidamente rotuladas, pois esse processo requer um especialista treinado e com muito tempo disponível.

De modo a produzir resultados relevantes e desafiadores, foram pesquisadas bases de imagens tomográficas públicas contendo um número grande de amostras e *ground truth* devidamente segmentado para vários órgãos. Como esses dois requisitos não foram completamente satisfatórios, necessitou-se ajustar as bases selecionadas. Esse ajuste consistiu na realização de uma etapa de pré-processamento sobre as imagens de tomografias, seguida de uma etapa de confecção das máscaras de *ground truth* para cada classe de osso, conforme detalhado nas próximas seções.

4.2.1 Seleção das bases tomográficas

Foram selecionadas duas bases de imagens tomográficas distintas para esta dissertação: a primeira disponibilizada pela National Library of Medicine, referentes ao projeto *Visible Human Project* (VHP), e a segunda, pelo Instituto de Pesquisa Contra o Câncer Digestivo (IRCAD) [46] localizado na França.

A base de imagens do Projeto VHP contém dois conjuntos tomográficos: um do corpo humano masculino com 1.875 imagens de tamanho 512x512 *pixels*, e o segundo de um corpo

humano feminino com 1.730 imagens também de 512x512 *pixels*. Ambos os conjuntos possuem resolução de 16 bits, sendo que cada pixel contém informações de cor em escala de cinza armazenadas em 10 bits.

A seleção dessa base deu-se pela sua exclusividade, i.e., ela possui tomografias de corpo completo. Isso nos permite explorar os desafios na segmentação de todos os diferentes tipos de ossos, desafios estes que, em outros trabalhos, não foram colocados em prática [14, 24, 43]; tais trabalhos focaram em um tipo de osso ou uma única área do corpo como, por exemplo, a região abdominal.

A base de dados disponibilizada pela IRCAD, denominada 3D-IRCADb (*3D Image Reconstruction for Comparison of Algorithm Database*), contém 20 estudos clínicos envolvendo lesões e tumores de diversos pacientes anônimos. Esses estudos foram disponibilizados com a finalidade de comparar diferentes algoritmos de segmentação, contendo assim o *ground truth* de vários órgãos da região abdominal.

Os 20 estudos clínicos somam 2.823 tomografias registradas no formato DICOM; esse formato é específico para imagens médicas e o tamanho das imagens é de 512 x 512 *pixels*. Essa base foi selecionada para o projeto visto que ela já contém máscaras *ground truth* dos ossos.

A base tomográfica do corpo humano masculino do VHP somada às bases do IRCAD compõem o conjunto de imagens necessárias para a condução do treinamento supervisionado das topologias de rede *deep learning* selecionadas para o projeto, enquanto que a base feminina do VHP será utilizada para testes e coleta dos dados necessários para o cálculo dos coeficientes de desempenho.

4.2.2 Pré-processamento

A etapa de pré-processamento tem por objetivo converter todas as imagens de diferentes bases de imagens médicas para um formato padronizado e disponibilizá-las para o processo de treinamento e teste. Além do mais, é preciso criar as máscaras para a segmentação de cada classe de osso.

A **Tabela 4** mostra o padrão de imagem tanto para tomografia quanto para a confecção das máscaras *ground truth*.

Tabela 4 – Padrão de imagem para tomografia e *ground truth*.

	Tomografia	***Máscara (Ground Truth)***
Dimensões	512 x 512 x 1	512 x 512 x 1
Valores de pixels aceitos	[0, 255]	0 (fundo) ou 255 (osso)
Tipo	Inteiro de 8 bits sem sinal	Inteiro de 8 bits sem sinal
Formato	PNG	PNG

A primeira etapa desse processo consiste em converter todas as imagens para uma única extensão. Tanto as imagens tomográficas quanto as máscaras *ground truth* da base de dados do IRCAD foram convertidas do formato DICOM para o formato PNG, utilizando a biblioteca Python *pydicom*.

Padronizando os formatos, o passo seguinte foi equalizar todas as imagens numa escala de cor monocromática de 8 *bits* com tons de cinza de 0 a 255. As imagens de tomografias do *VHP* foram todas disponibilizadas em 16 *bits*; porém, com informações de apenas 10 *bits* para cada *pixel*. A equalização para 8 *bits* juntamente com uma correção de gamma foi feita utilizando a Equação 12 para cada *pixel p* da imagem, onde *p* é o valor do *pixel* em escala monocromática de 0 a 255, 1.023 é o limite da escala de 8 bits (0 a 1.023), 65.535 é o limite de escala para 16 bits (0 a 65.535) e 0.6 é um fator de ajuste de intensidade escolhido arbitrariamente. Essa equação baseia-se na fórmula de correção de gamma utilizando a lei da potência aplicada em processamento de imagens [64].

$$p = \left(\frac{p * 1.023}{65.535}\right)^{0.6} \qquad \text{(Equação 12)}$$

As tomografias do IRCAD foram ajustadas para o formato DICOM com 16 *bits* sinalizados. Cada *pixel* registrado nesse padrão encontra-se em uma escala chamada *Hounsfield Units* (HU), que corresponde a uma medida de radiodensidade normalizada entre água destilada e ar; essa medida foi criada por Godfrey Hounsfield. O formato DICOM armazena, além dos valores de cada *pixel* em HU, os dados referentes tanto ao estudo clínico como ao equipamento utilizado na captação da tomografia. Essas informações adicionais permitiram a conversão de HU para uma escala de intensidade, utilizando os dados de *slope* e *intercep* contidos internamente no próprios arquivo DICOM, aplicando a seguinte equação linear (Equação 13):

$$HU = IN * slope + intercept \quad \text{(Equação 13)}$$

Onde *pixel value* é o valor do *pixel* captado pelo equipamento tomográfico, o *slope* é o coeficiente angular e *interpect* é o coeficiente linear de uma reta.

Além da aplicação da Equação 13, com valores para HU próximos a -1000, que correspondem a ar, e valores muito acima de 1000 a materiais estranhos como metais e outros objetos, todos os *pixels* foram limitados entre a faixa de -1024 a 1024 antes de serem normalizados para valores em 8 *bits* de 0 a 255.

4.2.3 Criação das máscaras

Essa foi uma das tarefas mais demoradas, repetitivas e de fundamental importância para o treinamento das redes *deep learning* em modo supervisionado uma vez que o aprendizado das redes *deep learning* se faz comparando o resultado de saída com a máscara *ground truth* esperada ajustando-se o valor dos pesos a cada iteração.

Como já mencionado, foi utilizado o software Photoshop™ para a confecção das máscaras pois dispõe de diversas ferramentas de seleção e camadas de aplicação de filtros. Aqui, deve-se notar que, quase na totalidade dos casos de confecção de máscara, foi utilizada a aplicação de uma camada de *threshold*, a qual cria uma imagem binária a partir de um valor de cor definido. Como os ossos têm maior claridade quando comparados a outras regiões, essa camada ajuda na delimitação dessas partes com maior facilidade.

A ferramenta de seleção *Quick Selection* foi fundamental para seleções difíceis, i.e., quando a qualidade das imagens era baixa e a seleção de regiões nas quais os pixels pertencentes a duas ou mais classes praticamente se tocavam, como no caso das costelas e esterno.

O procedimento padrão adotado na confecção de cada máscara **(**Figura 33**)** em aplicar, inicialmente, uma camada de *threshold* para eliminar a maior parte do fundo e outros órgãos, seguido do uso de ferramentas de seleção para segmentar cada classe de osso, selecioná-la e registrá-la em seu próprio arquivo como uma imagem binária.

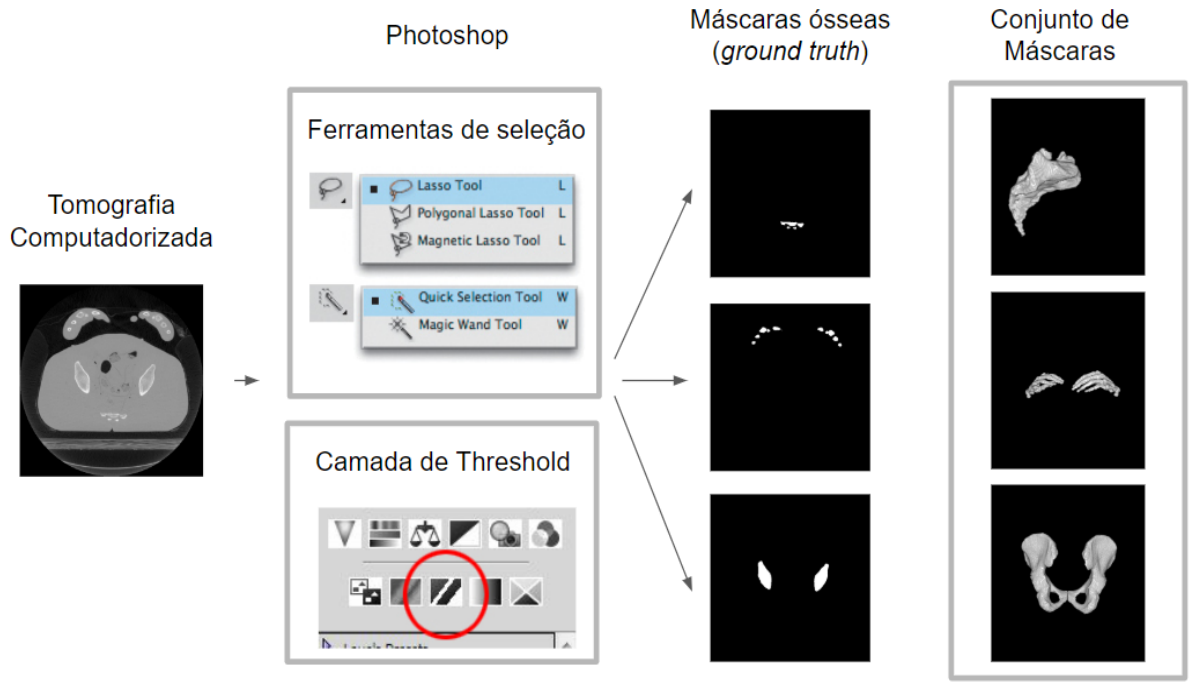


Figura 33 – Etapas da criação de *ground truth* para bases de imagens tomográficas.

Fonte: Compilação do autor

Esse procedimento foi aplicado para todas as 6.418 imagens tomográficas. Apenas para a base de imagens do VHP, as 3.595 tomografias geraram **8.482** máscaras ósseas enquanto as 20 bases de imagens do IRCAD geraram **5.534** máscaras. A quantidade exata de máscaras para cada classe pode ser vista na **Tabela 5**.

Tabela 5 – Quantidade de máscaras de *ground truth* por classe de osso nas bases dos corpos humanos masculino e feminino do *Visible Human Project* juntamente com as 20 bases do IRCAD.

Classes	***VHP masculino***	***VHP feminino***	***IRCAD***
clavícula	94	86	17
crânio	174	160	0
pés	152	162	0
fêmur	486	423	42
fíbula	401	345	0
mãos	143	167	0
bacia	219	213	255
úmero	240	280	12
mandíbula	106	82	0
patela	52	52	0
rádio	129	216	0
costelas	390	366	1.929
sacro	157	127	65
escápula	176	164	100
esterno	210	188	407
tíbia	410	351	0
ulna	146	198	0
vértebras	619	598	2.707
Total	**4.304**	**4.178**	**5.534**

O tempo para a confecção de apenas uma máscara variou de um a dez minutos, dependendo do caso. A região abdominal foi a mais difícil devido à quantidade de ruído nas imagens, além da interseccionalidade entre os ossos das vértebras e costelas e entre os ossos da costela e do esterno.

O processo de criação do *ground truth* tanto das bases tomográficas do VHP quanto das bases do IRCAD levou aproximadamente oito meses para ser concluído. Todo o processo de segmentação manual foi realizado somente pelo próprio autor desta dissertação recorrendo-se a um especialista em tomografias computadorizadas do CETAC (Centro de Tomografia Computadorizada LTDA) de Curitiba e a um átlas médico [1] em caso de dúvidas.

4.3 Seleção das arquiteturas das redes *deep learning*

Baseado em uma revisão literária apresentada no Capítulo 3, juntamente com testes preliminares, foram selecionadas as seguintes topologias de redes *deep learning* para compor este projeto: U-Net, DenseNet, ResNet, DeepLab e FCN.

Como uma rede pode ser montada de inúmeras maneiras e conter uma infinidade de variações, escolheram-se arquiteturas já utilizadas em artigos relacionados à segmentação de imagens. Algumas adaptações foram feitas para adequar as topologias escolhidas à proposta deste trabalho como, por exemplo, aceitar um tensor de tamanho 512x512x1 (altura, largura e profundidade) referente a uma tomografia computadorizada monocromática na entrada e produzir na saída um tensor de tamanho 512x512x19 (altura, largura e número de classes, 18 ossos e o fundo) referente à segmentação resultante.

A topologia U-Net escolhida foi baseada em [26], sendo inseridas as camadas de *Dropout* [59] e de *Max Pooling* e *Batch Normalization* [60] entre cada camada convolutiva, seguida pela camada de ativação ReLU para atenuar problemas de *overfitting* cf. a Figura 34. A rede U-Net é a topologia mais utilizada em segmentações médicas, sendo que não poderia ficar de fora da seleção neste trabalho.

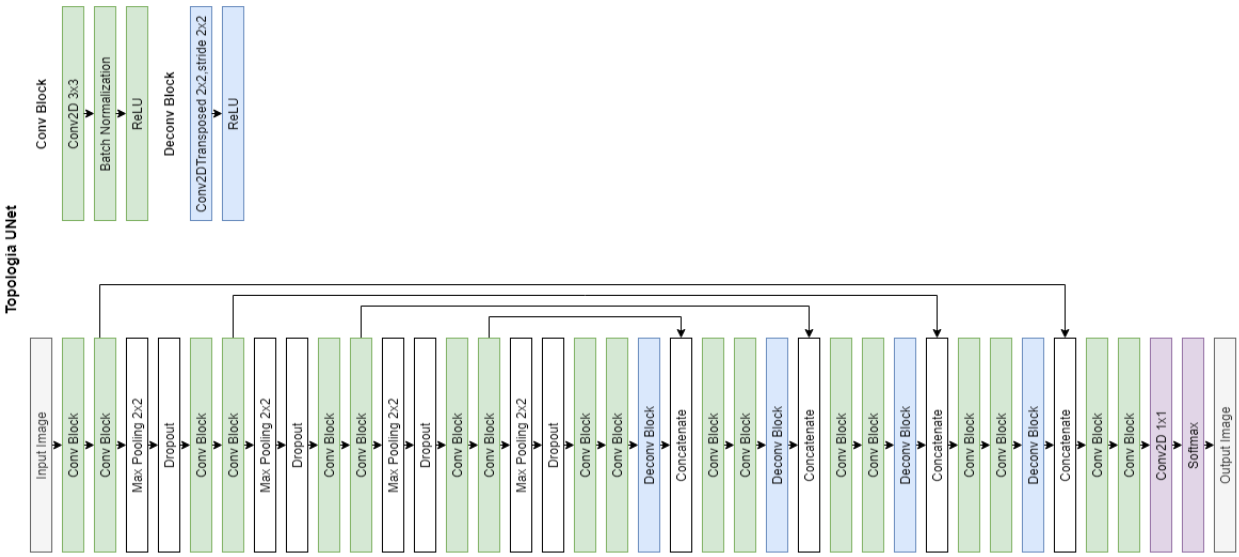


Figura 34 – Topologia U-Net.

Fonte: Compilação do autor

Já a topologia DenseNet foi selecionada a partir do modelo FC-DenseNet103 (Figura 35), usada no trabalho [61] para lidar com o problema de segmentação semântica.

Selecionou-se essa topologia por esta aperfeiçoar a técnica de como combinar características entre diferentes camadas em relação à topologia U-Net.

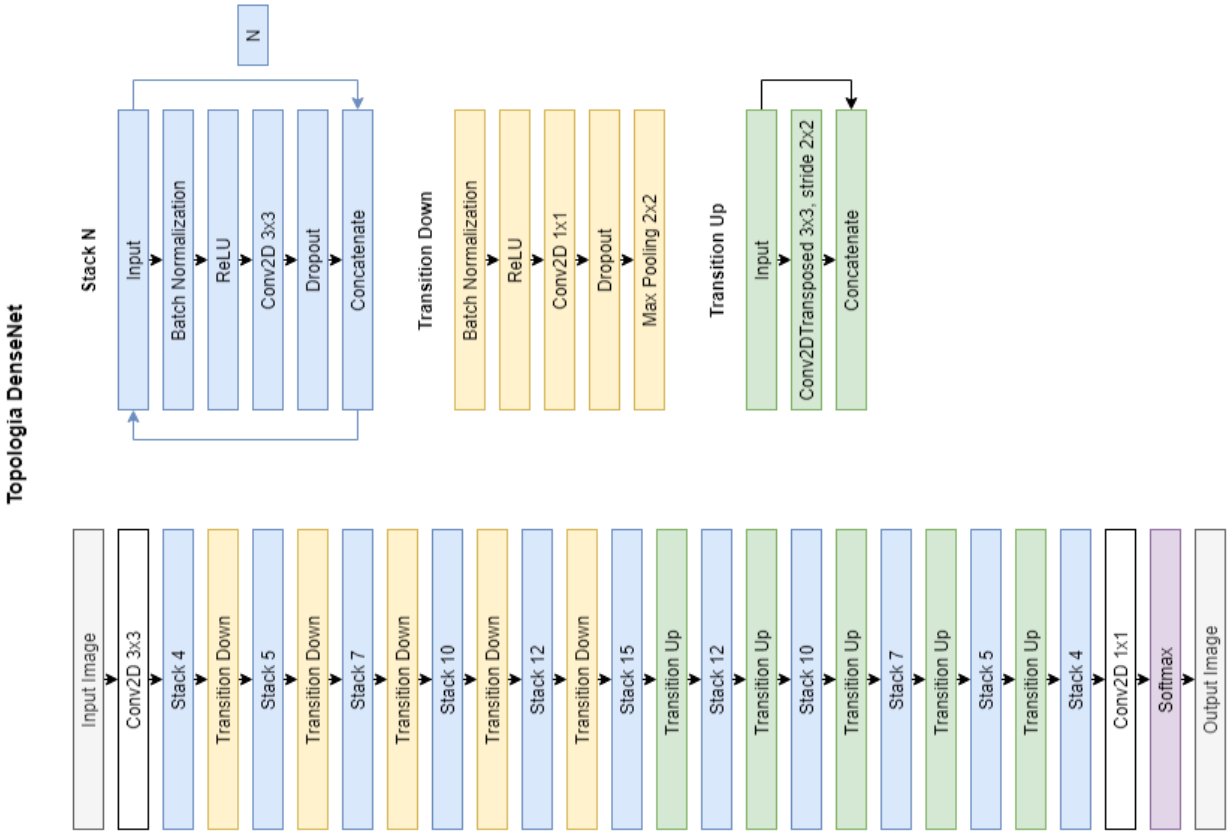


Figura 35 – Topologia DenseNet baseada na arquitetura FC-DenseNet103.

Fonte: Compilação do autor

A arquitetura ResNet foi proposta para a comparação de várias arquiteturas diferentes [62] na tarefa de rastreamento de pessoas usando vista aérea em ambientes lotados, sendo sua estrutura apresentada cf. a **Figura 36**. Ela foi selecionada por possuir uma capacidade de extensão indefinida em termos de concatenação de camadas de aprendizado uma vez que sua técnica de *skip connection* impede que os gradientes zerem durante a etapa de *backpropagation*.

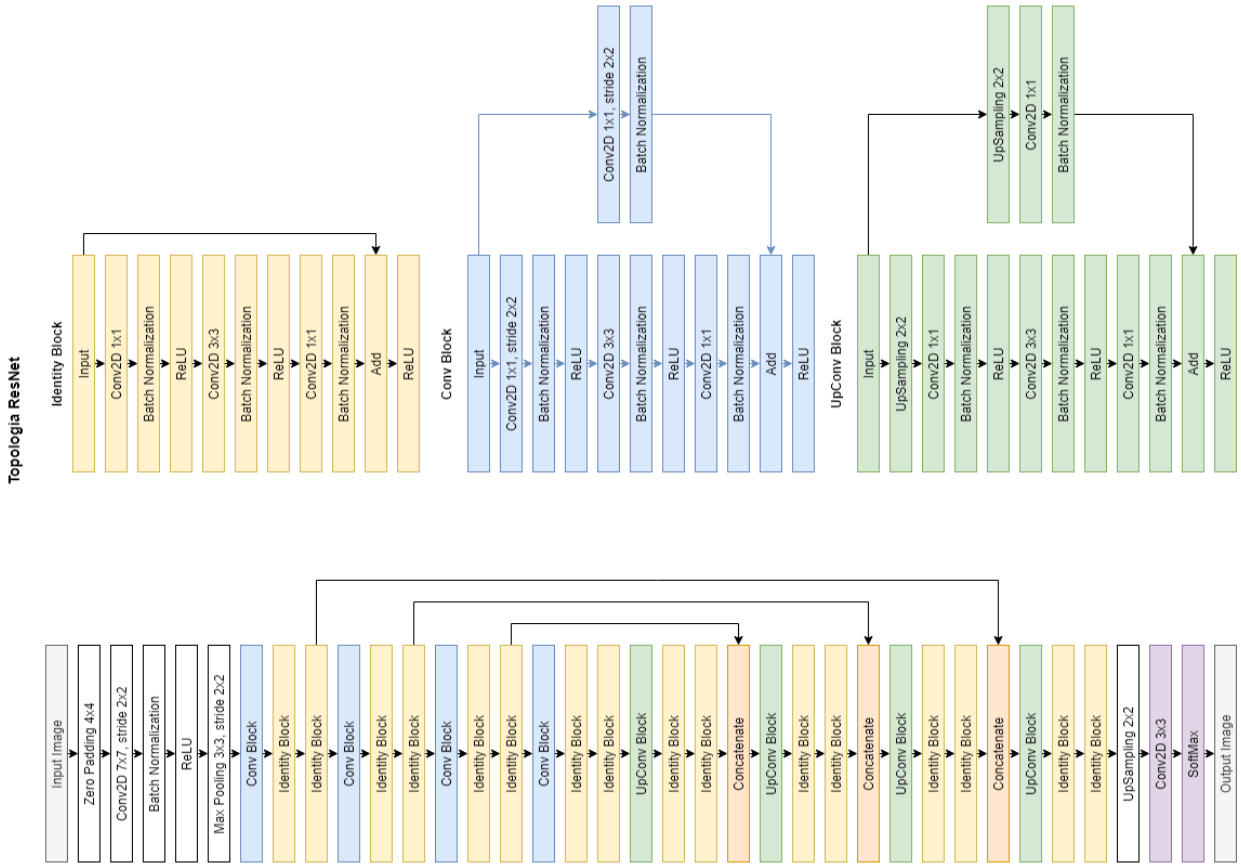


Figura 36 – Topologia ResNet.

Fonte: Compilação do autor

O modelo *DeepLab*, estabelecido para ser usado neste projeto, foi o avanço mais recente dessa topologia nomeada *DeepLab* V3 *Plus* [35] contendo, além de camadas de *Depthwise Separable Convolution* [63], um refinamento na parte *decoder* da arquitetura, cf. a Figura 37, atingindo um desempenho superior na segmentação dos *data sets* PASCAL VOC 2012 e *Cityscapes*. Essa arquitetura foi selecionada pela complexa capacidade de combinar características entre as camadas intermediárias resultando numa segmentação mais precisa.

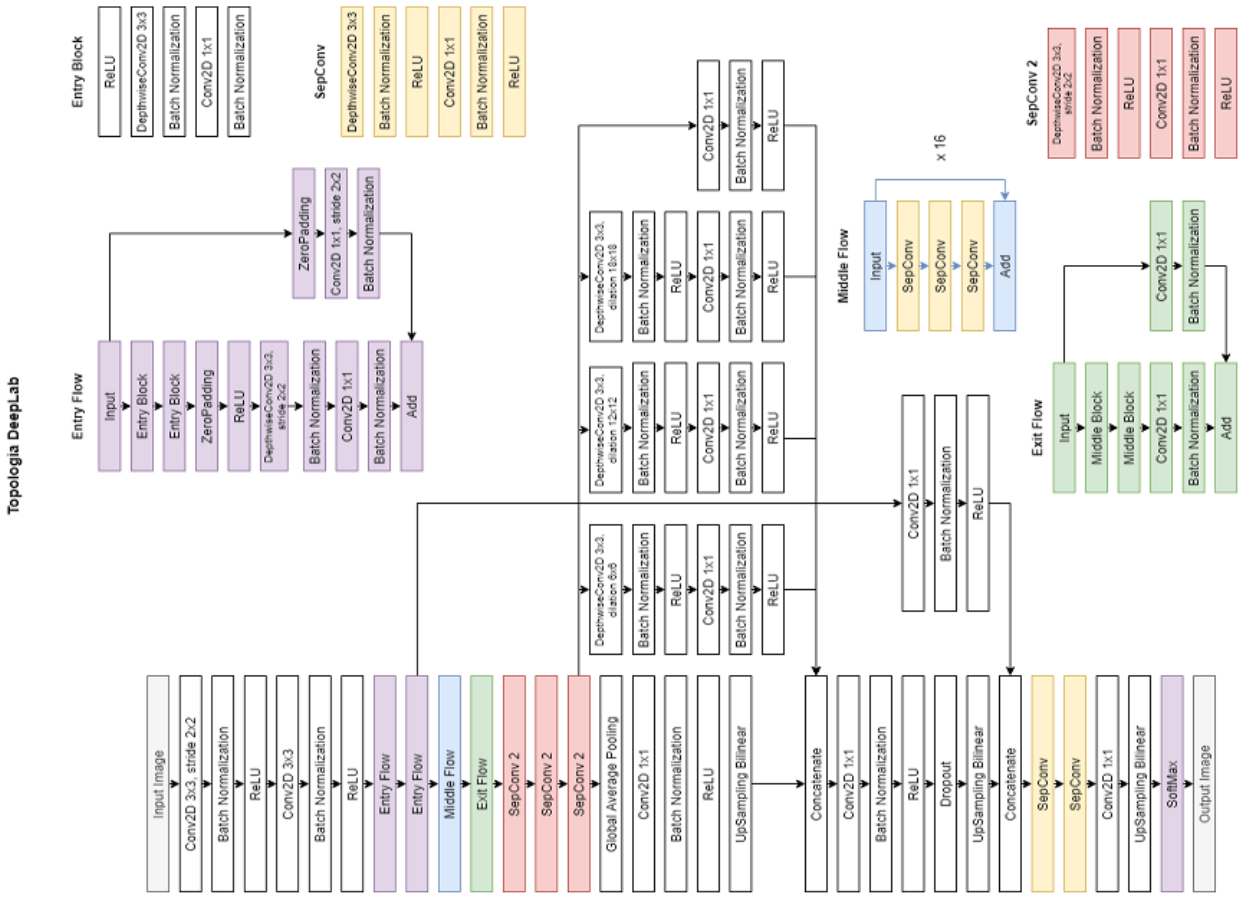


Figura 37 – Topologia DeepLab baseada no modelo DeepLab V3+.

Fonte: Compilação do autor

Por fim, a topologia FCN (*Fully Convolution Network*) proposta em modelo [46], contendo como corpo da parte *encoder* a topologia VGG16 e refinamento da segmentação de saída por meio da concatenação das camadas de deconvolução com as camadas de *max pooling* cf. a **Figura 38**. Foi selecionada por possuir uma arquitetura multirresolução que combina características de diferentes camadas de forma simples e eficiente na etapa de *upsampling* na obtenção da segmentação final.

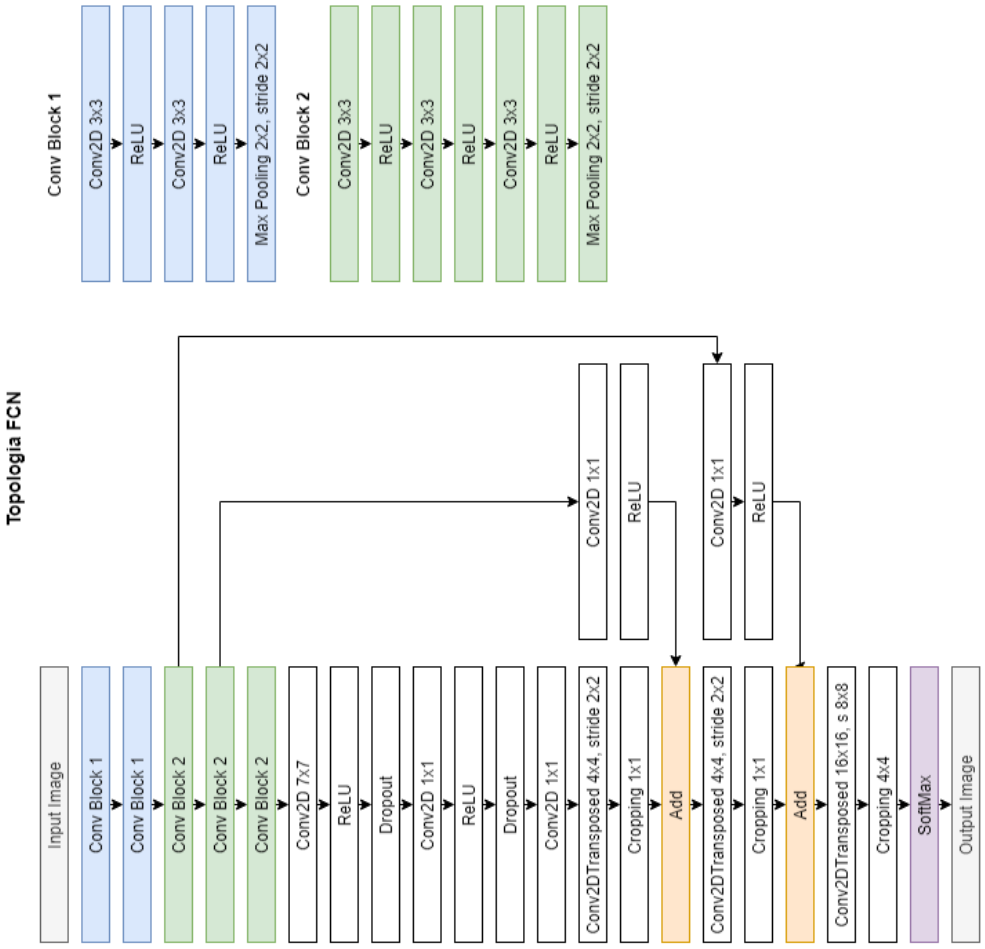


Figura 38 – Topologia FCN.

Fonte: Compilação do autor

A última camada para todas as topologias foi a *Softmax* a qual é responsável por prever a probabilidade de cada *pixel* pertencer a uma determinada classe de osso. Anterior a essa camada, obrigatoriamente, fez-se necessário utilizar uma camada de Convolução ou Convolução Transposta para reduzir o número de filtros de saída ao número de classes a serem segmentadas, sendo para o nosso projeto 19 classes, i.e., 18 ossos e fundo.

É importante frisar que os resultados e conclusões finais deste projeto são completamente dependentes da forma como foram montadas as arquiteturas. A adição de camadas extras ou até simples mudanças na composição da rede podem alterar significativamente os resultados, logo se deve evitar conclusões de forma genérica a respeito de cada topologia apresentada nesta seção. A arquitetura ResNet, aqui proposta, pode apresentar desempenho fraco, mas pode ter um desempenho diferenciado em outro experimento colocando-se o dobro de camadas.

4.4 Método de treinamento, testes e coleta de dados

Assumindo que as bases de treinamento foram devidamente montadas e as topologias de redes *deep learning* definidas, as próximas etapas a realizar são o treinamento supervisionado e coleta de dados. Visando evitar resultados enviesados foi utilizado o método de validação cruzada com cinco *folds* para cada topologia de rede *deep learning*, usando o seguinte protocolo em cada treinamento:

- Para cada treinamento: 100 épocas com a rede de pesos gravada a cada 10 épocas.
- Base de treinamento, composta pela base do corpo masculino do VHP (1.865 amostras) somada às 20 bases do IRCAD (2.823 amostras): 80% para treinamento e 20% para validação.
- Algoritmo de otimização de pesos para todos os treinamentos: *Adam (adaptive moment estimation)*.
- Tamanho máximo do lote (*batch*) de imagens de entrada da rede: quatro (devido a limitações da GPU da máquina utilizada).
- Taxa de aprendizado dos pesos (*learning rate*): 0.001.
- *Data augmentation* com 15% de probabilidade de aplicação para cada técnica à imagem de tomografia de entrada:
    - Transformações de forma:
        - Rotação
        - Translação
        - Escalamento
        - Espelhamento horizontal
        - Deformação elástica
    - Transformação de intensidade de cor
        - Saturação
        - Contraste.

- Função custo para cálculo dos gradientes da rede no processo de atualização dos pesos – *Weighted Cross Entropy* (Equação 14):

$$L(l,q) = -1\frac{1}{M}\sum_{x=1}^{M} w_n(x)\left[\sum_{n=1}^{N} ln_{(x)}\log\left(p_n(x)\right)\right] \qquad \text{(Equação 14)}$$

onde *M* representa a quantidade de *pixels* da imagem segmentada, *N* o número de classes, *wn(x)* é o valor do mapa de pesos aplicado ao *pixel x*, *ln* representa o valor do *ground truth* e *pn* a probabilidade prevista pela camada *Softmax* na saída da rede *deep learning*.

Sobre a escolha dos parâmetros propostos no protocolo de treinamento, cabe salientar que foram feitos testes preliminares para chegar aos valores usados neste projeto. A escolha de 100 épocas foi motivada pelo tempo de conclusão de cada treinamento que, em média, leva de um a quatro dias, dependendo da arquitetura da rede em questão e disponibilidade de GPU. Soma-se ainda ao tempo para a conclusão de todos os 25 treinamentos necessários o tempo de processamento da validação cruzada de cada experimento.

A divisão do conjunto de treinamento se deu da seguinte forma: 80% para treinamento e 20% para validação. Essa divisão baseou-se em estudos empíricos feitos por [65]. A função custo *Weighted Cross Entropy* foi a mesma utilizada em La Rosa [14], seguindo o propósito de atenuar a influência da classe "fundo" durante o treinamento juntamente com o algoritmo Adam de aprendizado.

Foram testadas três taxas de aprendizado (0.01, 0.001 e 0.0001) e um esquema de *learning schedule* que reduzia a taxa de 0.01 para 0.001 após 80 épocas. Como deixar um valor fixo em 0.001 e aplicar o *learning schedule* apresentaram praticamente o mesmo efeito, optou-se pela forma mais simples, ou seja, manter apenas uma taxa constante de aprendizado de 0.001.

A chance de 15% de escolha para cada transformação foi obtida dividindo-se 100% pelo número total de transformações, que foram sete, apresentando diferentes imagens para as redes durante cada época para enriquecer o aprendizado.

Todo esse processo compõe um total de 25 treinamentos, sendo os primeiros cinco para a rede U-Net, os seguintes cinco para a rede DenseNet e assim por diante para as redes ResNet, DeepLab e FCN.

Após cada treinamento, segue-se o processo de coleta de dados na base de testes constituída pelo conjunto tomográfico do corpo feminino do VHP, onde a rede treinada executa a seguinte sequência para cada imagem:

- A rede treinada faz a segmentação da tomografia obtendo 1.730 imagens contendo 19 canais sendo 18 para cada classe de osso mais uma para o fundo.
- Com a máscara de *ground truth* para cada uma das 19 classes calcula-se o número de VP (Verdadeiros Positivos), VN (Verdadeiros Negativos), FP (Falsos Positivos) e FN (Falsos Negativos).

- Todos os dados calculados para cada classe são salvos em uma linha de um arquivo ".csv" tendo a primeira coluna com o nome identificador da tomografia, seguida das demais informações para cada classe.

Um exemplo de coleta de resultado pode ser observado na Figura 39 para a tomografia cvf1204f.png. Para a classe "mandíbula", foram encontrados 4.355 Verdadeiros Positivos (VP), 256.441 Verdadeiros Negativos (VN), 1.348 Falsos Positivos (FP) e 0 Falso Negativo (FN), sendo que a soma de todos esses valores deve resultar em 262.144, exatamente o tamanho de 512x512 *pixels* contidos em cada canal da imagem de saída de tamanho 512x512x19.

input_name	TP_background	TP_clavicle	TP_cranium	TP_feet	TP_femur	TP_fibula	TP_hands	TP_hips	TP_humerus	TP_mandible
datasets/vhp/female/inputs/cvf1204f.png	251189	0	0	0	0	0	0	0	0	4355

TP_patella	TP_radius	TP_ribs	TP_sacrum	TP_scapula	TP_sternum	TP_tibia	TP_ulna	TP_vertebras
0	0	0	0	0	0	0	0	2926

TN_background	TN_clavicle	TN_cranium	TN_feet	TN_femur	TN_fibula	TN_hands	TN_hips	TN_humerus	TN_mandible
7281	262144	261210	262144	262144	262144	262144	262144	262144	256441

TN_patella	TN_radius	TN_ribs	TN_sacrum	TN_scapula	TN_sternum	TN_tibia	TN_ulna	TN_vertebras
262144	262130	261810	262118	262144	262144	262144	262139	258205

FP_background	FP_clavicle	FP_cranium	FP_feet	FP_femur	FP_fibula	FP_hands	FP_hips	FP_humerus	FP_mandible
0	0	934	0	0	0	0	0	0	1348

FP_patella	FP_radius	FP_ribs	FP_sacrum	FP_scapula	FP_sternum	FP_tibia	FP_ulna	FP_vertebras
0	14	334	26	0	0	0	5	1013

FN_background	FN_clavicle	FN_cranium	FN_feet	FN_femur	FN_fibula	FN_hands	FN_hips	FN_humerus	FN_mandible
3674	0	0	0	0	0	0	0	0	0

FN_patella	FN_radius	FN_ribs	FN_sacrum	FN_scapula	FN_sternum	FN_tibia	FN_ulna	FN_vertebras
0	0	0	0	0	0	0	0	0

Figura 39 – Exemplos de dados provenientes do resultado da segmentação da tomografia cvf1204f.png contendo VP, VN, FP e FN para cada uma das 19 classes.

Fonte: Compilação do autor

O coeficiente Dice para a mandíbula, na tomografia cvf1204f.png da Figura 39, pode ser facilmente obtido aplicando-se a Equação 11 da seguinte forma:

$$Dice(mandíbula) = \frac{2 * 4.355(VP)}{2 * 4.355(VP) + 1.348(FP) + 0(FN)} = 0.865$$

Esse processo foi repetido cinco vezes para cada uma das cinco topologias de redes *deep learning* selecionadas, totalizando 25 arquivos que serviram como base para o processo seguinte destinado às análises estatísticas dos resultados e conclusões.

4.5 Análise estatística dos resultados

A última etapa do fluxo de trabalho foi executar uma análise sobre os 25 arquivos de dados gerados na etapa anterior procurando verificar se há diferenças estatísticas significativas nos resultados das segmentações para as cinco topologias de rede *deep learning*– U-Net, DenseNet, ResNet, DeepLab e FCN selecionadas neste projeto.

Esse processo iniciou-se com a codificação de um *script* em Python que calcula os coeficientes Dice para cada classe de osso em cada tomografia da base de testes tomando como entrada o arquivo de dados (cf. formato exibido na Figura 39 da seção anterior) resultante de cada treinamento.

Calculado os coeficientes Dice para cada classe, para cada tomografia, para cada um dos cinco arquivos pertencentes a cada rede, o *script*, então, faz a consolidação dos resultados individuais de cada topologia calculando a média entre as mesmas tomografias nos seus arquivos salvando os dados num outro arquivo ".csv".

A Figura 40 demonstra o fluxo de cálculo pegando como exemplo a tomografia cvf1204f.png e a classe clavícula para essa imagem nos arquivos pertencentes à rede U-Net. Cada arquivo ".csv" contém, no total, 1.723 linhas (tomografia cvf1005f.png à tomografia cvf2727f.png), sendo que cada linha contém o total de VP, VN, FP e FN para 19 classes (18 classes de ossos mais a classe "fundo").

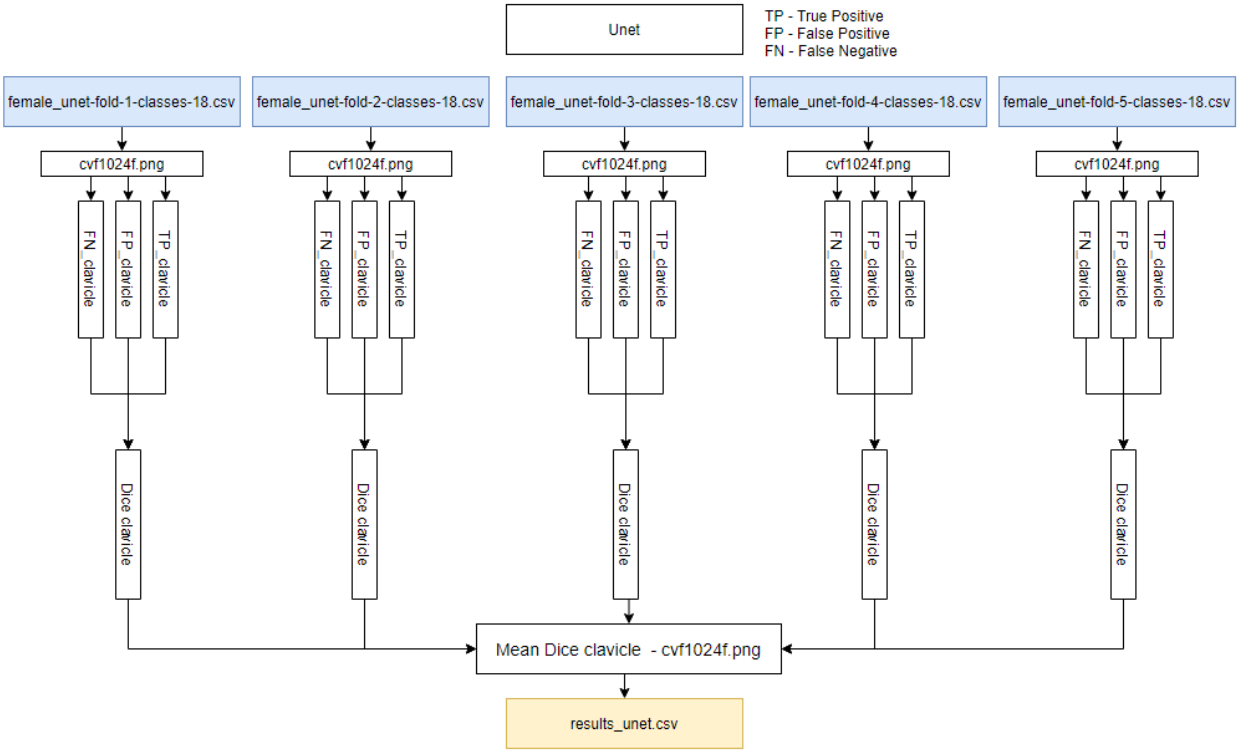


Figura 40 – Fluxo de cálculo da média do coeficiente Dice para a classe clavícula para a rede U-Net para a tomografia cvf1204f.png.

Fonte: Compilação do autor

O cálculo do Dice global para cada é feito através da média entre as médias de todas as tomografias do passo anterior (cf. **Figura 41**) exemplificando para a rede U-Net.

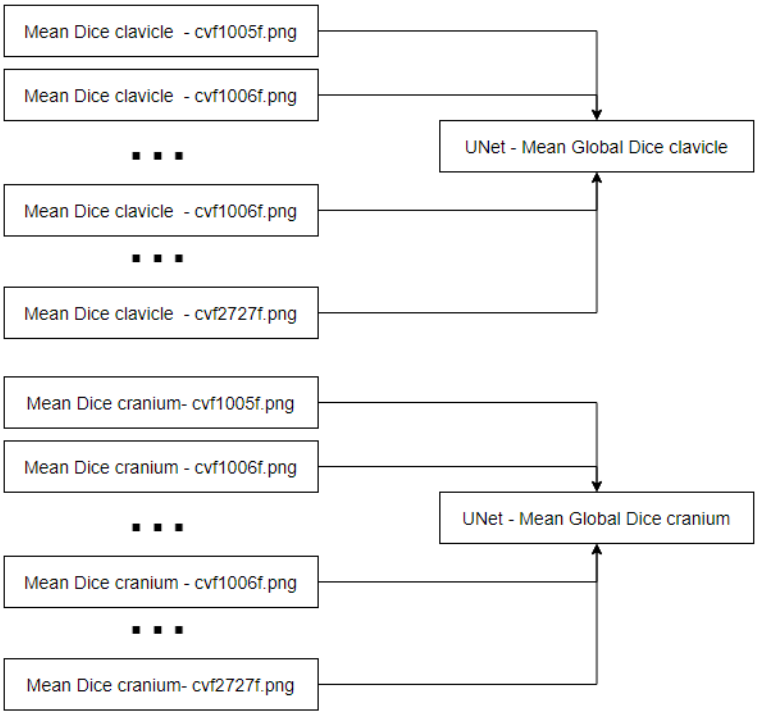


Figura 41 - Cálculo dos Dices globais para a rede U-Net.

Fonte: Compilação do autor

A comparação entre o desempenho das redes *deep learning* foi feita usando-se o *software* de análise estatística SPSS da IBM™, tomando como parâmetros os coeficientes Dice médios de cada tomografia calculados anteriormente. Essa análise deve responder à questão de pesquisa e embasar a conclusão desta dissertação, cujo detalhadamente é dado no próximo capítulo.

4.6 Considerações finais

Este capítulo descreveu os passos usados no desenvolvimento desta dissertação, detalhando a confecção das bases de dados e seleção das arquiteturas de redes *deep learning*, bem como a metodologia de treinamento, testes, coleta de dados e análise estatística. Esta última concerne à análise, sobre os resultados obtidos, empregada para verificar se as segmentações ósseas têm dependência, em termos de significância estatística, quanto à topologia de rede *deep learning* selecionada.

No próximo capítulo são apresentados os resultados obtidos com a aplicação da metodologia aqui proposta, colocando em destaque detalhes do processo de treinamento,

testes, coleta de dados, análises quantitativas e estatísticas dos resultados, bem como o fechamento deste projeto de dissertação.

5 Resultados experimentais

Neste capítulo serão analisados e discutidos os resultados dos experimentos. Tais experimentos avaliarão cinco redes *deep learning*: U-Net, DenseNet, ResNet, DeepLab e FCN aplicadas sobre um conjunto de treinamento de imagens tomográficas com foco na segmentação. Os resultados serão usados para responder à seguinte questão de pesquisa: Qual das topologias avaliadas possui o melhor desempenho em segmentar tomografias computadorizadas de eixo axial em 19 classes de ossos? Deve-se salientar que, das 19 classes, 18 delas são grupos de ossos e uma é fundo/*background*. A motivação para tal questionamento está em verficar se a complexidade das topologias das redes selecionadas e o número de parâmetros de treinamento impactam no resultado final da segmentação de imagens tomográficas.

Para responder a essa indagação, foi utilizada como métrica de avaliação de desempenho individual, para cada rede *deep learning,* o coeficiente Dice. Tal coeficiente mede o grau com que as máscaras de segmentação previstas se sobrepõem às máscaras de *ground truth* criadas. Nessa linha, foram formuladas as seguintes hipóteses:

H_0: Todas as redes *deep learning* possuem coeficientes Dice médios iguais entre si.

H_1: Uma rede *deep learning,* pelo menos, possui um coeficiente Dice médio diferente.

Caso a hipótese H_0 seja verdadeira, pode-se afirmar que todas as redes tiveram o mesmo desempenho na tarefa de segmentação independente da topologia utilizada. Entretanto, caso a hipótese H_0 não seja verdadeira, pelo menos uma das redes possui uma diferença significativa nos valores médios dos resultados. O próximo passo é fazer um estudo estatístico para cada par de redes a fim de descobrir quais são similares, quais possuem diferenças e qual foi a rede que obteve o melhor desempenho.

Na direção de resolver o problema formulado foram criados dois grupos de experimentos. O primeiro grupo envolve treinamento, testes e cálculo do coeficiente Dice individual para cada rede *deep learning*, *e* o segundo abrange testes estatísticos para comparar os resultados obtidos. Tais testes foram realizados em duas partes: a primeira parte avaliou os testes de forma global e a segunda, entre pares.

As próximas subseções estão organizadas da seguinte forma, *vis-à-vis* à apresentação dos resultados:

- Breve descrição da métrica de avaliação Dice e critérios utilizados durante o levantamento dos dados de treinamento e teste.
- Treinamento, testes e levantamento dos coeficientes Dice das redes *deep learning*.
- Testes estatísticos sobre os resultados do item anterior, os resultados propriamente ditos e discussão.
- Considerações finais.

5.1 Métrica de avaliação

Para avaliar o desempenho individual da segmentação de imagens tomográficas de ossos do corpo humano foi utilizado o coeficiente Dice. Esse coeficiente mede o grau de sobreposição entre as máscaras geradas na saída das redes *deep learning* e as máscaras de *ground truth,* dada pela Equação 11.

Como tal métrica de avaliação é baseada em valores binários (v*erdadeiro* ou f*also*), faz-se necessário converter as máscaras expressas em probabilidades e geradas pela camada de saída *Softmax* de todas as redes em máscaras binárias. A conversão foi feita atribuindo-se *verdadeiro* para o *pixel* com maior valor de probabilidade no eixo dos canais (Figura 42), indicando que ele pertence àquela classe de um osso, enquanto enquanto aos demais atribuiu-se *falso*, indicando que eles não pertencem a nenhuma classe.

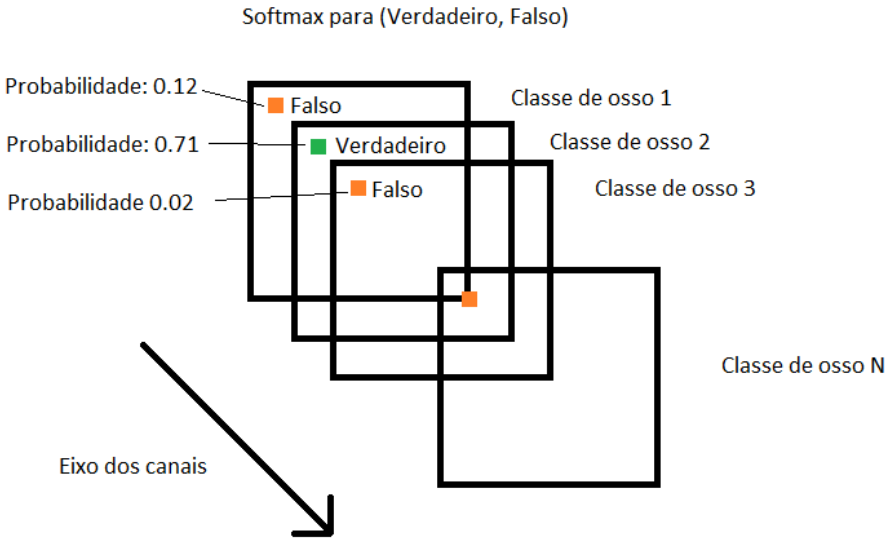


Figura 42 – Conversão das máscaras de saída de *Softmax* para máscaras binárias.

Fonte: Compilação do autor

Existem diversas situações em que a presença da classe avaliada simplesmente não existe numa tomografia como, por exemplo, a classe “pés” ou a classe “fêmur” quando se está avaliando a região da cabeça. Nesses casos, somente valores *verdadeiros negativos* são aceitáveis como corretos para tais classes; porém, o coeficiente Dice não contempla essas situações, gerando uma divisão por zero. Como em tais momentos não se tem certeza se o modelo previu *falso* corretamente, fruto de aprendizado ou de um problema de *overfitting* da classe dominante “fundo”, decidiu-se desconsiderar as medições dos coeficientes para cada classe fora da faixa de imagens indicada na **Tabela 6**. Isso implica que, para a segmentação das 1.730 amostras do corpo feminino do VHP, gerando imagens contendo 19 canais totalizando 32.870 máscaras, apenas 4.178 destas foram consideradas na avaliação.

Tabela 6 – Faixa na com a presença de cada classe na base tomográfica VHP do corpo feminino.

Classe	***Início***	***Fim***	***Qtd.***
clavícula	cvf1234f.png	cvf1319f.png	86
crânio	cvf1005f.png	cvf1164f.png	160
pés	cvf2566f.png	cvf2727f.png	162
fêmur	cvf1824f.png	cvf2246f.png	423
fíbula	cvf2258f.png	cvf2602f.png	345
mãos	cvf1678f.png	cvf1844f.png	167
bacia	cvf1701f.png	cvf1913f.png	213
úmero	cvf1247f.png	cvf1526f.png	280
mandíbula	cvf1131f.png	cvf1212f.png	82
patela	cvf2193f.png	cvf2244f.png	52
rádio	cvf1507f.png	cvf1722f.png	216
costelas	cvf1261f.png	cvf1626f.png	366
sacro	cvf1722f.png	cvf1848f.png	127
escápula	cvf1233f.png	cvf1396f.png	164
esterno	cvf1309f.png	cvf1496f.png	188
tíbia	cvf2236f.png	cvf2586f.png	351
ulna	cvf1517f.png	cvf1714f.png	198
vértebras	cvf1146f.png	cvf1743f.png	598
		Total	**4.178**

Tomando como exemplo a tomografia cvf1447f.png e consultando-se a tabela, verifica-se que o valor 1447 encontra-se dentro das faixas apenas para as classes úmero (1247 a 1526), costelas (1261 a 1626), vértebras (1146 a 1743) e esterno (1309 a 1496), o que implica que os coeficientes Dice, para essa imagem, serão calculados centrados apenas nessas classes cf. **Figura 43.**

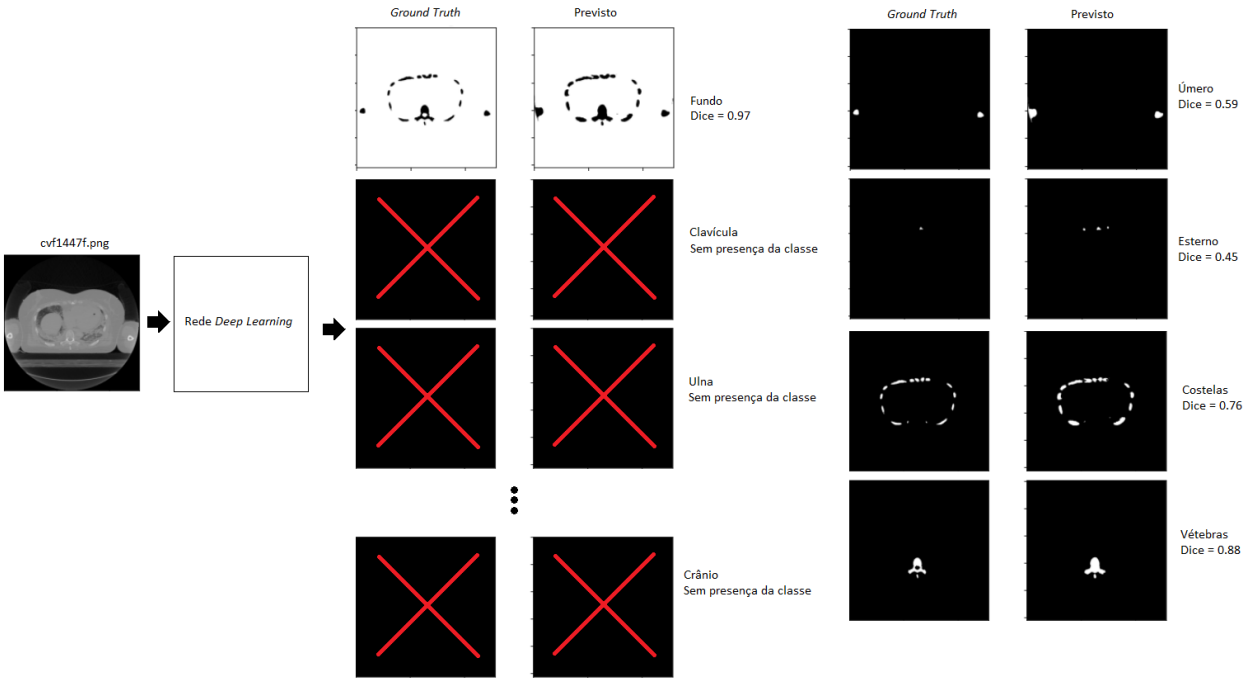


Figura 43 – Cálculo dos coeficientes Dice apenas para as classes contidas na tomografia.

Fonte: Compilação do autor

5.2 Treinamento, testes e levantamento dos coeficientes Dice

Todas as cinco redes *deep learning* foram treinadas utilizando-se da base de imagens masculinas do VHP, *somando-as* às bases tomográficas do IRCAD, totalizando 6.418 amostras, antes da segmentação. Deve-se notar que o cálculo dos coeficientes Dice sobre os resultados da base de testes do corpo feminino do VHP foi feito para 4.178 amostras válidas (**Tabela 6**).

A quantidade de amostras por classe é desbalanceada (cf. **Figura 44**). Tal desequilíbrio é oriundo de fatores como o tamanho dos ossos do corpo humano, que são diferentes entre si, e também da escassez de bases de imagens tomográficas de corpo completo, ou até mesmo de outras regiões diferentes do tórax e abdômen disponíveis para estudos. Esse desbalanceamento favorece a classificação para parte das redes, cujos grupos de classes são majoritários como, por exemplo, a classe “fundo”, na medida em que os modelos são apresentados mais vezes a essas classes durante o processo de treinamento que gera, potencialmente, problemas de *overfitting*.

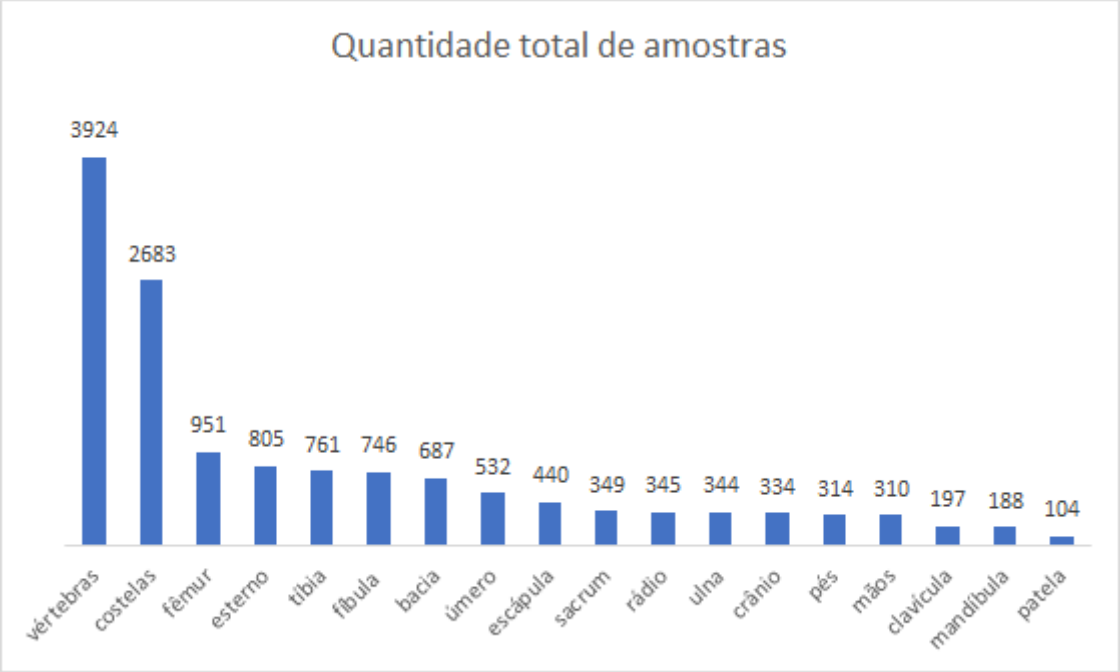


Figura 44 – Quantidade de amostras por classe na base de treinamento dado pelo conjunto de imagens do corpo masculino do VHP somado às 20 bases do IRCAR.

Fonte: Compilação do autor

Em termos metodológicos, cada rede foi treinada e validada em um esquema de validação cruzada conhecido como K-Fold (cf. **Figura 45**) em que se divide a base de dados em dois conjuntos mutuamente exclusivos, sendo um utilizado no treino e outro na validação. O valor de K indica o número de partições que foram geradas sobre a base de treinamento; neste trabalho o valor de K foi 5.

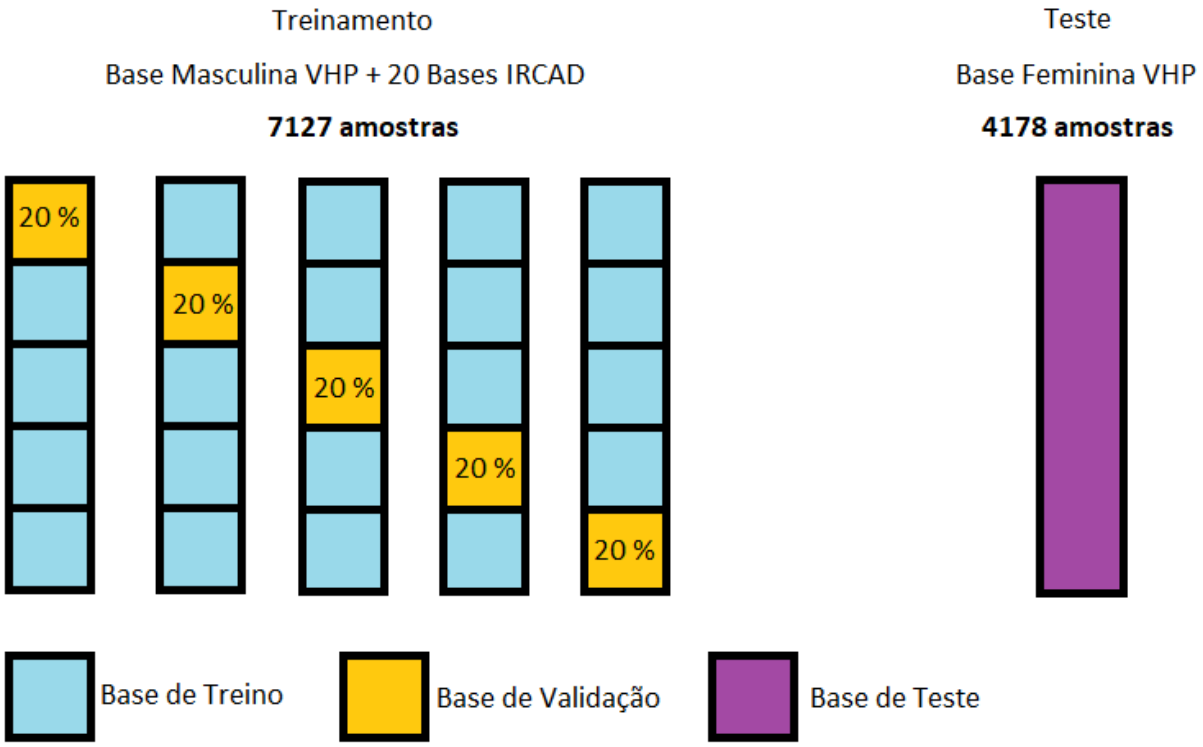


Figura 45 – Esquema de validação cruzada do tipo K-Fold com K igual a 5.

Fonte: Compilação do autor

Deve-se notar que, durante o treinamento, foram aplicadas diversas técnicas de *data augmentation* para melhorar a generalização do aprendizado de cada rede *deep learning*. Uma abordagem técnica aplicada consistiu em realizar diversas transformações como, por exemplo, translação, escalonamento, rotação, giro vertical, deformação elástica e mudança de saturação às imagens de entrada com probabilidade de 15% para cada transformação acontecer.

Outra técnica utilizada para diminuir o problema de *overfitting*, gerado pelo desbalanceamento das amostras, consistiu na aplicação de diferentes pesos na função custo – *Weighted Cross Entropy* para cada classe de osso. Cada peso foi calculado usando-se a Equação 15, onde *freq(c)* é a frequência dos *pixels* de uma determinada classe de osso dividida pelo número total de *pixels* nas imagens em que a classe aparece, e $freq_m$ é a frequência mediana dos *pixels* de todas as classes; o **Quadro 4** mostra os valores calculados em questão.

$$\alpha_c = \frac{freq_m}{freq(c)} \qquad \text{(Equação 15)}$$

Todos os pesos foram calculados antes do treinamento de forma global e com aplicação da função custo durante a etapa de *backpropagation*, em que os pesos da rede em questão são atualizados com os gradientes calculados pela função custo da Equação 14.

Quadro 4 - Pesos atribuídos a cada classe de osso para a função custo Weighted Cross Entropy para diminuir o overfitting dado pelo desbalanceamento das amostras.

Classe	**fundo**	**clavícula**	**crânio**	**pés**	**fêmur**	**fíbula**	**mãos**
Peso	0.00033	3.851657	0.189668	0.681439	0.236127	2.563772	1.884704
Classe		**bacia**	**úmero**	**mandíbula**	**patela**	**rádio**	**costelas**
Peso		0.194719	1	1.740797	6.942333	6.094191	0.09294
Classe		**sacrum**	**escápula**	**esterno**	**tíbia**	**ulna**	**vértebras**
Peso		0.653063	1.145436	2.256164	0.431686	4.130787	0.041536

Os parâmetros globais de treinamento de cada rede estão descritos na **Tabela 7**. Deve-se notar que todas as redes foram treinadas por 100 épocas e com a técnica de otimização Adam para atualizar os parâmetros e taxa de aprendizado de **1e-4.** O número de lotes utilizados variou de rede para rede, dependendo do volume de memória em GPU disponibilizado no momento e da complexidade da rede, cujo mínimo foi de dois lotes e máximo de quatro lotes; registramos que o máximo permitido nas redes DenseNet e FCN não ultrapassou dois lotes.

Tabela 7 – Parâmetros de treinamento usados com cada rede *deep learning*.

Parâmetros	***Valor***
Otimizador	Adam
Taxa de aprendizado	0.0001
Épocas	100
Tamanho de lote	2 ou 4
Função custo	*Weighted Cross Entropy*

Para cada tarefa de treinamento concluída, realizou-se a coleta de dados prevendo-se, com o modelo treinado, a máscara de saída para cada tomografia do conjunto feminino do VHP e calculando-se o número de VP (Verdadeiro Positivo), VN (Verdadeiro Negativo), FP

(Falso Positivo) e FN (Falso Negativo) para cada classe de osso em relação à sua máscara *ground truth*. Sobre esses valores aplicou-se, então, a Equação 11 para se conseguir o coeficiente Dice.

Obtendo-se as medições dos coeficientes Dice, descartaram-se aquelas nas tomografias fora da faixa da Figura *7 - Representação da operação realizada em um nó da rede.*

para cada classe e calculou-se a média global entre os cinco treinamentos, obtendo-se os valores exibidos na **Tabela 8**.

Tabela 8 – Média global dos coeficientes Dice para cada classe. A cor verde destaca os coeficientes que superaram o valor de 0.700.

Dice	U-Net	DenseNet	ResNet	DeepLab	FCN	Média
clavícula	0.7130	0.6744	0.4494	0.6468	0.5280	0.6023
crânio	0.8092	0.8170	0.7296	0.8239	0.6779	0.7715
pés	0.6475	0.3444	0.3062	0.3403	0.5878	0.4452
fêmur	0.8761	0.7748	0.6718	0.8298	0.7435	0.7792
fíbula	0.7368	0.5357	0.4482	0.6728	0.5547	0.5896
mãos	0.5411	0.4810	0.2125	0.2404	0.2676	0.3485
bacia	0.7940	0.7433	0.6638	0.7609	0.7150	0.7354
úmero	0.6453	0.6575	0.4426	0.7285	0.5912	0.6130
mandíbula	0.7991	0.7559	0.6845	0.7961	0.6695	0.7410
patela	0.6951	0.7623	0.1798	0.2106	0.4704	0.4636
rádio	0.3620	0.3700	0.3149	0.3926	0.3058	0.3491
costelas	0.5687	0.5701	0.4128	0.4984	0.4304	0.4961
sacro	0.4957	0.4069	0.3294	0.3889	0.4931	0.4228
escápula	0.6489	0.5623	0.4200	0.5637	0.4774	0.5345
esterno	0.5355	0.5394	0.3072	0.5152	0.3405	0.4476
tíbia	0.8051	0.7683	0.5648	0.7517	0.7195	0.7219
ulna	0.5122	0.5663	0.2978	0.5708	0.4133	0.4721
vértebras	0.7710	0.6746	0.5727	0.7166	0.6563	0.6782
Global	**0.6642**	**0.6113**	**0.4449**	**0.5804**	**0.5357**	**0.5673**

Nas imagens da **Figura 46** e **Figura 47**, pode-se observar que a segmentação que atingiu, de um lado, o maior coeficiente de todas as classes de ossos foi o fêmur, com 0.7729 e, de outro lado, o pior coeficiente de todas as classes de ossos foi a mão, com 0.3485. Essa constatação pode ser observada nas imagens tridimensionais geradas por cada rede.

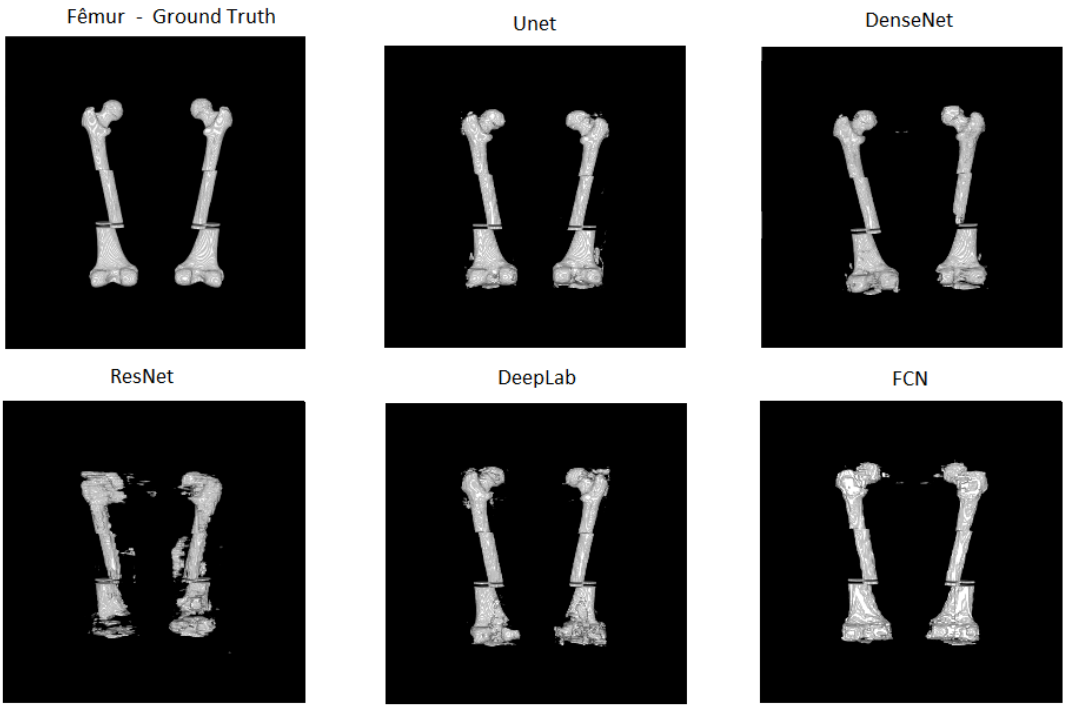


Figura 46 - Segmentação automática dos ossos do fêmur.

Fonte: Compilação do autor

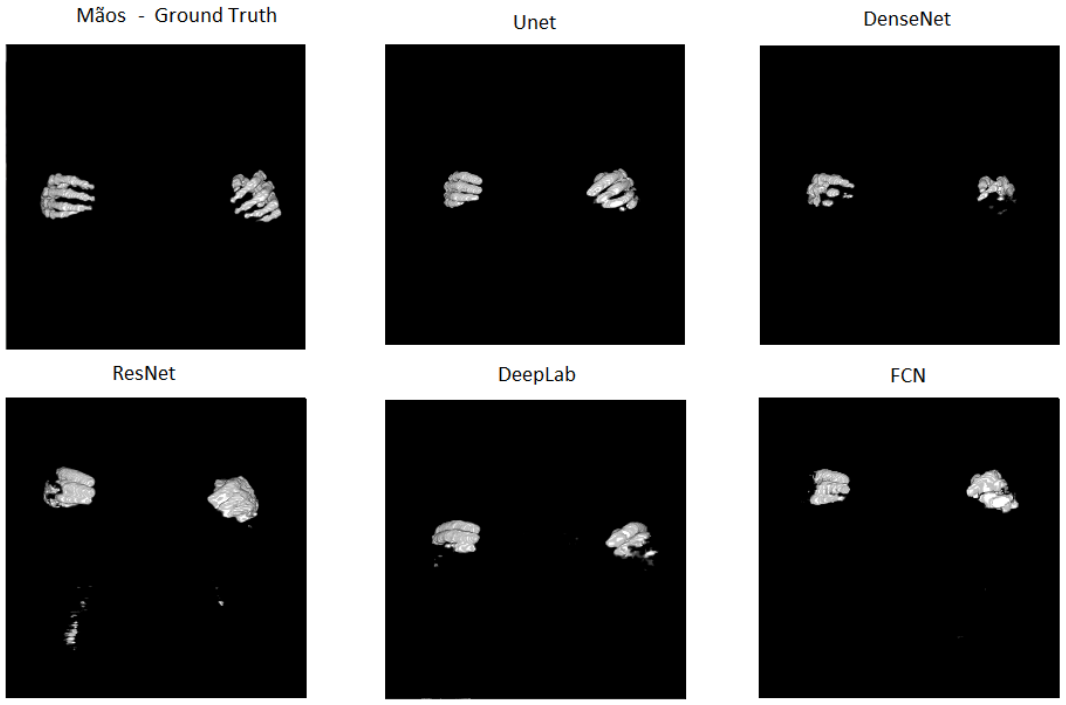


Figura 47 – Segmentação automática dos ossos das mãos. É possível observar que praticamente não houve qualquer segmentação dos ossos por parte das redes.

Fonte: Compilação do autor

A maior média global de todas as redes *deep learning* foi a U-Net. Ela permitiu obter um coeficiente Dice de 0.6642. Todavia, não há como afirmar, sem o apoio de testes estatísticos, que a rede *deep learning* U-Net foi a que gerou o coeficiente Dice com uma significância estatística importante. Outras redes *deep learning* podem ter apresentado resultados muito similares à U-Net na média quando comparados com todo o conjunto tomográfico.

5.3 Testes estatísticos e análise de resultados

Dada a proximidade dos valores obtidos com algumas das redes *deep learning*, foi necessário realizar vários testes estatísticos. Assim, partindo-se dos coeficientes médios Dice para cada *rede deep learning*, coletados na seção anterior para as 4.178 amostras válidas, de acordo com a **Tabela 6**, realizaram-se vários testes estatísticos utilizando-se o software SPSS™ da IBM para validar ou refutar as hipóteses $\mathbf{H_0}$ e $\mathbf{H_1}$. Tal processo dará o suporte necessário para responder à questão proposta: Há diferença significativa de desempenho quanto às diferentes topologias de rede *deep learning* estudadas *vis-à-vis* na segmentação de imagens de tomografias computadorizadas em 18 classes de ossos?

Para isso realizou-se o teste para a análise de variância ANOVA de um fator (ANOVA *one-way*). A normalidade dos dados foi avaliada pelos testes de Kolmogorov-Smirnov e Shapiro-Wilk e a homogeneidade de variância, pelo teste de Levene. Aplicou-se também o pré-processamento de reamostragem (*bootstrapping*) com 1.000 amostras com índice de confiabilidade IC de 95% para aumentar a confiança nos resultados (Haukoos & Lewis, 2005). Empregou-se a correção de Welch sobre os resultados, antevendo-se a possibilidade de heterogeneidade de variância nas amostras.

Os resultados preliminares indicaram que as amostras não apresentam uma distribuição normal (Kolmogorov-Smirnov = 0.10, *p-valor* < 0.001, Shapiro-Wilk = 0.95, *p-valor* < 0.001), bem como não possuem homogeneidade de variância (Levene (4, 20885) *p-valor* < 0.001). Já os resultados da análise de variância ANOVA (cf. **Quadro 5**) mostraram que há diferença estatística significativa entre as médias dos coeficientes Dice para as redes *deep learning* [Welch's F (4, 10434.86) = 614.84, *p-valor* < 0.001] refutando a hipótese nula $\mathbf{H_0}$ e corroborando com a hipótese alternativa $\mathbf{H_1}$.

Quadro 5 - Teste ANOVA com correção de Welch demostrando que as médias dos coeficientes Dice entre algumas das redes é estatisticamente diferente.

ANOVA

Dice

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	103.016	4	25.754	578.074	.000
Within Groups	930.458	20885	.045		
Total	1033.474	20889			

Robust Tests of Equality of Means

Dice

	Statistic[a]	df1	df2	Sig.
Welch	614.836	4	10434.859	.000

a. Asymptotically F distributed.

Para descobrir quais pares de topologias de redes *deep learning* possuem diferença estatística, foi realizado o teste não paramétrico *post-hoc* de Games-Howell, no qual se verificou que apenas as redes *deep learning* DenseNet e DeepLab são similares (*p-valor* = 0.8890 > 0.05) enquanto todas as demais combinações são estatisticamente diferentes (p-valor < 0.01) conforme valores exibidos na **Tabela 9**.

Tabela 9 – Teste de Games-Howell para verificar que apenas as redes *deep learning* DenseNet e DeepLab possuem semelhança estatística.

P-valor	U-Net	DenseNet	ResNet	DeepLab	FCN
U-Net		0.0000	0.0000	0.0000	0.0000
DenseNet	0.0000		0.0000	0.8890	0.0000
ResNet	0.0000	0.0000		0.0000	0.0000
DeepLab	0.0000	0.8890	0.0000		0.0000
FCN	0.0000	0.0000	0.0000	0.0000	

Por fim, analisando as médias dos coeficientes Dice dadas pela **Tabela 10**, juntamente com os resultados estatísticos obtidos, pode-se inferir que a rede *deep learning* U-Net possui média dos coeficientes Dice significativamente superior (0.6854) *vis-à-vis* às demais topologias *deep learning* analisadas. A U-Net pode ser considerada a rede *deep learning* que obteve o melhor desempenho na segmentação automática executada sobre o corpo feminino do VHP. Esse resultado responde à nossa última questão de pesquisa, que encerra a seguinte

indagação: Dentre as redes *deep learning* U-Net, DenseNet, ResNet, DeepLab e FCN existe alguma que possui desempenho superior frente às demais ?

Tabela 10 – Estatísticas sobre 4.178 amostras para o coeficiente Dice e número de parâmetros de treinamento.

Dice

Rede	Contagem	Soma	Média	Variância	Desvio Padrão	Parâmetros
U-Net	4.178	2863.752	0.6854	0.0408	∓0.2021	31,053,965
DenseNet	4.178	2600.513	0.6224	0.0475	∓0.2179	9,426,355
ResNet	4.178	1997.242	0.4780	0.0417	∓0.2042	2,754,771
DeepLab	4.178	2619.416	0.6270	0.0533	∓0.2308	41,257,123
FCN	4.178	2352.172	0.5630	0.0394	∓0.1984	134,455,833

Em termos estatísticos, a rede U-Net possui valores médios superiores às demais. Nessa linha, quando se confronta a quantidade de parâmetros de treinamento com a média dos coeficientes Dice, pode-se verificar que a rede ResNet, que possui a menor quantidade de parâmetros (2,754,771), obteve o pior desempenho seguida da rede FCN, que possui a maior quantidade (134,455,833). Isso indica que a forma como são extraídas as características das imagens de entrada e as camadas ocultas são mais importantes que a quantidade de filtros das camadas convolutivas.

O desempenho superior das redes *deep learning* U-Net, DenseNet e DeepLab está na abordagem que cada uma delas usa para concatenar as características em diferentes níveis. A rede *deep learning* U-Net concatena características em camadas de níveis de profundidade similares enquanto a rede *deep learning* DenseNet concatena recursivamente características em camadas de níveis anteriores. Já a rede *deep learning* DeepLab utiliza a abordagem *Atrous Spatial Pyramid Pooling*, que concatena características de camadas de diferentes níveis utilizando diferentes abordagens de convolução, como *atrous convolution* e *depthwise separable convolution*.

O gráfico *box-and-whisker* da Figura 48 exibe as distribuições dos coeficientes Dice para todas as tomografias do corpo feminino do VHP. Tal gráfico ajuda a visualizar tanto a similaridade entre as médias das redes *deep learning* DenseNet e DeepLab quanto a superioridade da rede *deep learning* U-Net *vis-à-vis* às demais. Da mesma forma, auxilia a visualizar o baixo desempenho das redes *deep learning* ResNet e FCN.

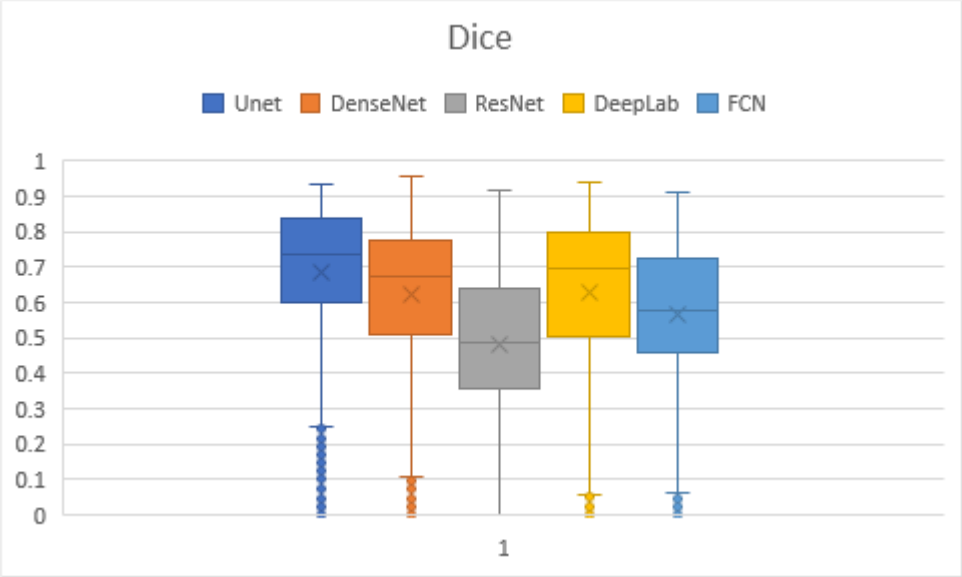


Figura 48 – Gráfico *box-and-whisker* para o coeficiente Dice relativo aos desempenhos de redes *deep learning* na segmentação de imagens tomográficas.

Fonte: Compilação do autor

Finalmente, o **Quadro 6** resume as principais conclusões quanto à segmentação propriamente dos ossos do corpo humano utilizando-se de redes *deep learning*.

Quadro 6 - Principais conclusões e lições aprendidas.

No.	Principais Conclusões
1	A rede *deep learning* **U-Net** apresentou média dos coeficientes Dice (0.6854) estatisticamente superior às demais redes *deep learning* encerrando uma forma mais simples de concatenação de características entre camadas validando a hipótese de pesquisa $\mathbf{H_1}$.
2	As redes *deep learning* **DenseNet** e **DeepLab** foram as únicas topologias a apresentarem similaridade estatística entre as médias de seus coeficientes Dice (0.6224 e 0.6270); porém, encerrando formas mais complexas de colocar em prática a concatenação e extração de características.
3	As redes *deep learning* **ResNet** e **FCN,** apesar de encerrarem a menor e a maior quantidade de parâmetros de treinamento (2,740,435 e 134,455,833), obtiveram as médias estatisticamente mais baixas dos coeficientes Dice (0.4780 e 0.5630). Isso sugere que a forma com que se extraem as características das imagens de entrada nas camadas é mais importante que a quantidade de filtros utilizados para extrair as características delas.

5.4 Principais dificuldades

Segmentação de órgãos em Tomografias Computadorizadas, utilizando redes *deep learning,* envolve diversos desafios. Esses desafios se apresentaram, em particular, durante a análise dos resultados, ou seja, surgiram a partir de *pixels* classificados erroneamente, cujas causas principais estão descritas abaixo.

- Baixa quantidade das amostras tomográficas para o treinamento;
- Formato dos ossos;
- Excesso da presença da classe “fundo” nas tomografias; e
- Bordas muito próximas entre ossos de classes diferentes.

Nas seções a seguir serão abordados, com mais profundidade, os quatro tópicos descritos acima, os quais devem ser levados em conta, principalmente, em trabalhos futuros.

5.4.1 Baixa quantidade de amostras de treinamento

Uma das dificuldades para o aprendizado das redes *deep learning*, quando treinadas em modo supervisionado, é a necessidade de serem apresentadas a muitas amostras durante o treinamento, para se obter uma classificação eficaz. Outro problema que acompanha a baixa quantidade de amostras é o desbalanceamento da quantidade de amostras por classe. A **Figura 49** mostra que costelas e vértebras somadas representam 47,15% do total de amostras, a patela possui menos que 1%. O desequilíbrio decorre de fatores particulares como, por exemplo, o fato de os tamanhos dos ossos serem diferentes entre si, além da falta de bases tomográficas de corpo completo ou até mesmo de outras regiões diferentes do tórax e abdômen disponíveis para estudo.

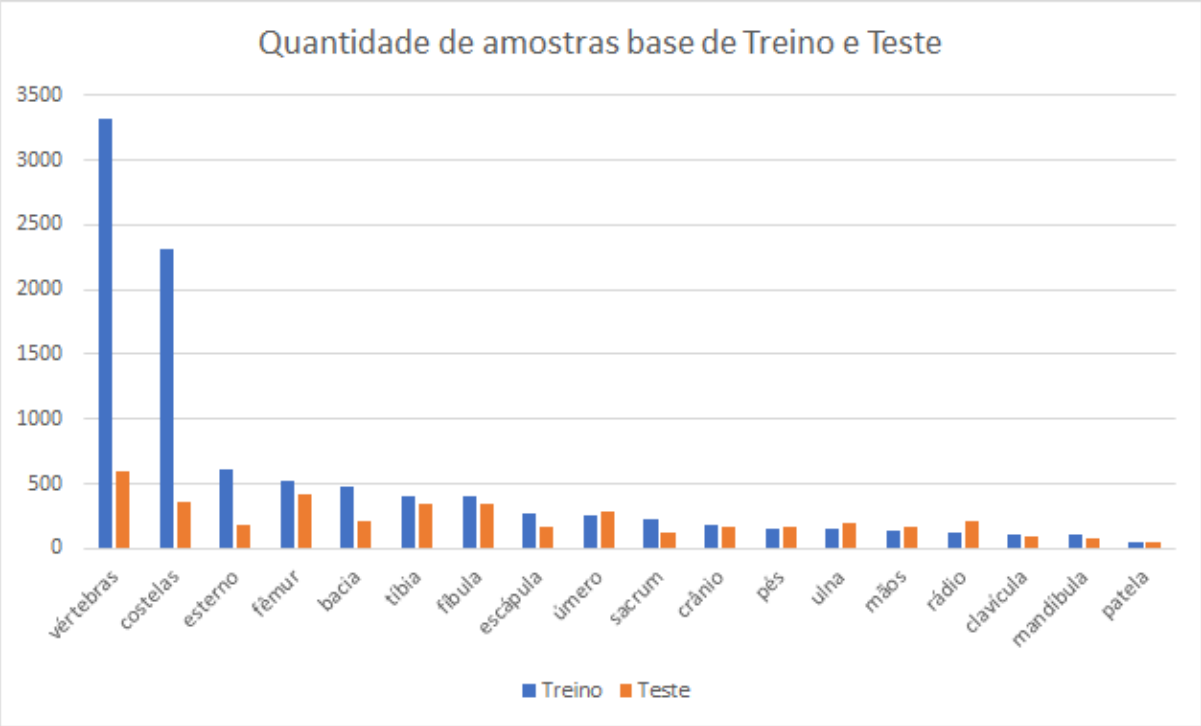


Figura 49 – Quantidade de amostras por classe utilizada para treinamento (VHP masculino + IRCAD) e para testes (VHP feminino). Além da baixa quantidade de amostras para classes como patela, mandíbula e clavícula, é possível notar o desbalanceamento na quantidade individual.

Fonte: Compilação do autor

O problema de desbalanceamento acaba criando uma tendência quanto à classificação das redes a grupos cujas classes possuem maiores quantidades de amostras. Este foi explorado em [66]. É importante ressaltar que tal fator sozinho não é suficiente para uma

correlação direta entre os coeficientes Dice e a quantidade de amostras por classe, como apontado na **Tabela 11.**

Tabela 11 – Coeficiente Dice por quantidade de amostras para cada classe de osso.

Classe	*Amostras*	*Dice*
mãos	310	0.3485
rádio	345	0.3491
sacro	349	0.4228
pés	314	0.4452
esterno	805	0.4476
patela	104	0.4636
ulna	344	0.4721
costelas	2685	0.4961
escápula	440	0.5345
fíbula	746	0.5896
clavícula	197	0.6023
úmero	532	0.6130
vértebras	3924	0.6782
tíbia	761	0.7219
bacia	687	0.7354
mandíbula	188	0.7410
crânio	334	0.7715
fêmur	951	0.7792

Pode-se verificar que, apesar das classes costela e vértebra possuírem um alto número de amostras, estas não têm o melhor coeficiente Dice. Isso ocorre em razão de outros fatores, que devem ser analisados em conjunto como, por exemplo, a anatomia dos ossos em sua extensão, proximidade com outros ossos, dentre outros.

5.4.2 Anatomia dos ossos

Diferentes corpos podem apresentar os mesmos ossos com características diferentes. O conjunto tomográfico do osso esterno de corpo feminino do VHP possui uma particularidade em que o processo xifoide difere-se de todos os demais conjuntos conforme a Figura 50.

Figura 50 – Processo xifoide do osso esterno feminino de forma bifurcada é completamente diferente dos demais conjuntos.

Fonte: Compilação do autor

Nessa ilustração, a rede terá extrema dificuldade em prever uma situação que nunca foi vista durante a sua etapa de treinamento, o que pode ser observado visualmente na **Figura 51**. Isso mostra o quão longe a segmentação de todas as redes ficou da máscara de *ground truth*.

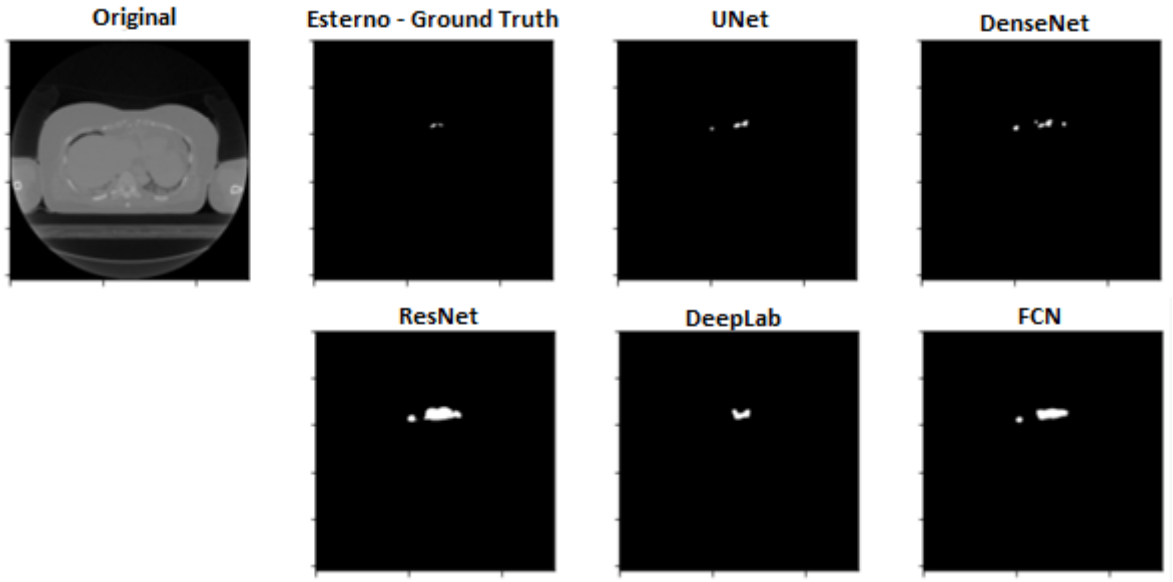


Figura 51 – Segmentação de uma tomografia do processo xifoide feminino do VHP. É notável a diferença entre o ground truth e os resultados da segmentação para todas as redes.

Fonte: Compilação do autor

Outro problema recorrente nos ossos está no formato das extremidades. Isso ocorre devido à necessidade de estabelecerem ligações com outros ossos. Contribui também para o problema o fato de que as cartilagens possuem formas demasiadamente diferentes em relação à sua extensão quando vistas sob o eixo axial, o que dificulta perceber exatamente seu início ou término. Essa questão prejudica tanto na criação das máscaras de *ground truth* quanto na própria segmentação executada pelas redes *deep learning*. Esse problema foi encontrado em todas as classes de ossos. Pode-se elucidar tal dificuldade analisando-se, por exemplo, os coeficientes Dice ao longo da segmentação da tíbia (cf. Figura 52), que mostram valores baixos nas extremidades iniciais e finais.

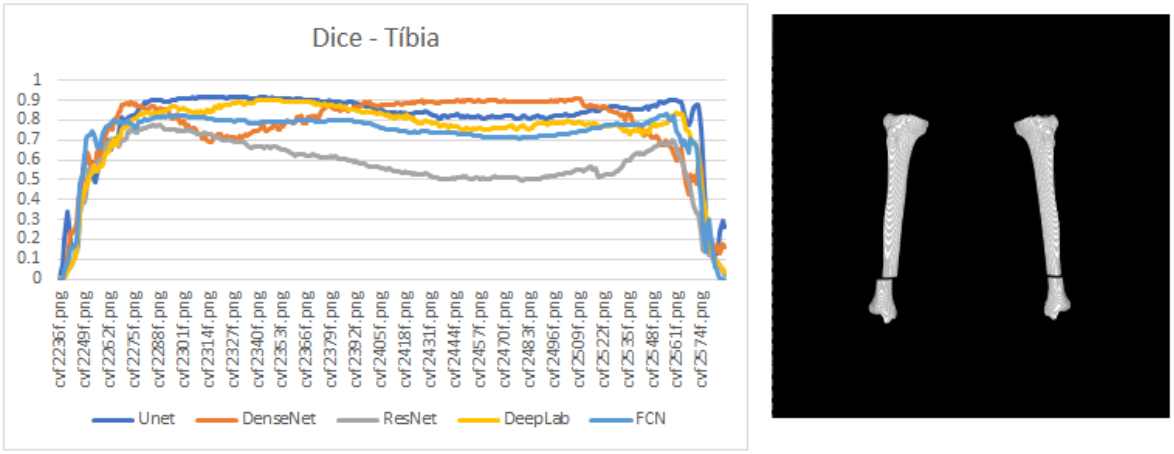


Figura 52 – Coeficientes Dice ao longo da segmentação da tíbia. As extremidades possuem os valores mais baixos.

Fonte: Compilação do autor

A forma irregular de alguns ossos como, por exemplo, das mãos e dos pés em toda sua extensão, assim como o baixo número de amostras também prejudicam o aprendizado das redes. Na comparação (cf. **Figura 53**) entre os coeficientes Dice das mãos e do fêmur, pode-se ver que a irregularidade da forma em cada tomografia, durante o deslocamento axial, acompanha a irregularidade dos valores.

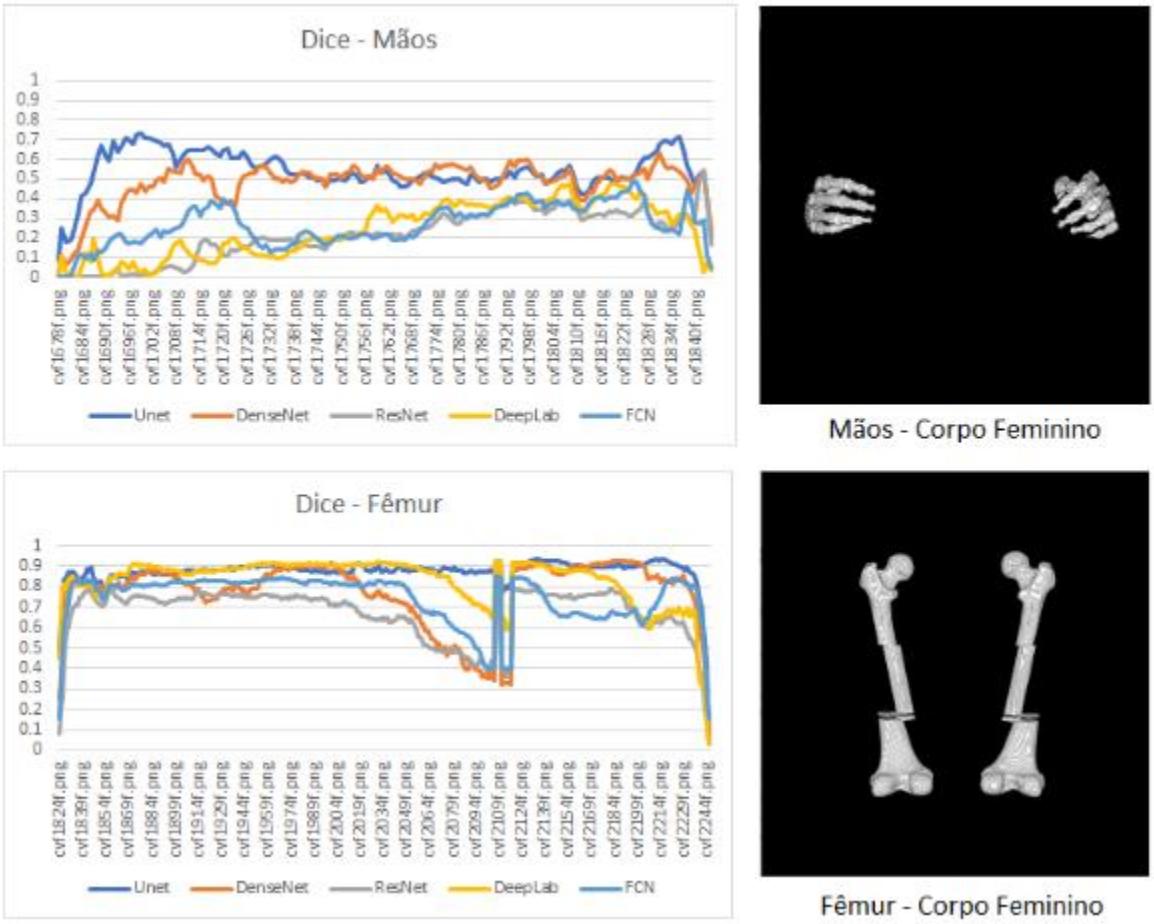


Figura 53 – Coeficientes Dice para a segmentação das mãos e dos ossos do fêmur em toda sua extensão. A irregularidade da forma acompanha a irregularidade dos valores.

Fonte: Compilação do autor

5.4.3 Excesso da classe “fundo”

A classe “fundo” predomina em todas as tomografias. Ela é a maior fonte de erros de segmentação. Esses erros ocorrem mesmo aplicando-se técnicas de balanceamento de pesos na função custo durante o processo de treinamento. Deve-se notar que essa técnica não é suficiente para mitigar esse tipo de problema, o que pode ser verificado em todas as matrizes de confusão nos apêndices; a classe “fundo” ficou em primeiro lugar nos erros de classificação.

5.4.4 Bordas muito próximas entre ossos de outras classes

Um problema comum na segmentação são as bordas que se tocam ou, mais precisamente, aquela região em que diferentes classes estão muito próximas ou sobrepostas em certos pontos. Essa região impõe uma dificuldade suplementar tanto para construir a máscara de *ground truth* quanto para segmentar, propriamente, usando as redes *deep learning*. A **Figura 54** mostra exemplos de ossos com bordas que praticamente se tocam:

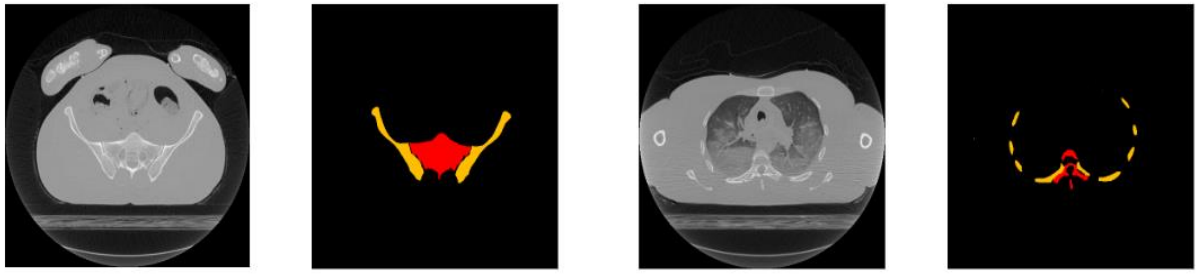


Figura 54 – Na tomografia (a) existe uma dificuldade significativa em separar precisamente as fronteiras entre o osso sacro e o osso da bacia, como mostra a imagem (b). O mesmo se aplica na tomografia (c), para os ossos da costela e vértebras, demonstrado em (d).
Fonte: Compilação do autor

Ocorrência do problema é frequente entre as bordas dos ossos das vértebras e costelas, ou das bordas das costelas com o osso esterno. Na **Figura 55** foi criada uma sobreposição de imagens com o intuito de mostrar a proximidade entre um osso da costela e os ossos das vértebras e também as dificuldades que as redes *deep learning* tiveram em separar ambos os ossos durante a segmentação.

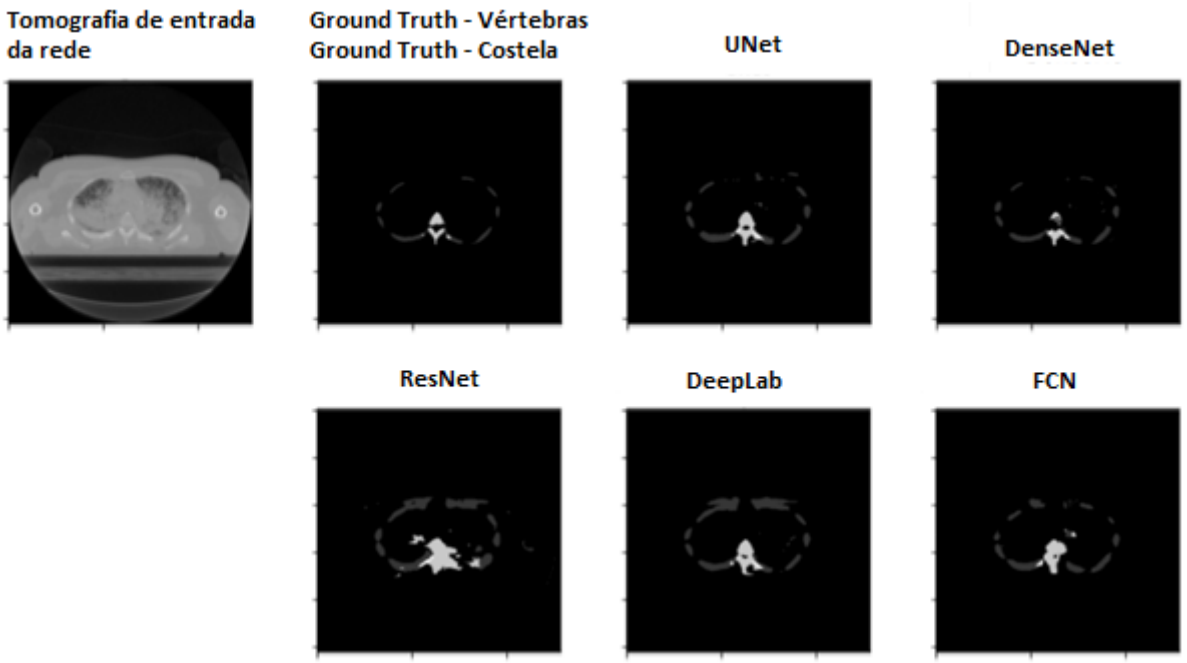


Figura 55 – As vértebras são apresentadas com a coloração cinza-clara e as costelas com a coloração cinza-escura nos resultados da segmentação. As redes apresentaram um grau de dificuldade em separar as bordas entre os ossos quando comparadas com o *ground truth*.
Fonte:Compilação do autor

Para concluir esta seção foram construídas o **Quadro 7** e o **Quadro 8**. Elas resumem os principais desafios tanto durante a realização desta pesquisa quanto no treinamento das redes *deep learning*.

Quadro 7 - Principais desafios durante a construção da pesquisa.

Nº	Principais desafios encontrados durante da pesquisa
1	A criação das máscaras de *ground truth*, devido à baixa qualidade das tomografias, demanda conhecimentos relativos ao formato e localização de cada osso.
2	A disponibilidade de computador com alta capacidade de memória de GPU é altamente relevante para a realização dos experimentos. Um treinamento com 16 GB de memória dedicada levou aproximadamente dois dias para ser concluído.
3	A escolha das topologias a serem estudadas é um desafio uma vez que, dentro dos próprios modelos, existem inúmeras variações, podendo ter impacto direto no resultado final. Uma rede *deep learning* ResNet, com 200 camadas, poderia gerar resultados bem superiores ao que foi realizado no contexto desta pesquisa.
4	A camada de ativação de saída da rede teve que ser customizada para executar a função *Softmax* no eixo do canal das imagens.

Quadro 8 - Principais desafios na tarefa de segmentação dos ossos.

Nº	Principais desafios encontrados na segmentação das redes *deep learning*
1	A baixa quantidade das amostras de treinamento, i.e., 4.688 exemplares, somando 1.865 exemplares da base masculina do VHP com 2.823 exemplares das 20 bases do IRCAD, juntamente com o desbalanceamento do número de amostras entre classes.
2	Ossos com formatos muito diferentes tanto nas extremidades quanto em sua extensão ou até mesmo para diferentes bases tomográficas.
3	O excesso de *pixels* da classe "fundo". Apesar da utilização de técnicas de balanceamento de pesos na função custo, ainda assim essa classe dominou os resultados frente às demais em relação à classificação incorreta.
4	As bordas muito próximas entre ossos de diferentes classes: todas as redes *deep learning* estudadas tiveram muita dificuldade em separar as costelas das vértebras e do osso esterno.

5.5 Considerações finais

Esta seção atingiu o objetivo de responder à questão de pesquisa proposta mostrando, através de análises estatísticas, que, além de haver diferença significativa na segmentação de tomografias computadorizadas entre 18 classes de ossos pelas redes *deep learning* (U-Net, DenseNet, ResNet, DeepLab e FCN), a rede U-Net apresentou um desempenho superior às demais.

Na metodologia empregada, utilizou-se um protocolo rígido de coleta de dados de forma a evitar enviesamento das análises replicando-se cinco vezes o treino de cada rede para cada parte da base de dados e eliminando amostras cujo resultado é ambíguo, condizente às radiografias nas quais a classe avaliada está ausente.

A próxima seção apresentará as conclusões finais e sugestões de pesquisas futuras baseadas nos principais desafios encontrados.

6 Conclusão

O objetivo desta dissertação foi avaliar diferentes topologias de redes *deep learning vis-à-vis* aos seus desempenhos na segmentação de um conjunto de imagens tomográficas do corpo humano feminino do *Visible Human Project* (VHP) em 18 classes de ossos distintos. Os resultados obtidos mostram que existem diferenças significativas na segmentação de imagens de tomografias computadorizadas de corte axial em 18 classes de ossos. Tais diferenças foram observadas para as seguintes topologias de rede *deep learning* experimentadas, a saber: U-Net, DenseNet, ResNet, DeepLab e FCN. A consecução desse objetivo seguiu a aplicação de um *framework* padrão envolvendo as seguintes etapas clássicas: treinamento, teste, coleta de dados e análise estatística intercruzada.

Subobjetivo 1: Um esforço importante do projeto foi a segmentação de *ground truth.* Tal segmentação foi executada sobre as bases de imagens tomográficas do VHP e do Instituto de Pesquisa contra o Câncer do Aparelho Digestivo na França (IRCAD), com vistas ao treinamento supervisionado das redes *deep learning*, levando aproximadamente oito meses de trabalho. Os resultados obtidos foram os seguintes:

- Criação de duas bases de *ground truth*, contendo a primeira 4.304 máscaras para o corpo masculino, e a segunda, 4.178 máscaras para o corpo feminino do VHP, distribuídas entre 18 classes de ossos.
- Criação de uma base de *ground truth* contendo 5.534 máscaras para o conjunto formado pelas 20 bases tomográficas do IRCAD, distribuídas entre nove classes de ossos.

Até o momento de finalização desta dissertação, as bases de *ground truth* criadas possuem a maior quantidade de classes de ossos já utilizadas para o treinamento supervisionado em trabalhos de segmentação médica, sendo 18 no total.

Subobjetivo 2: A experimentação foi realizada sobre um conjunto de cinco topologias de redes *deep learning*. Tal conjunto foi definido com base na literatura de redes *deep learning* relacionadas à segmentação de imagens em tomografia computadorizada do corpo humano; deve-se notar que os itens segmentados foram 18 ossos, a saber: crânio, mandíbula, clavícula, escápula, úmero, rádio, ulna, mãos, costelas, esterno, vértebras, sacro, bacia, fêmur, patela, tíbia, fíbula e pés. O resultado foi o seguinte conjunto de redes *deep learning*: U-Net, DenseNet, ResNet, DeepLab e FCN.

Em relação a trabalhos anteriores, a única comparação possível é o número de classes de ossos utilizados na segmentação, que foi 18, representando até o momento um feito inédito. Porém, a simples adição de outras classes vindas de outras bases implica segmentação manual das bases já segmentadas para essa nova classe, sendo necessário estudar técnicas de aprendizado incremental.

Subobjetivo 3: A realização dos experimentos representaram um esforço computacional importante com cada rede *deep learning* selecionada – U-Net, DenseNet, ResNet, DeepLab ou FCN. Em outras palavras, para cada topologia foram executadas as seguintes etapas: *treinamento* e *teste de segmentação de imagens.* O volume de memória e o cálculo necessário para operacionalizar o processo de treinamento e teste demandaram a seguinte arquitetura computacional: computação dinâmica em nuvem contendo de 16 a 24 GB de CPU com 12 a 16 GB de memória de GPU, podendo variar o fornecimento desse recurso entre as placas de vídeo NVIDIA K80s, T4s, P4s e P100s. O tempo de resposta para completar cada treinamento foi da ordem de 1 dia e 5 horas, sendo que cada treinamento consistia em um lote de imagens de 3.750 para treino e 937 para validação, ocupando 700 MB de memória em disco. Já o experimento final demandou a execução da etapa de treinamento cinco vezes para cada rede selecionada neste projeto, levando tempo considerável uma vez que cada treinamento levou em média de 1 a 2 dias.

Para evitar que os resultados fossem enviesados, todas as redes foram treinadas cinco vezes em partes distintas da base de treinamento e os resultados da segmentação coletados cinco vezes na base de testes. Coeficientes Dice para classes não presentes na tomografia foram ignorados evitando ambiguidade entre aprendizado correto ou problemas de *overfitting* da classe “fundo”.

Subobjetivo 4: O esforço final centrou-se na avaliação dos resultados em termos de desempenho da segmentação de imagens das tomografias da base de testes na seguinte verificação: se havia alguma topologia de rede *deep learning* que apresentava desempenho significativamente superior. No caso, dentre essas cinco topologias de redes, U-Net, DenseNet, ResNet, DeepLab e FCN, a arquitetura U-Net apresentou, de fato, desempenho superior às demais com um coeficiente Dice médio global de 0.6854. As redes DenseNet e DeepLab mostraram rendimento um pouco inferior; porém, estatisticamente similares de 0.6224 e 0.6270. As arquiteturas FCN e ResNet tiveram os piores desempenhos com coeficientes Dices de 0.5630 e 0.4780.

Cabe ressaltar que os resultados aqui obtidos foram gerados considerando-se, explicitamente, as arquiteturas de rede definidas no Capítulo 4. Essa observação é importante

visto que uma mesma topologia pode ser alterada de inúmeras maneiras impactando diretamente nas conclusões. Isso implica dizer que, uma topologia, em sua essência, não é melhor ou pior que a outra, mas sim a arquitetura de camadas derivada de sua concepção original.

Em termos gerais, a conclusão deste projeto de pesquisa é de que há diferença significativa nos resultados da segmentação de imagens de tomografias. Essa diferença está de acordo com a arquitetura utilizada e com a forma com que as redes *deep learning* combinam as características das imagens de entrada entre diferentes camadas nos níveis intermediários, unindo características de alto nível, tais como: formas, cores e contornos com características de baixo nível. Por exemplo: detalhes específicos de cada osso.

6.1 Trabalhos Futuros

O primeiro trabalho futuro que delineamos concerne ao estudo de técnicas que atenuem a influência da classe “fundo”, pois a aplicação de balanceamento na função custo *Weighted Cross Entropy* não permitiu reduzir influência do “fundo” na segmentação resultante. O segundo trabalho futuro seria incluir topologias tridimensionais com camadas convolutivas 3D e redes GAN (*Generative Adversarial Network).* O terceiro consistiria em colocar em sinergia técnicas de concatenação de características de diferentes topologias em uma única arquitetura, o que poderia, em essência, gerar uma de rede com capacidade ampliada de aprendizado. Por fim, um desafio ainda maior seria estudar como treinar, de forma incremental, uma rede *deep learning.* Tal estudo poderia permitir a adição de novas classes de diferentes bases de dados (ou de treinamento), sem que houvesse a necessidade explícita de mudar a base aqui produzida. Isso poderia representar um avanço na direção da construção de um atlas do corpo humano contendo a segmentação de todos os órgãos por um único modelo de rede.

7 Referências Bibliográficas

[1] P. FLECKENSTEIN, J. TRANUM-JENSEN, Anatomia em Diagnóstico Por Imagens: 2. ed.,São Paulo: Manole, 2004 [1] Chunran, Y., & Yuanyuan, W. (2018). Nodule on CT Images. 2–6.

[2] Barrow, H. G., Tenenbaum, J. M., & Park, M. (1978). RECOVERING INTRINSIC SCENE CHARACTERISTICS FROM IMAGES. 3–26.

[3] Hesamian, M. H., Jia, W., He, X., & Kennedy, P. (2019). Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. Journal of Digital Imaging, 32(4), 582–596. https://doi.org/10.1007/s10278-019-00227-x

[4] Chunran, Y., & Yuanyuan, W. (2018). Nodule on CT Images. 2–6.

[5] Shaziya, H., Shyamala, K., & Zaheer, R. (2018). Automatic Lung Segmentation on Thoracic CT Scans Using U-Net Convolutional Network. Proceedings of the 2018 IEEE International Conference on Communication and Signal Processing, ICCSP 2018, 643–647. https://doi.org/10.1109/ICCSP.2018.8524484.

[6] Kumar, A., Fulham, M., Feng, D., & Kim, J. (2019). Co-Learning Feature Fusion Maps from PET-CT Images of Lung Cancer. IEEE Transactions on Medical Imaging, 1–1. https://doi.org/10.1109/tmi.2019.2923601.

[7] Alves, J. H., Neto, P. M. M., & Oliveira, L. F. (2018). Extracting Lungs from CT Images Using Fully Convolutional Networks. Proceedings of the International Joint Conference on Neural Networks, 2018-July. https://doi.org/10.1109/IJCNN.2018.8489223.

[8] Jin, T., Cui, H., Zeng, S., & Wang, X. (2017). Learning Deep Spatial Lung Features by 3D Convolutional Neural Network for Early Cancer Detection. DICTA 2017 - 2017 International Conference on Digital Image Computing: Techniques and Applications, 2017-December, 1–6. https://doi.org/10.1109/DICTA.2017.8227454.

[9] Gerard, S. E., & Reinhardt, J. M. (2019). Pulmonary lobe segmentation using a sequence of convolutional neural networks for marginal learning. Proceedings - International Symposium on Biomedical Imaging, 2019-April(Isbi), 1207–1211. https://doi.org/10.1109/ISBI.2019.8759212.

[10] Farag, A., Lu, L., Roth, H. R., Liu, J., Turkbey, E., & Summers, R. M. (2017). A Bottom-Up Approach for Pancreas Segmentation Using Cascaded Superpixels and

(Deep) Image Patch Labeling. IEEE Transactions on Image Processing, 26(1), 386–399. https://doi.org/10.1109/TIP.2016.2624198.

[11] Man, Y., Huang, Y., Feng, J., Li, X., & Wu, F. (2019). Deep Q Learning Driven CT Pancreas Segmentation with Geometry-Aware U-Net. IEEE Transactions on Medical Imaging, 38(8), 1971–1980. https://doi.org/10.1109/TMI.2019.2911588.

[12] Huang, C. H., Xiao, W. T., Chang, L. J., Tsai, W. T., & Liu, W. M. (2018). Automatic tissue segmentation by deep learning: From colorectal polyps in colonoscopy to abdominal organs in CT exam. VCIP 2018 - IEEE International Conference on Visual Communications and Image Processing, 1–4. https://doi.org/10.1109/VCIP.2018.8698645.

[13] Zhou, Y., Wang, Y., Tang, P., Bai, S., Shen, W., Fishman, E. K., & Yuille, A. (2019). Semi-supervised 3D abdominal multi-organ segmentation via deep multi-planar co-training. Proceedings - 2019 IEEE Winter Conference on Applications of Computer Vision, WACV 2019, 121–140. https://doi.org/10.1109/WACV.2019.00020.

[14] La Rosa, F. (2017). A deep learning approach to bone segmentation in CT scans. 66. Retrieved from AMSLaurea Institutional Thesis Repository.

[15] Li, X., Chen, H., Qi, X., Dou, Q., Fu, C. W., & Heng, P. A. (2018). H-DenseU-Net: Hybrid Densely Connected U-Net for Liver and Tumor Segmentation from CT Volumes. IEEE Transactions on Medical Imaging, 37(12), 2663–2674. https://doi.org/10.1109/TMI.2018.2845918.

[16] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. Journal of Big Data, 6(1). https://doi.org/10.1186/s40537-019-0197-0.

[17] Schiemann, T., Tiede, U., & Höhne, K. H. (1997). Segmentation of the Visible Human for high-quality volume-based visualization. Medical Image Analysis, 1(4), 263–270. https://doi.org/10.1016/S1361-8415(97)85001-3.

[18] Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539.

[19] Shell, Adam, How to invest in artificial intelligence, 2020, Disponível em: < https://www.usatoday.com/story/money/2020/01/27/artificial-intelligence-how-invest/4542467002/.

[20] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (2013). Learning Internal Representations by Error Propagation. Readings in Cognitive Science: A Perspective from Psychology and Artificial Intelligence, (V), 399–421. https://doi.org/10.1016/B978-1-4832-1446-7.50035-2

[21] Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition.

[22] Springenberg, J. T., Dosovitskiy, A., Brox, T., & Riedmiller, M. (2015). Striving for simplicity: The all convolutional net. 3rd International Conference on Learning Representations, ICLR 2015 - Workshop Track Proceedings, 1–14.

[23] Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., & Terzopoulos, D. (2019). Image Segmentation Using Deep Learning : A Survey. 1–23.

[24] Kuok, Chan-Pang and Hsue, Jin-Yuan and Shen, Ting-Li and Huang, Bing-Feng and Chen, Chi-Yeh and Sun, Y.-N. (2018). Segmentation from 3D CT Images. Pacific Neighborhood Consortium Annual Conference and Joint Meetings (PNC), (c), 1–6.

[25] Lee, M. J., Hong, H., Shim, K. W., & Park, S. (2019). MGB-NET: Orbital bone segmentation from head and neck ct images using multi-graylevel-bone convolutional networks. Proceedings - International Symposium on Biomedical Imaging, 2019-April(Isbi), 692–695. https://doi.org/10.1109/ISBI.2019.8759424

[26] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9351, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

[27] Wang, C., Song, H., Chen, L., Li, Q., Yang, J., Hu, X. T., & Zhang, L. (2019). Automatic Liver Segmentation Using Multi-plane Integrated Fully Convolutional Neural Networks. Proceedings - 2018 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2018, 518–523. https://doi.org/10.1109/BIBM.2018.8621257

[28] Salehinejad, H., Valaee, S., Dowdell, T., & Barfett, J. (2018). IMAGE AUGMENTATION USING RADIAL TRANSFORM FOR TRAINING DEEP NEURAL NETWORKS Department of Electrical & Computer Engineering , University of Toronto , Toronto , Canada Department of Medical Imaging , St . Michael ' s Hospital , University of Toronto , Toro. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 3016–3020. https://doi.org/10.1109/ICASSP.2018.8462241

[29] Badrinarayanan, V., Kendall, A., Cipolla, R., & Member, S. (n.d.). SegNet : A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. 1–14.

[30] Huang, G., & Weinberger, K. Q. (n.d.). Densely Connected Convolutional Networks.

[31] Gibson, E., Giganti, F., Hu, Y., Bonmati, E., Bandula, S., Gurusamy, K., … Barratt, D. C. (2018). Automatic Multi-Organ Segmentation on Abdominal CT with Dense V-

Networks. IEEE Transactions on Medical Imaging, 37(8), 1822–1834. https://doi.org/10.1109/TMI.2018.2806309

[32] Chen, X., Zhang, R., & Yan, P. (2019). Feature fusion encoder decoder network for automatic liver lesion segmentation. Proceedings - International Symposium on Biomedical Imaging, 2019-April(Isbi), 430–433. https://doi.org/10.1109/ISBI.2019.8759555

[33] Wang, L., Chen, R., Wang, S., Zeng, N., Huang, X., & Liu, C. (2019). Nested Dilation Network (NDN) for Multi-Task Medical Image Segmentation. IEEE Access, 7, 44676–44685. https://doi.org/10.1109/ACCESS.2019.2908386

[34] Wu, Z., Shen, C., & van den Hengel, A. (2019). Wider or Deeper: Revisiting the ResNet Model for Visual Recognition. Pattern Recognition, 90, 119–133. https://doi.org/10.1016/j.patcog.2019.01.006

[35] Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11211 LNCS, 833–851. https://doi.org/10.1007/978-3-030-01234-2_49

[36] Xia, K., Yin, H., Qian, P., Jiang, Y., & Wang, S. (2019). Liver semantic segmentation algorithm based on improved deep adversarial networks in combination of weighted loss function on abdominal CT images. IEEE Access, 7, 96349–96358. https://doi.org/10.1109/ACCESS.2019.2929270

[37] Nguyen, N. Q., & Lee, S. W. (2019). Robust Boundary Segmentation in Medical Images Using a Consecutive Deep Encoder-Decoder Network. IEEE Access, 7, 33795–33808. https://doi.org/10.1109/ACCESS.2019.2904094

[38] Goodfellow, I. J., Pouget-abadie, J., Mirza, M., Xu, B., & Warde-farley, D. (n.d.). Generative Adversarial Nets. 1–9.

[39] Isola, P., Efros, A. A., Ai, B., & Berkeley, U. C. (n.d.). Image-to-Image Translation with Conditional Adversarial Networks.

[40] Dou, Q. I., Ouyang, C., Chen, C., Chen, H. A. O., Glocker, B. E. N., Zhuang, X., … Member, S. (2020). PnP-AdaNet : Plug-and-Play Adversarial Domain Adaptation Network at Unpaired Cross-Modality Cardiac Segmentation. 99065–99076.

[41] Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J., & Greenspan, H. (2018). Synthetic data augmentation using GAN for improved liver lesion classification. Proceedings -

International Symposium on Biomedical Imaging, 2018-April, 289–293. https://doi.org/10.1109/ISBI.2018.8363576

[42] Ge, Y., Wei, D., Xue, Z., Wang, Q., Zhou, X., Zhan, Y., & Liao, S. (2019). UNPAIRED MR TO CT SYNTHESIS WITH EXPLICIT STRUCTURAL CONSTRAINED ADVERSARIAL LEARNING Shanghai United Imaging Intelligence Co ., Ltd , Shanghai , China Institute for Medical Imaging Technology , School of Biomedical Engineering , Shanghai Jiao Tong. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), (Isbi), 1096–1099.

[43] Fang, L., Liu, J., Liu, J., & Mao, R. (2018). Automatic Segmentation and 3D Reconstruction of Spine Based on FCN and Marching Cubes in CT Volumes. 2018 10th International Conference on Modelling, Identification and Control (ICMIC), (Icmic), 1–5.

[44] Sangeetha, V., & Prasad, K. J. R. (2006). Syntheses of novel derivatives of 2-acetylfuro[2,3-a]carbazoles, benzo[1,2-b]-1,4-thiazepino[2,3-a]carbazoles and 1-acetyloxycarbazole-2- carbaldehydes. Indian Journal of Chemistry - Section B Organic and Medicinal Chemistry, 45(8), 1951–1954. https://doi.org/10.1002/chin.200650130

[45] Visible Korean (acessado em 22 de março de 2020), http://vkh3.kisti.re.kr/

[46] IRCAD France - Research Institute against Digestive Cancer (acessado em 23 de março de 2020) ,https://www.ircad.fr/

[46] Shelhamer, E., Long, J., & Darrell, T. (2017). Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(4), 640–651. https://doi.org/10.1109/TPAMI.2016.2572683

[47] Jiang, H., Shi, T., Bai, Z., & Huang, L. (2019). AHCNet: An Application of Attention Mechanism and Hybrid Connection for Liver Tumor Segmentation in CT Volumes. IEEE Access, 7, 24898–24909. https://doi.org/10.1109/ACCESS.2019.2899608

[48] Ahmad, M., Ai, D., Xie, G., Qadri, S. F., Song, H., Huang, Y., … Yang, J. (2019). Deep Belief Network Modeling for Automatic Liver Segmentation. IEEE Access, 7, 20585–20595. https://doi.org/10.1109/ACCESS.2019.2896961

[49] Shrestha, U., & Salari, E. (2018). Automatic Tumor Segmentation Using Machine Learning Classifiers. IEEE International Conference on Electro Information Technology, 2018-May, 153–158. https://doi.org/10.1109/EIT.2018.8500205

[50] Wang, Z. H., Liu, Z., Song, Y. Q., & Zhu, Y. (2019). Densely connected deep U-Net for abdominal multi-organ segmentation. Proceedings - International Conference on Image

Processing, ICIP, 2019-September, 1415–1419. https://doi.org/10.1109/ICIP.2019.8803103

[51] Rafiei, S., Nasr-Esfahani, E., Soroushmehr, S. M. R., Karimi, N., Samavi, S., & Najarian, K. (2018). Liver segmentation in ct images using three dimensional to two dimensional fully convolutional network. ArXiv, 2067–2071.

[52] Truong, T. N., Dam, V. D., & Le, T. S. (2018). Medical Images Sequence Normalization and Augmentation: Improve Liver Tumor Segmentation from Small Data Set. Proceedings - 2018 3rd International Conference on Control, Robotics and Cybernetics, CRC 2018, 1–5. https://doi.org/10.1109/CRC.2018.00010

[53] Van De Leemput, S. C., Meijs, M., Patel, A., Meijer, F. J. A., Van Ginneken, B., & Manniesing, R. (2019). Multiclass brain tissue segmentation in 4D CT using convolutional neural networks. IEEE Access, 7, 51557–51569. https://doi.org/10.1109/ACCESS.2019.2910348

[54] Chen, S., Yang, H., Fu, J., Mei, W., Ren, S., Liu, Y., … Chen, H. (2019). U-Net Plus: Deep Semantic Segmentation for Esophagus and Esophageal Cancer in Computed Tomography Images. IEEE Access, 7, 82867–82877. https://doi.org/10.1109/ACCESS.2019.2923760

[55] Trullo, R., Petitjean, C., Ruan, S., Dubray, B., Nie, D., & Shen, D. (2017). Segmentation of Organs at Risk in thoracic CT images using a SharpMask architecture and Conditional Random Fields. Proceedings - International Symposium on Biomedical Imaging, 1003–1006. https://doi.org/10.1109/ISBI.2017.7950685

[56] Trullo, R., Petitjean, C., Nie, D., Shen, D., & Ruan, S. (2017). Fully automated esophagus segmentation with a hierarchical deep learning approach. Proceedings of the 2017 IEEE International Conference on Signal and Image Processing Applications, ICSIPA 2017, 503–506. https://doi.org/10.1109/ICSIPA.2017.8120664

[57] Tang, Z., Chen, K., Pan, M., Wang, M., & Song, Z. (2019). An Augmentation Strategy for Medical Image Processing Based on Statistical Shape Model and 3D Thin Plate Spline for Deep Learning. IEEE Access, 7, 133111–133121. https://doi.org/10.1109/ACCESS.2019.2941154

[58] Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, 1–15.

[59] Cicuttin, A., Crespo, M. L., Mannatunga, K. S., Garcia, V. V., Baldazzi, G., Rignanese, L. P., … Zorzi, N. (2016). A programmable System-on-Chip based digital pulse

processing for high resolution X-ray spectroscopy. 2016 International Conference on Advances in Electrical, Electronic and Systems Engineering, ICAEES 2016, 15, 520–525. https://doi.org/10.1109/ICAEES.2016.7888100

[60] Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. 32nd International Conference on Machine Learning, ICML 2015, 1, 448–456.

[61] Jegou, S., Drozdzal, M., Vazquez, D., Romero, A., & Bengio, Y. (2017). The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2017-July, 1175–1183. https://doi.org/10.1109/CVPRW.2017.156

[62] Liciotti, D., Paolanti, M., Pietrini, R., Frontoni, E., & Zingaretti, P. (2018). Convolutional Networks for Semantic Heads Segmentation using Top-View Depth Data in Crowded Environment. Proceedings - International Conference on Pattern Recognition, 2018-August, 1384–1389. https://doi.org/10.1109/ICPR.2018.8545397

[63] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017-January, 1800–1807. https://doi.org/10.1109/CVPR.2017.195

[64] Dhawan, V., Sethi, G., Lather, V. S., & Sohal, K. (2013). Power Law Transformation and Adaptive Gamma Correction : A Comparative Study. 7109, 118–123.

[65] Gholamy, A., Kreinovich, V., & Kosheleva, O. (2018). Why 70/30 or 80/20 Relation Between Training and Testing Sets : A Pedagogical Explanation. Departmental Technical Reports (CS)

[66] Guo, X., Yin, Y., Dong, C., Yang, G., & Zhou, G. (2008). On the class imbalance problem. Proceedings - 4th International Conference on Natural Computation, ICNC 2008, 4(October), 192–201. https://doi.org/10.1109/ICNC.2008.871

8 Apêndice A

O apêndice a seguir apresenta dados referentes aos resultados da segmentação óssea em tomografias computadorizadas de eixo axial para as seguintes redes *deep learning*: U-Net, DenseNet, ResNet, DeepLab e FCN. Cada osso recebeu uma ficha técnica contendo as seguintes informações:

- Breve descrição do osso.
- Quadros estatísticos com a contagem de amostras, soma dos valores, média, variância e desvio padrão dos coeficientes Dice para cada uma das redes avaliadas.
- Gráficos com os coeficientes Dice, resultado da segmentação individual de cada tomografia ao longo do osso.
- Linhas das *matrizes de confusão* para cada classe informando a quantidade média de acertos e erros na classificação dos *pixels* para cada rede *deep learning*.
- Resultado tridimensional da segmentação de cada rede *deep learning* juntamente com o *ground truth* para comparação visual.

Notas sobre os resultados:

- Todas as estatísticas foram calculadas usando-se apenas as tomografias na qual a classe existe de acordo com o "quadro de índices tomográficos". Nele são colocados as referências das imagens inicial e final de cada classe de osso. Isso significa, por exemplo, que para a classe *clavícula* os coeficientes foram calculados apenas levando-se em conta as tomografias que seguem sequencialmente da imagem cvf1234f.png até a imagem cvf1319f.png.

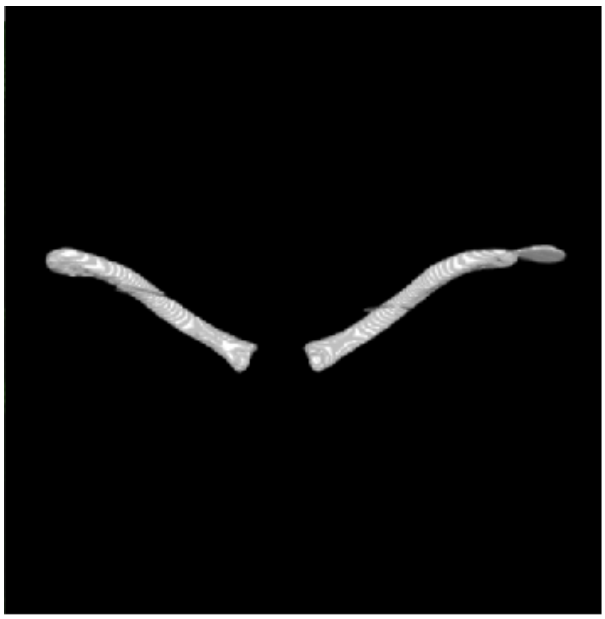
Tabela de índices tomográficos

Classe	***Início***	***Fim***	***Qtd.***
clavícula	cvf1234f.png	cvf1319f.png	86
crânio	cvf1005f.png	cvf1164f.png	160
pés	cvf2566f.png	cvf2727f.png	162
fêmur	cvf1824f.png	cvf2246f.png	423
fíbula	cvf2258f.png	cvf2602f.png	345
mãos	cvf1678f.png	cvf1844f.png	167
bacia	cvf1701f.png	cvf1913f.png	213
úmero	cvf1247f.png	cvf1526f.png	280
mandíbula	cvf1131f.png	cvf1212f.png	82
patela	cvf2193f.png	cvf2244f.png	52
rádio	cvf1507f.png	cvf1722f.png	216
costelas	cvf1261f.png	cvf1626f.png	364
sacro	cvf1722f.png	cvf1848f.png	127
escápula	cvf1233f.png	cvf1396f.png	164
esterno	cvf1309f.png	cvf1496f.png	188
tíbia	cvf2236f.png	cvf2586f.png	351
ulna	cvf1517f.png	cvf1714f.png	198
vértebras	cvf1146f.png	cvf1743f.png	598

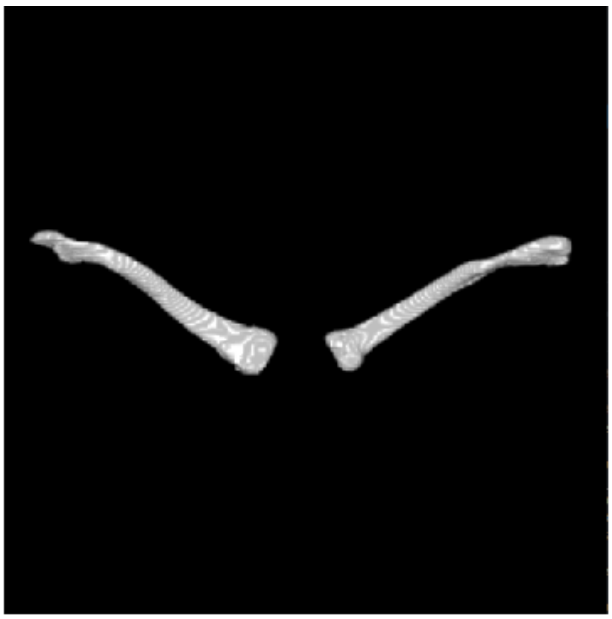
- Todas as estatísticas são calculadas a partir da média dos cinco experimentos para cada rede *deep learning*.
- Todas as estatísticas são referentes ao conjunto tomográfico do corpo feminino do *Visible Human Project* (VHP).

8.1 Clavícula

A clavícula é um osso alongado em formato de “S” que liga os membros superiores ao tronco. A clavícula masculina é mais espessa e curva enquanto a feminina, mais curta e menos curva.



Clavícula - Corpo Masculino

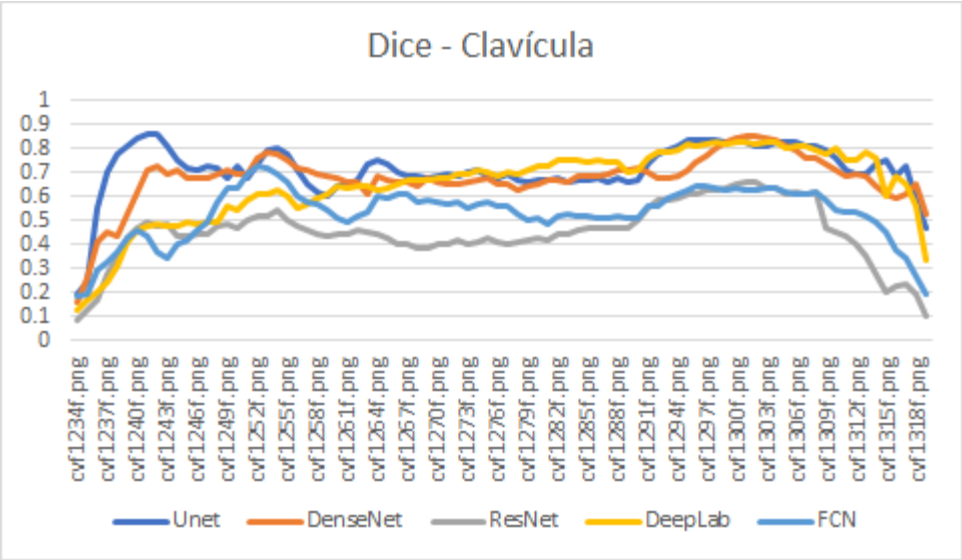


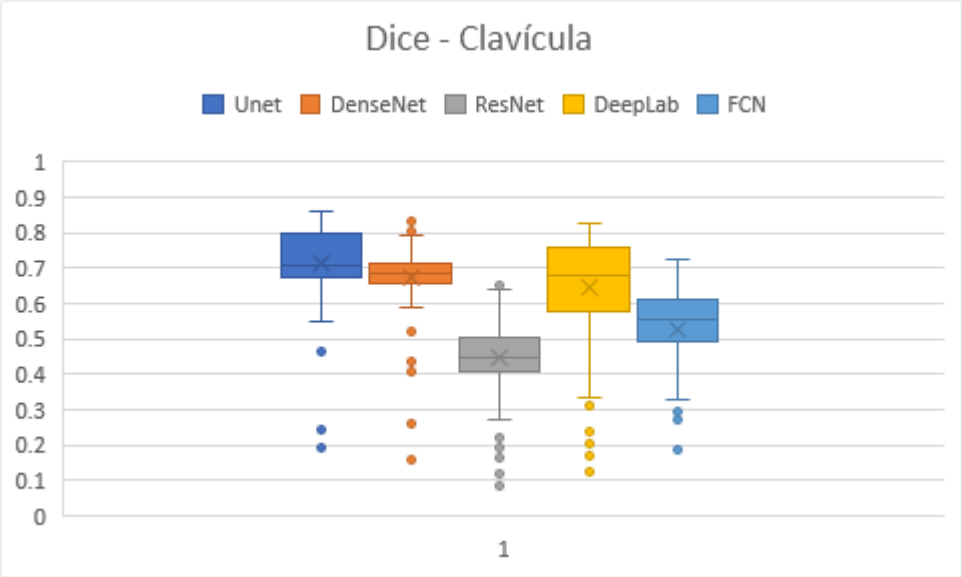
Clavícula - Corpo Feminino

Quadro estatístico contendo a contagem de tomografias, soma, média, variância e desvio padrão para os coeficientes Dice na segmentação da clavícula.

Dice					
Rede	*Contagem*	*Soma*	*Média*	*Variância*	*Desvio Padrão*
U-Net	86	61.299	0.7128	0.0114	∓0.1063
DenseNet	86	57.996	0.6744	0.0117	∓0.1074
ResNet	86	38.648	0.4494	0.0158	∓0.1249
DeepLab	86	55.628	0.6468	0.0249	∓0.1570
FCN	86	45.419	0.5281	0.0129	∓0.1127

Desempenho das redes *deep learning* U-Net, DenseNet, ResNet, DeepLab e FCN para os coeficientes Dice na segmentação da clavícula do corpo feminino.

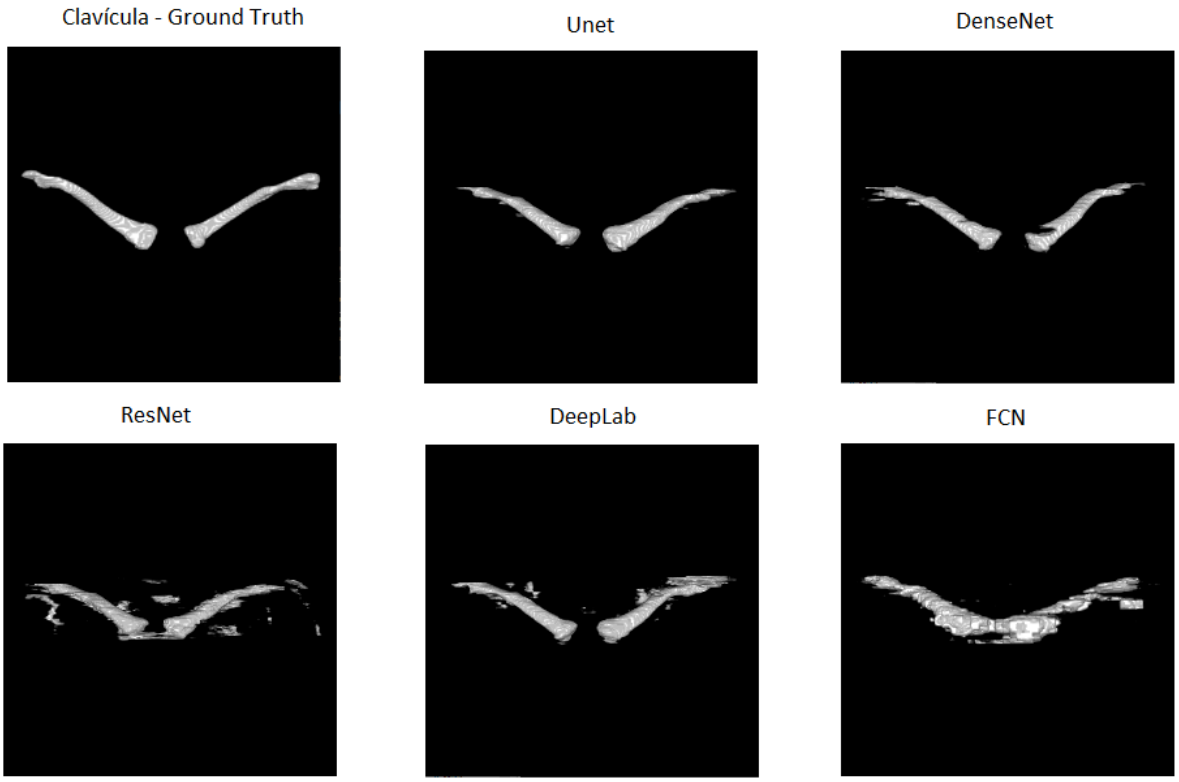




Quadro com as linhas da *matriz de confusão* para a segmentação da clavícula do corpo feminino do VHP.

Matriz Confusão	U-Net	DenseNet	ResNet	DeepLab	FCN
fundo	45236.6	95711.2	235540.6	163237	84626.6
clavícula	**62583.2**	**56587.8**	**60361.6**	**52636.4**	**55285.8**
crânio	922.6	2639	2766.6	0	525.8
pés	244	9729.6	3006.6	16.4	214.6
fêmur	136	44501	9452.4	76721.8	635.8
fíbula	168	4865.4	73.6	3.4	0
mãos	0	482.2	1287.2	250	955.4
bacia	2648.8	5118.8	4010.6	2132	350.2
úmero	6	6755.4	1759.6	228.6	98.2
mandíbula	2.4	799.2	189.4	0	69.2
patela	0	134	184	71.8	0
rádio	0.4	547	51.4	1.8	0
costelas	2620.4	610	4466.2	1841.4	3920
sacro	0	0	0	0	95.6
escápula	2567.2	4760.4	9049.6	7431.4	3813.8
esterno	2752.4	79.4	1392.6	819.8	3354.8
tíbia	78.6	67840.6	1855	33338.4	715.6
ulna	0.2	1646	162.8	1.2	86.4
vértebras	42.8	386.2	530.8	18.6	406.6

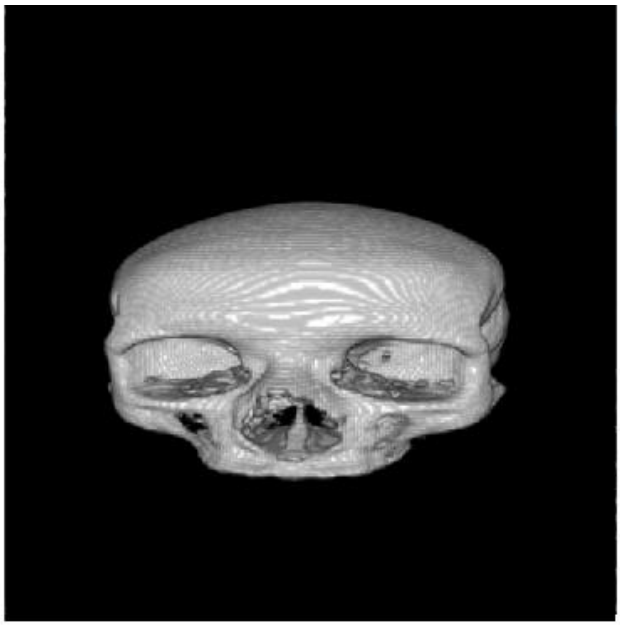
Resultado da segmentação das redes U-Net, DenseNet, ResNet, DeepLab e FCN juntamente com o *ground truth* para a clavícula feminina.



8.2 Crânio

O crânio é uma estrutura óssea complexa formada de vários ossos menores, cuja função é proteger o cérebro assim como dar suporte às áreas faciais.

Crânio - Corpo Masculino

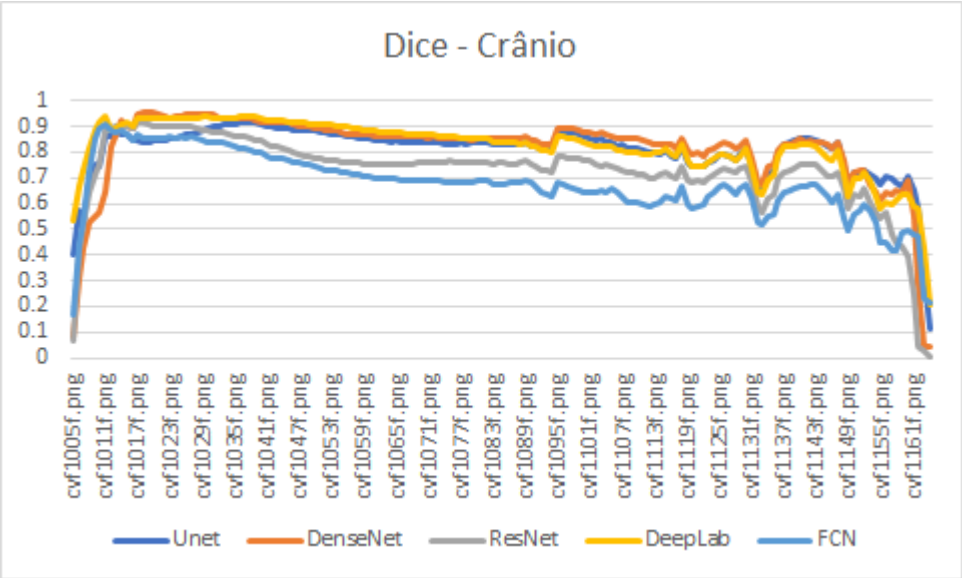


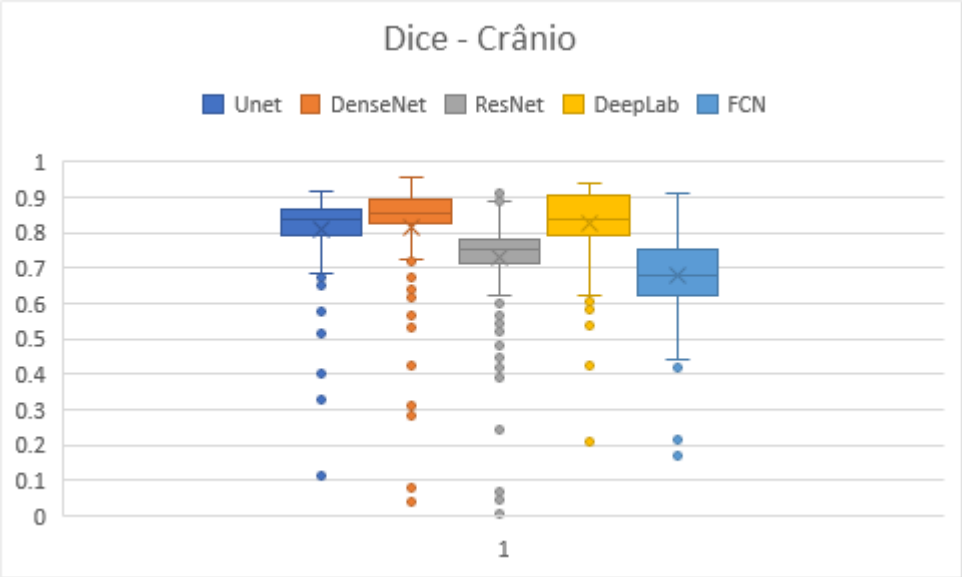
Crânio - Corpo Feminino

Quadro estatístico contendo a contagem de tomografias, soma, média, variância e desvio padrão para os coeficientes Dice na segmentação do crânio.

Dice					
Rede	*Contagem*	*Soma*	*Média*	*Variância*	*Desvio Padrão*
U-Net	86	72.626	0.8445	0.0059	∓0.1031
DenseNet	86	73.616	0.8560	0.0192	∓0.1525
ResNet	86	68.135	0.7923	0.0129	∓0.1533
DeepLab	86	76.291	0.8871	0.0036	∓0.1086
FCN	86	64.549	0.7506	0.0112	∓0.1257

Desempenho das redes *deep learning* U-Net, DenseNet, ResNet, DeepLab e FCN para os coeficientes Dice na segmentação do crânio do corpo feminino.

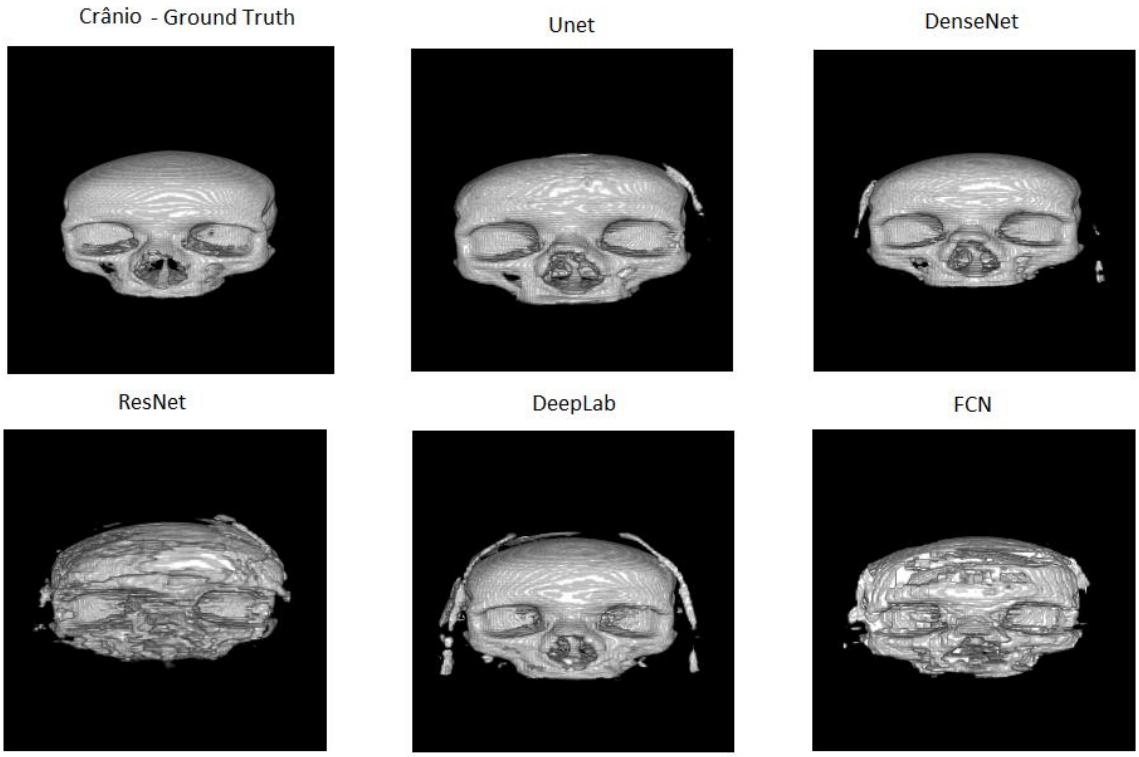




Quadro com as linhas da *matriz de confusão* para a segmentação do crânio do corpo feminino do VHP.

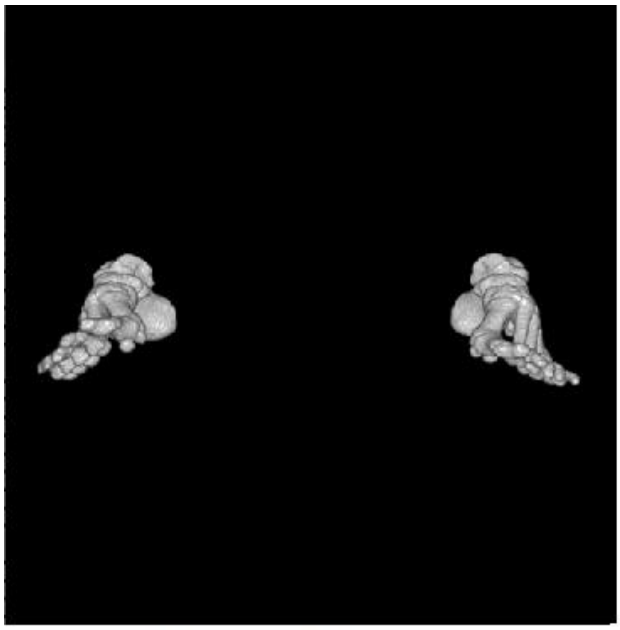
Matriz Confusão	U-Net	DenseNet	ResNet	DeepLab	FCN
fundo	817264.8	1501903.4	1501903.4	886664.4	2273168
clavícula	0	0	0	0	8.4
crânio	**2106724.4**	**2002632.4**	**2002632.4**	**1749514.6**	**2105683**
pés	37521.8	245606.2	245606.2	319257.2	91195.8
fêmur	21.4	73736.6	73736.6	7682.4	3652.4
fíbula	2104.6	7763.4	7763.4	4.4	6.4
mãos	1391.6	1723.4	1723.4	0	11678.4
bacia	4.8	65.6	65.6	119.2	815.4
úmero	0	305.8	305.8	0	7.6
mandíbula	6019.6	2198.4	2198.4	307.6	10445.8
patela	0	1043.6	1043.6	0	2389
rádio	0	126.4	126.4	0	1.4
costelas	0	453.2	453.2	0	2869.2
sacro	2	1.2	1.2	0	343.6
escápula	1.6	211.6	211.6	0	203.8
esterno	0	0	0	0	0
tíbia	1288.8	20093	20093	2448	295.4
ulna	0	28.8	28.8	0	3.6
vértebras	6530.8	7865	7865	12149.8	17672.8

Resultado da segmentação das redes U-Net, DenseNet, ResNet, DeepLab e FCN juntamente com o *ground truth* para o crânio feminino.

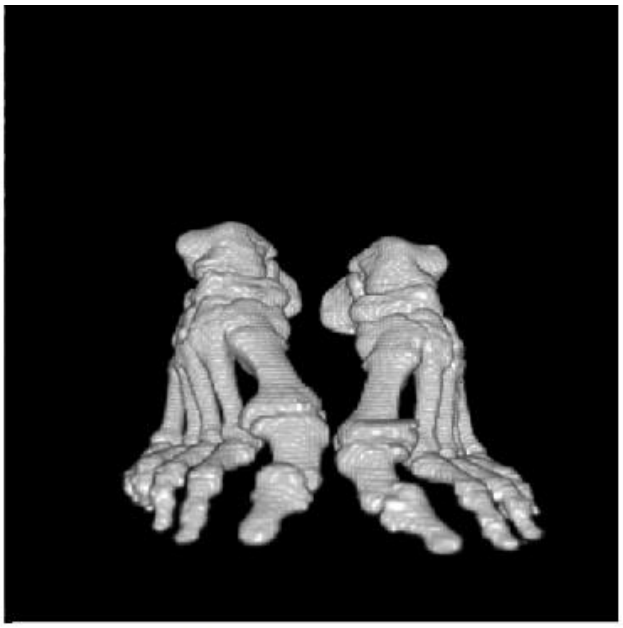


8.3 Pés

Os pés, compostos pelos ossos do tarso, metatarso e falanges, têm a função tanto de sustentar quanto de ajudar no deslocamento do corpo.



Pés - Corpo Masculino

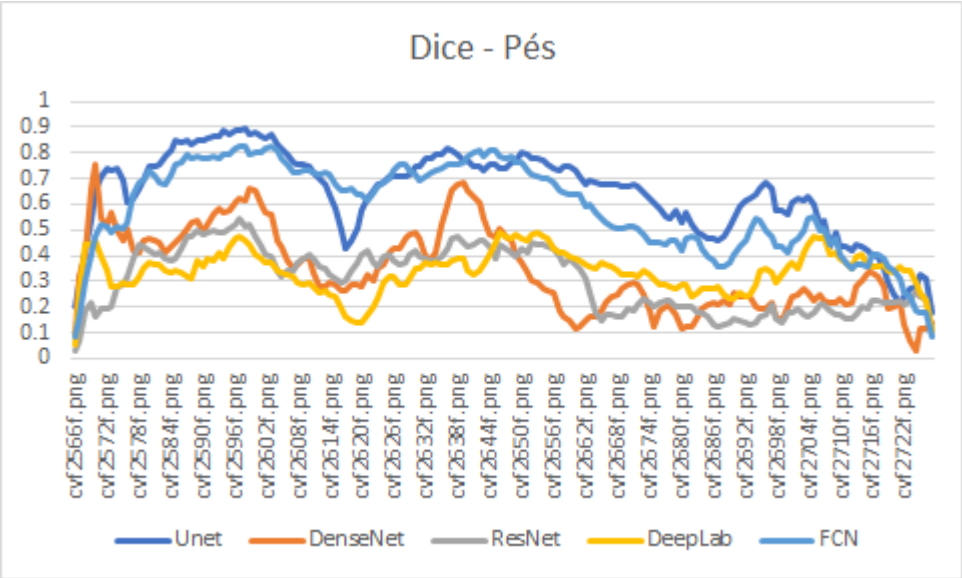


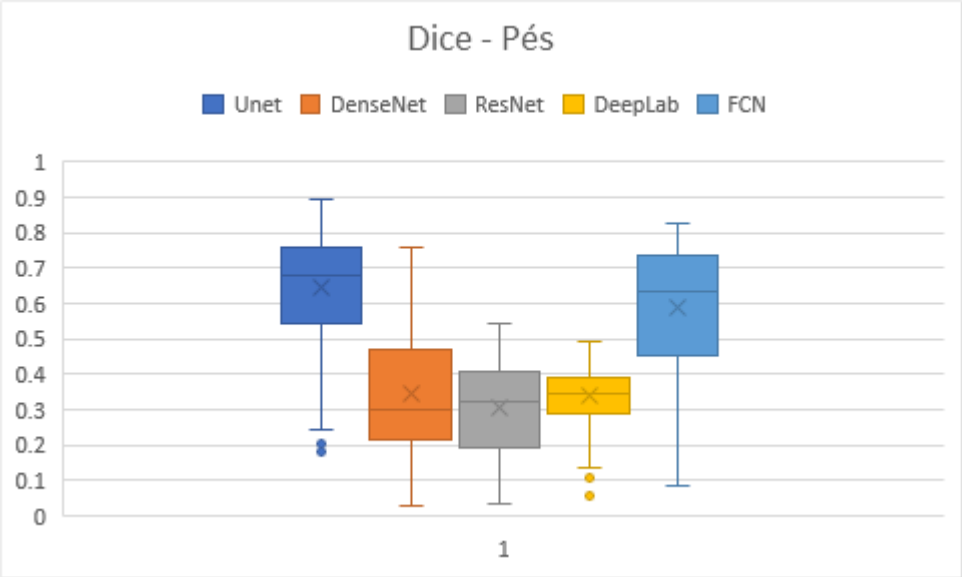
Pés - Corpo Feminino

Quadro estatístico contendo a contagem de tomografias, soma, média, variância e desvio padrão para os coeficientes Dice na segmentação dos pés.

Dice					
Rede	*Contagem*	*Soma*	*Média*	*Variância*	*Desvio Padrão*
U-Net	162	104.898	0.6475	0.0273	∓0.1646
DenseNet	162	55.796	0.3444	0.0261	∓0.1612
ResNet	162	49.603	0.3062	0.0152	∓0.1228
DeepLab	162	55.136	0.3403	0.0064	∓0.0799
FCN	162	95.217	0.5878	0.0313	∓0.1763

Desempenho das redes *deep learning* U-Net, DenseNet, ResNet, DeepLab e FCN para os coeficientes Dice na segmentação dos pés do corpo feminino.

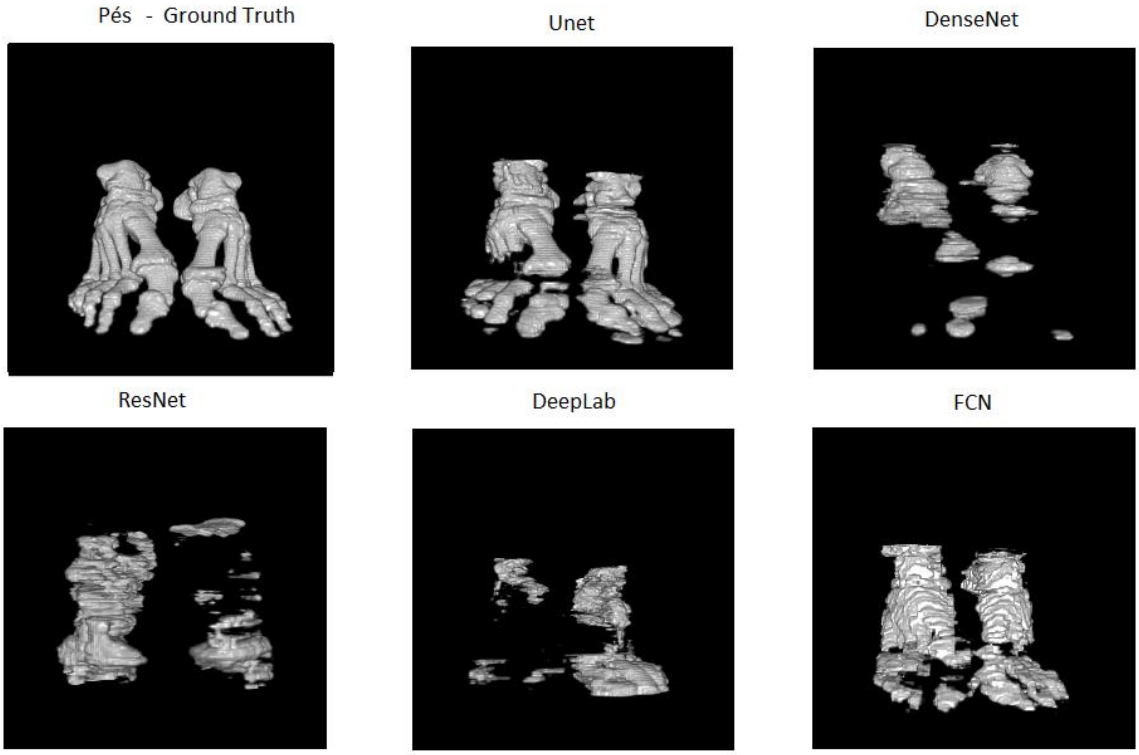




Quadro com as linhas da *matriz de confusão* para a segmentação dos pés corpo feminino do VHP.

Matriz Confusão	U-Net	DenseNet	ResNet	DeepLab	FCN
fundo	297602.8	32449	229838.4	93386	435016.6
clavícula	0	0	0	0	0
crânio	237.4	7.6	976.8	39.4	4729
pés	606152.6	230705.2	252186.6	165166.2	630645.8
fêmur	107.2	112.6	7920	10884.6	15291
fíbula	3953.8	16.6	8876	832.6	10209
mãos	31594.8	2926.6	4124.2	128.2	28918.6
bacia	0	0	0	1.8	11.6
úmero	0.4	0	1.8	0.4	88.4
mandíbula	0	25.8	10.8	4	10.4
patela	0	174.8	31003.6	18354.6	13856
rádio	1103.8	0	22.4	0	392.6
costelas	0	0	0	0	3
sacro	0	0	0	0	0
escápula	0	0	0	0	28.2
esterno	0	0	0	0	0
tíbia	7805	2758.6	15660.8	5547.4	27357.2
ulna	4057.8	98.4	0	0	1808.4
vértebras	0	0	0	2	0.8

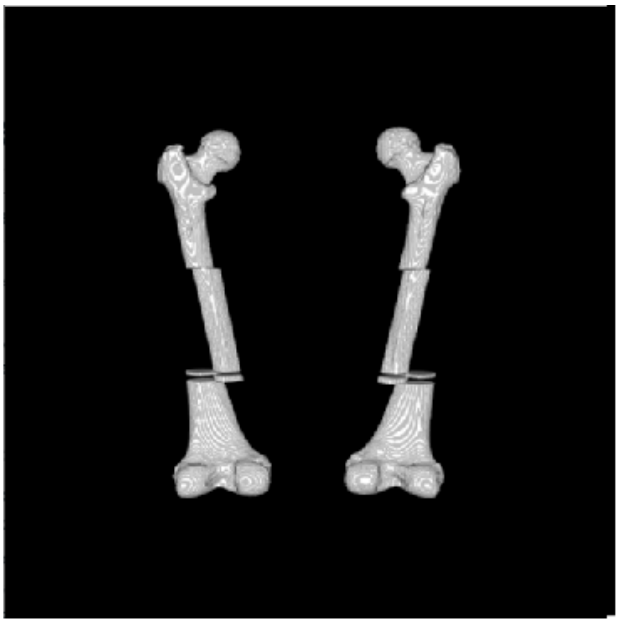
Resultado da segmentação das redes U-Net, DenseNet, ResNet, DeepLab e FCN juntamente com o *ground truth* para os pés femininos.



8.4 Fêmur

O osso do fêmur é o osso mais longo e volumoso do corpo tendo como principal função ajudar no suporte do corpo bem como na movimentação das pernas.

Fêmur - Corpo Masculino

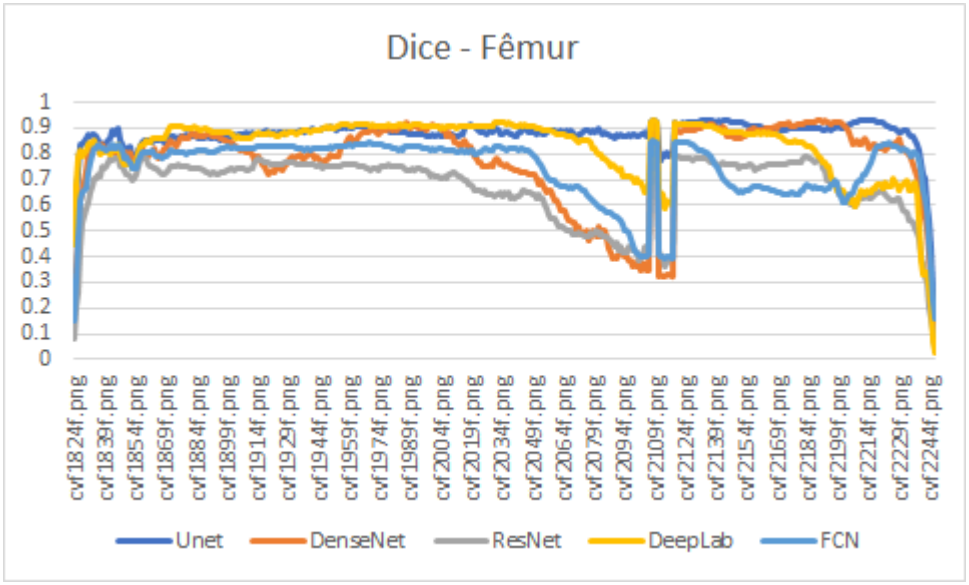


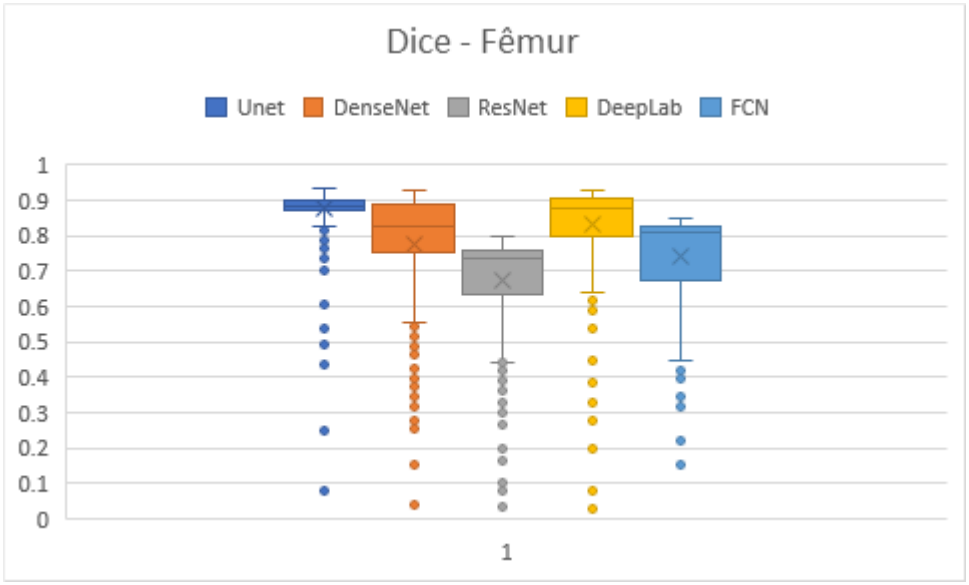
Fêmur - Corpo Feminino

Quadro estatístico contendo a contagem de tomografias, soma, média, variância e desvio padrão para os coeficientes Dice na segmentação do fêmur.

Dice					
Rede	*Contagem*	*Soma*	*Média*	*Variância*	*Desvio Padrão*
U-Net	423	370.583	0.8761	0.0048	∓0.0695
DenseNet	423	327.761	0.7748	0.0267	∓0.1631
ResNet	423	284.152	0.6718	0.0170	∓0.1304
DeepLab	423	350.997	0.8298	0.0150	∓0.1222
FCN	423	314.492	0.7435	0.0148	∓0.1216

Desempenho das redes *deep learning* U-Net, DenseNet, ResNet, DeepLab e FCN para os coeficientes Dice na segmentação do fêmur do corpo feminino.

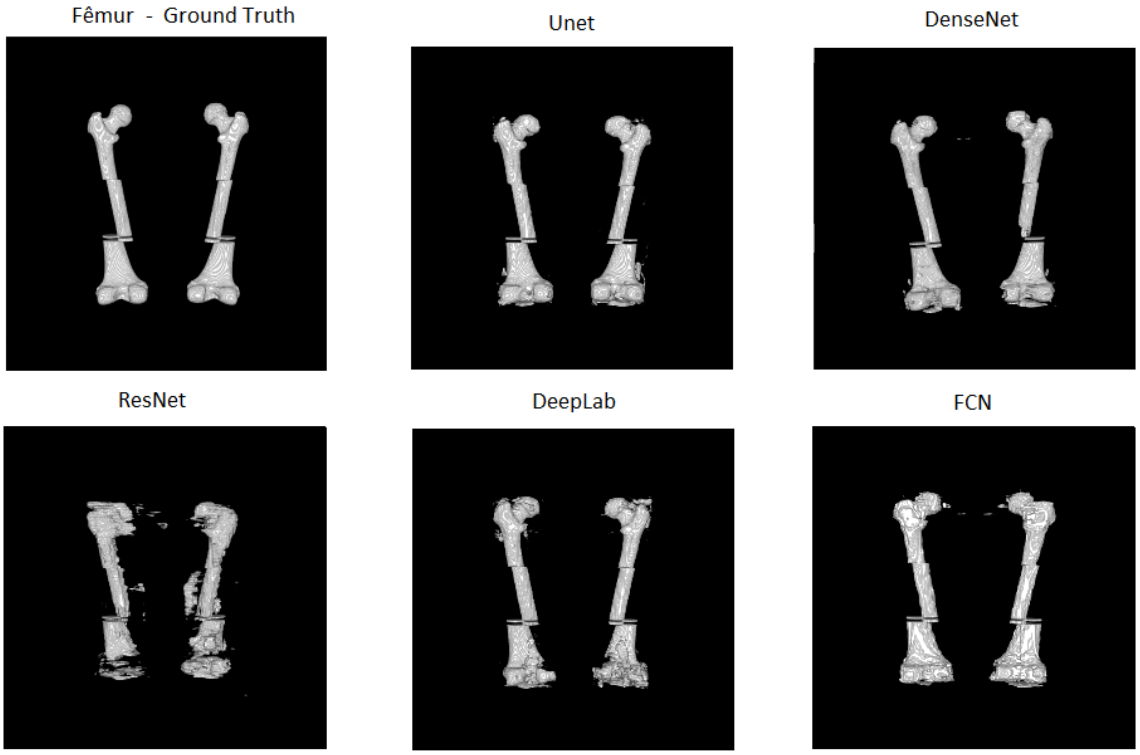




Quadro com as linhas da *matriz de confusão* para a segmentação do fêmur do corpo feminino do VHP.

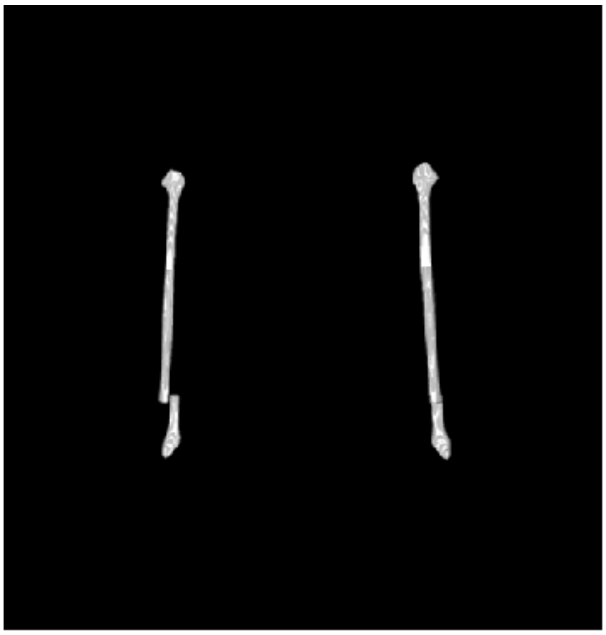
Matriz Confusão	U-Net	DenseNet	ResNet	DeepLab	FCN
fundo	398865.4	242881.4	982873.4	295198.4	484506
clavícula	101.6	3.4	5.8	0	42
crânio	4215.6	13745.8	11135.8	23	2254.6
pés	12293.8	24191.6	25561.2	8031.2	4763.4
fêmur	**1413359.2**	**1210547.6**	**1131308.4**	**891413.4**	**1205004.6**
fíbula	7595.6	35015	18209.6	5900.6	1694.4
mãos	15.2	0	130	1.4	0.4
bacia	854.6	605.8	1908.8	883.6	2549.8
úmero	1469.2	1031.4	89.4	0	5793.4
mandíbula	159.6	236.6	6875.4	0	260.2
patela	0	98	405.2	0.6	76.8
rádio	0	0	0	0	0
costelas	3.6	37.8	56.4	14.6	138.6
sacro	1	0	0	0	134.6
escápula	1213.8	1522.2	82.4	4.2	1058
esterno	0	0	0	0	0
tíbia	124342.4	140877.6	103986	81475.8	56398.4
ulna	49.4	0	0	0	0
vértebras	1564.4	2386.6	13125	1500	4183

Resultado da segmentação das redes U-Net, DenseNet, ResNet, DeepLab e FCN juntamente com o *ground truth* para o fêmur feminino.

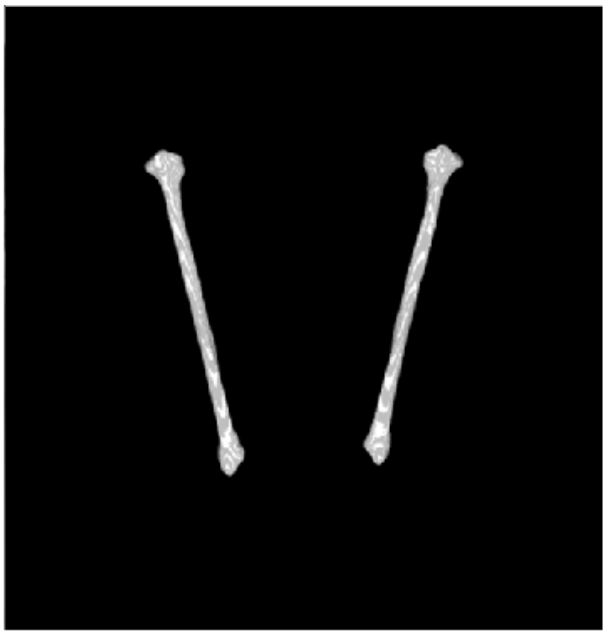


8.5 Fíbula

A fíbula, que se localiza ao lado da tíbia, é um osso da perna que tem como função a fixação dos músculos.



Fíbula - Corpo Masculino

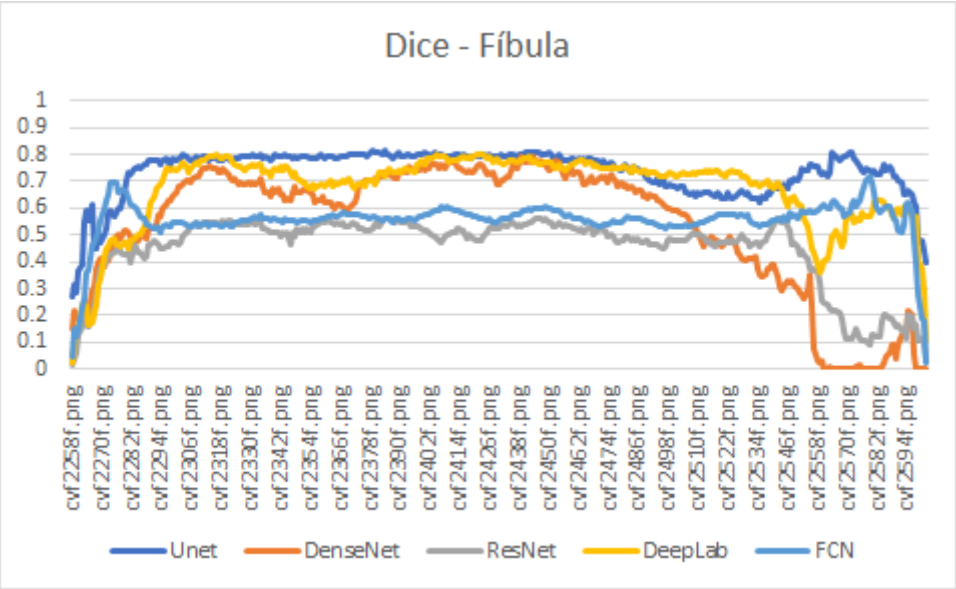


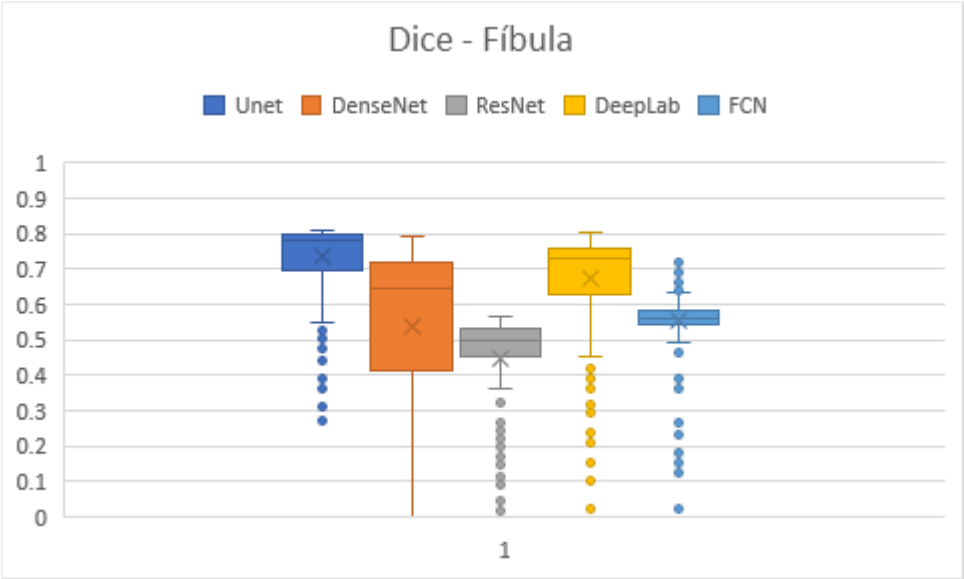
Fíbula - Corpo Feminino

Quadro estatístico contendo a contagem de tomografias, soma, média, variância e desvio padrão para os coeficientes Dice na segmentação da fíbula.

Dice					
Rede	*Contagem*	*Soma*	*Média*	*Variância*	*Desvio Padrão*
U-Net	345	254.201	0.7368	0.0084	∓0.0915
DenseNet	345	184.807	0.5357	0.0591	∓0.2427
ResNet	345	154.612	0.4482	0.0181	∓0.1343
DeepLab	345	232.111	0.6728	0.0203	∓0.1421
FCN	345	191.357	0.5547	0.0062	∓0.0789

Desempenho das redes *deep learning* U-Net, DenseNet, ResNet, DeepLab e FCN para os coeficientes Dice na segmentação da fíbula do corpo feminino.

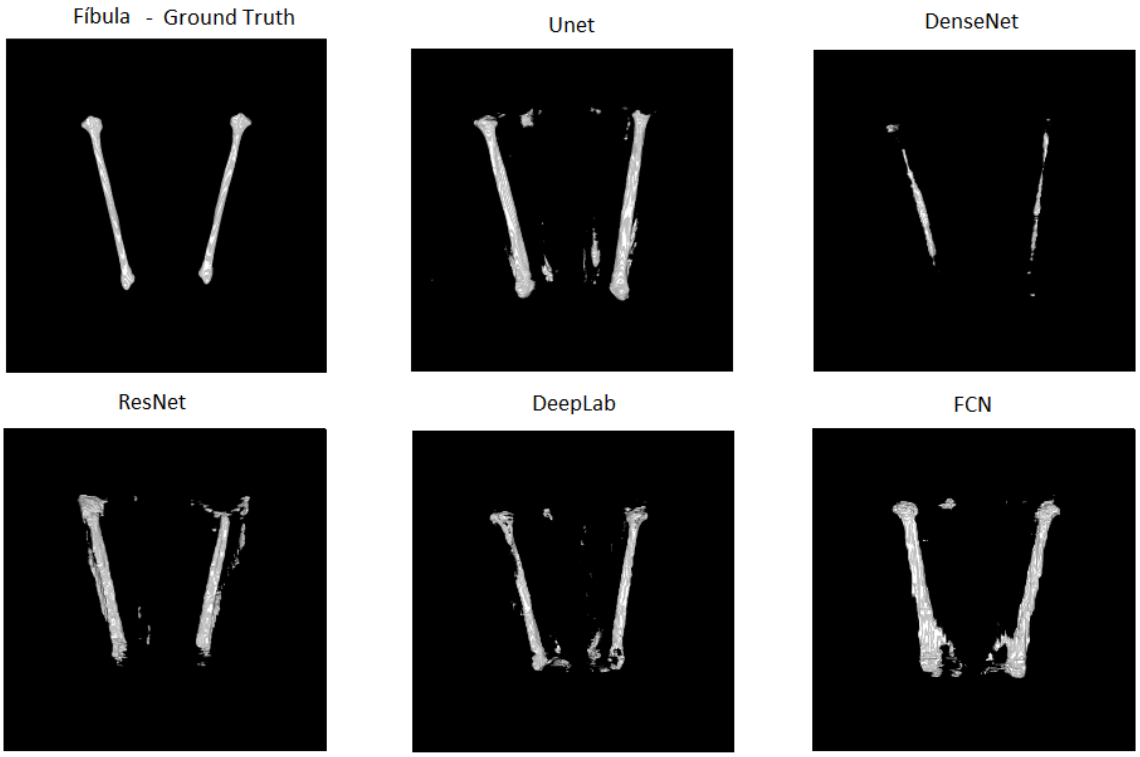




Quadro com as linhas da *matriz de confusão* para a segmentação da fíbula do corpo feminino do VHP.

Matriz Confusão	U-Net	DenseNet	ResNet	DeepLab	FCN
fundo	120523.8	35797.4	267145	80322.2	252961
clavícula	0.4	0	0	0	1.4
crânio	1438	0	1	1	489.8
pés	6192.2	1380.8	1126.2	4687.8	4055.6
fêmur	398.6	3532.8	514.2	1180.6	15959
fíbula	**167845.8**	**89380.2**	**112169.4**	**113931.2**	**168928**
mãos	462.6	0	29.6	0	477.8
bacia	0	0	0	0	2.6
úmero	5.6	0	0.6	0	111.2
mandíbula	19	0	0.8	0	10
patela	0	0	0	0	0
rádio	17501.4	0	112.8	550	11096.2
costelas	0	0	0	0	0
sacro	0	0	0	0	0
escápula	17.6	0	0	44	17.6
esterno	0	0	0	0	0
tíbia	3955.6	728.6	331.4	563.2	9435.6
ulna	141.6	0	11.4	0	0.4
vértebras	20.2	0	0	21.4	8

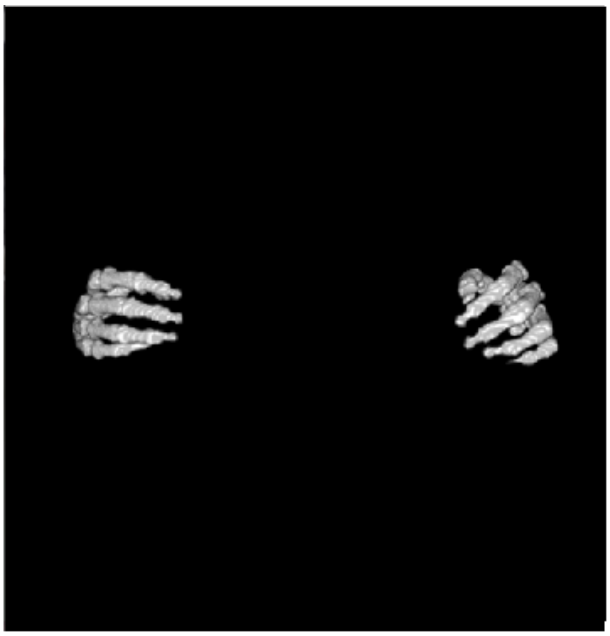
Resultado da segmentação das redes U-Net, DenseNet, ResNet, DeepLab e FCN juntamente com o *ground truth* para a fíbula feminina.



8.6 Mãos

As mãos são compostas de vários conjuntos de ossos agrupados em carpo, metacarpo e falanges, cuja função é ajudar na sua articulação.

Mãos - Corpo Masculino

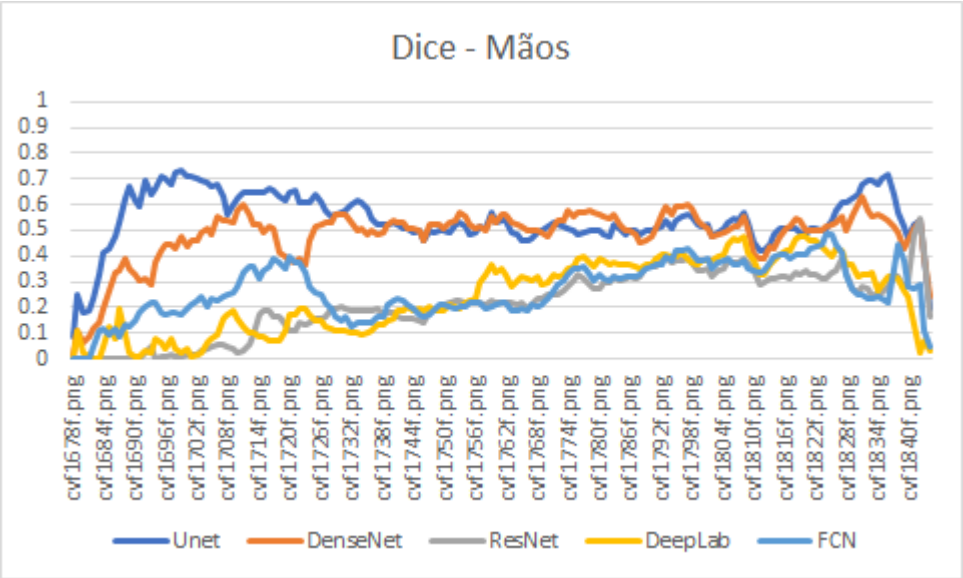


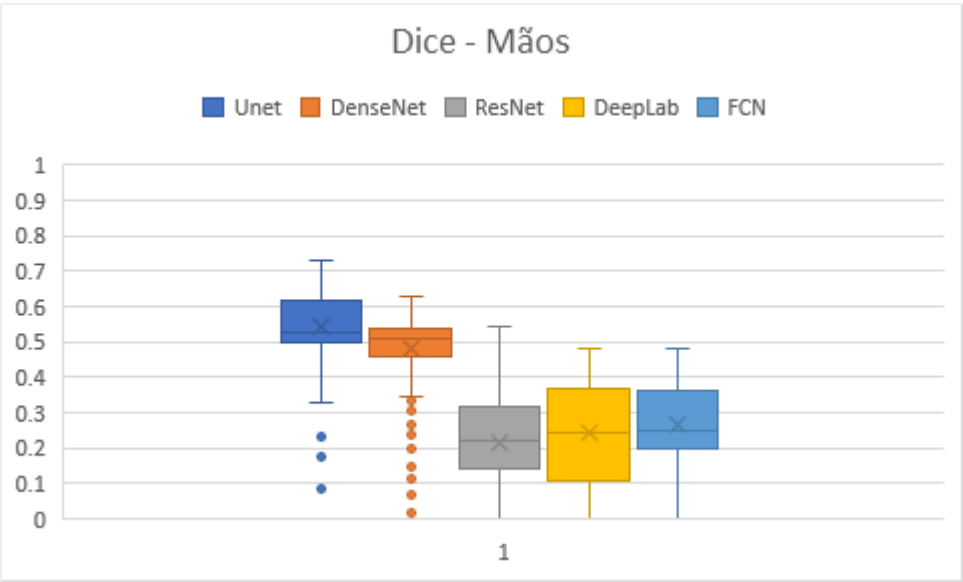
Mãos - Corpo Feminino

Quadro estatístico contendo a contagem de tomografias, soma, média, variância e desvio padrão para os coeficientes Dice na segmentação das mãos.

Dice					
Rede	*Contagem*	*Soma*	*Média*	*Variância*	*Desvio Padrão*
U-Net	167	90.366	0.5411	0.0109	∓0.1040
DenseNet	167	80.324	0.4810	0.0110	∓0.1046
ResNet	167	35.487	0.2125	0.0160	∓0.1260
DeepLab	167	40.149	0.2404	0.0204	∓0.1424
FCN	167	44.682	0.2676	0.0112	∓0.1057

Desempenho das redes *deep learning* U-Net, DenseNet, ResNet, DeepLab e FCN para os coeficientes Dice na segmentação das mãos do corpo feminino.

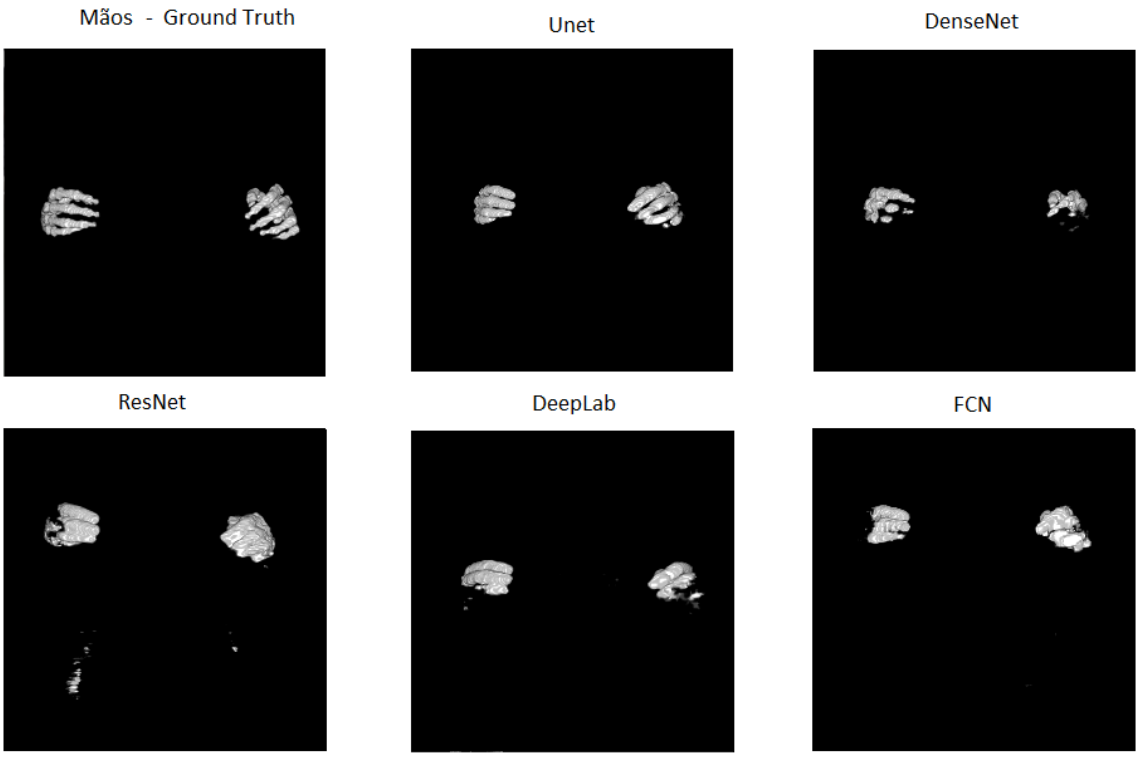




Quadro com as linhas da *matriz de confusão* para a segmentação das mãos do corpo feminino do VHP.

Matriz Confusão	U-Net	DenseNet	ResNet	DeepLab	FCN
fundo	105367.2	68979.6	233566.6	37168.4	107001.8
clavícula	0	0	0	0	0
crânio	52.2	1269.8	81571.6	0.2	3726.2
pés	27847.2	21916.8	79053.2	152.4	15986
fêmur	0	40.8	2757.4	26.6	58.6
fíbula	0	0	147.8	0	0
mãos	**105670.6**	**85858.4**	**40986.6**	**25207.4**	**42659.2**
bacia	0	0	0	0	17.2
úmero	36	0	873.4	0.4	0.4
mandíbula	3.8	3.6	84.8	3.2	19.6
patela	0	3986.8	9033.8	630.6	524.4
rádio	447.6	153.6	290	0	277.4
costelas	0.4	0.6	1	0	0.2
sacro	0	0	0	0	0
escápula	0	0	87.6	0	0.2
esterno	0	0	0	0	0
tíbia	12.8	254.4	5393.8	5.2	1277.8
ulna	3602.2	6228	546.8	15.2	1002
vértebras	0	48.4	30	0	0

Resultado da segmentação das redes U-Net, DenseNet, ResNet, DeepLab e FCN juntamente com o *ground truth* para as mãos femininas.



8.7 Bacia

Os ossos da bacia, ou pélvis, encontrados na cintura, possuem as funções tanto de dar sustentação ao corpo como também movimentação e manutenção do equilíbrio.

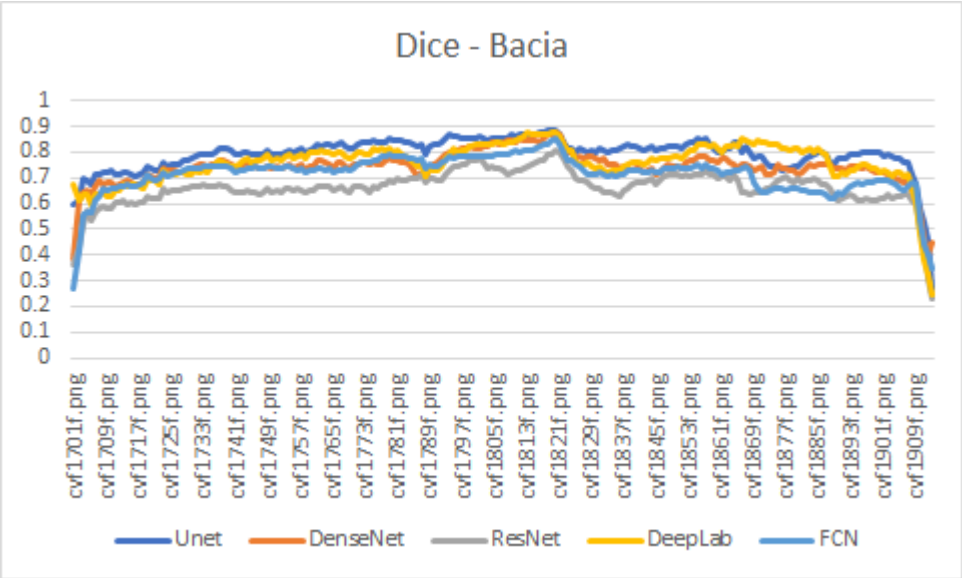
Bacia - Corpo Masculino

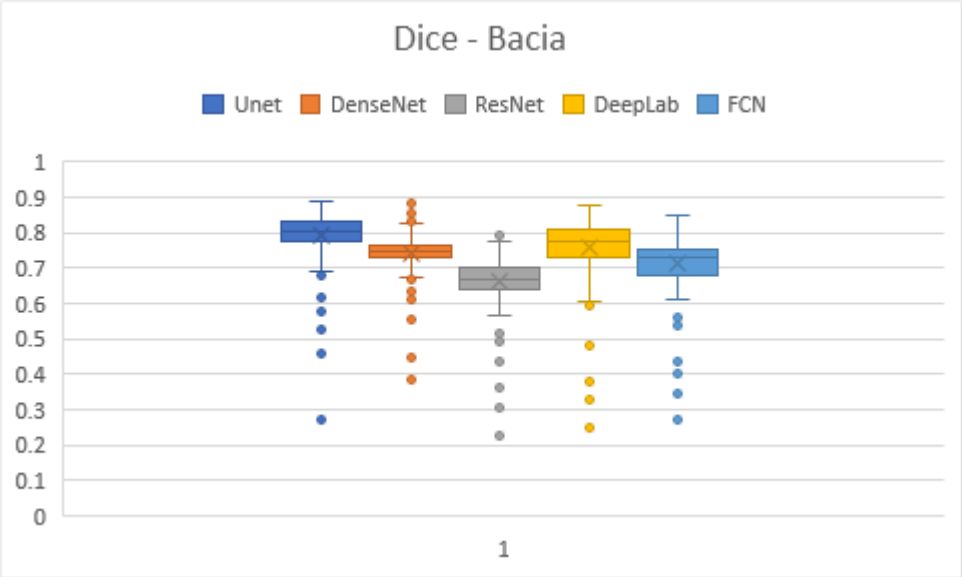
Bacia - Corpo Feminino

Quadro estatístico contendo a contagem de tomografias, soma, média, variância e desvio padrão para os coeficientes Dice na segmentação da bacia.

Dice					
Rede	*Contagem*	*Soma*	*Média*	*Variância*	*Desvio Padrão*
U-Net	213	169.128	0.7940	0.0047	∓0.0685
DenseNet	213	158.324	0.7433	0.0043	∓0.0653
ResNet	213	141.395	0.6638	0.0049	∓0.0700
DeepLab	213	162.072	0.7609	0.0068	∓0.0820
FCN	213	152.291	0.7150	0.0058	∓0.0760

Desempenho das redes *deep learning* U-Net, DenseNet, ResNet, DeepLab e FCN para os coeficientes Dice na segmentação da bacia do corpo feminino.

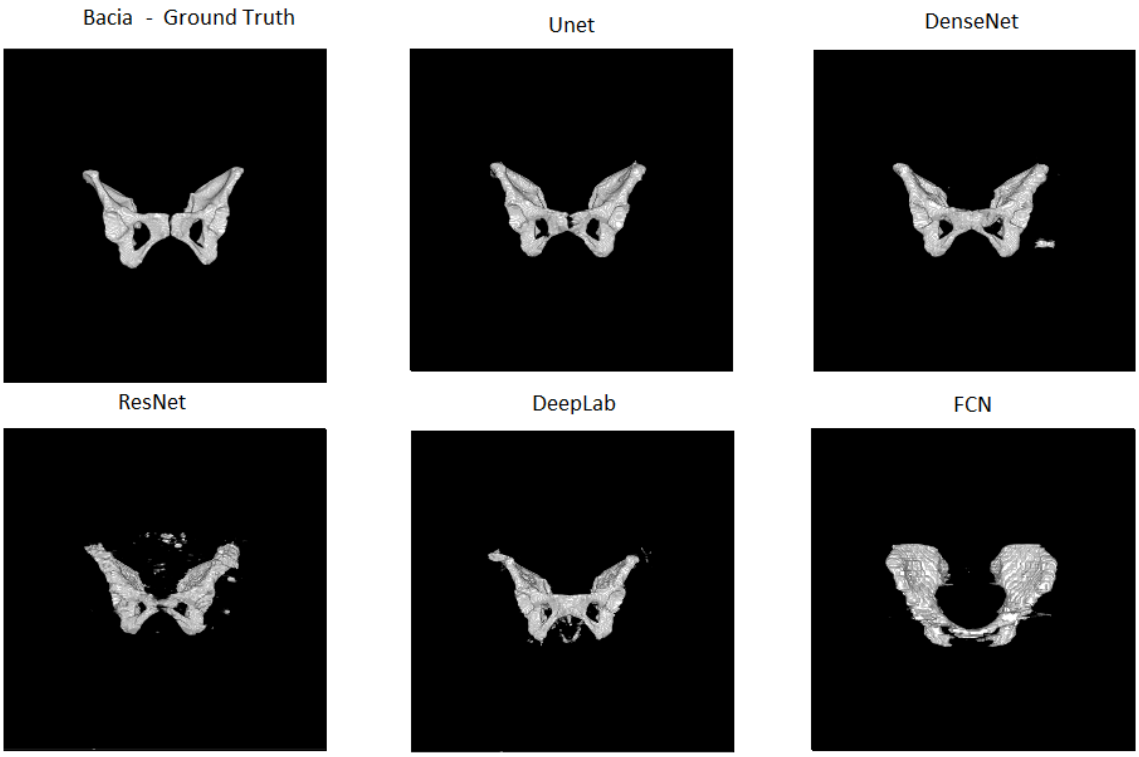




Quadro com as linhas da *matriz de confusão* para a segmentação da bacia do corpo feminino do VHP.

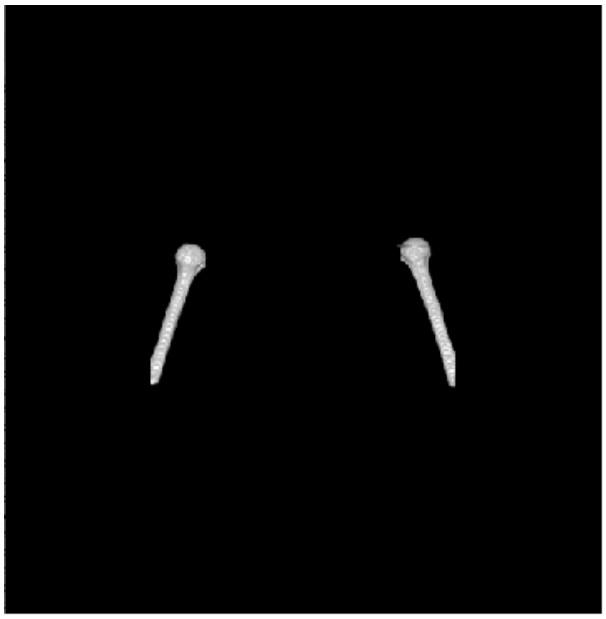
Matriz Confusão	U-Net	DenseNet	ResNet	DeepLab	FCN
fundo	362002.2	508165.6	821927.8	421105.2	521937.8
clavícula	100.6	1007	65	0	58.4
crânio	7005.8	13309.8	10317.6	0	8731.2
pés	17969.2	24686	14056.8	5189.8	959.2
fêmur	2823.4	52779.2	15366.6	1849.8	12454.2
fíbula	436.8	4178.6	8285	1562.4	0.8
mãos	159.2	0	1189.2	6117.2	3637
bacia	**764579.8**	**763931.6**	**742057.6**	**622999**	**752293.2**
úmero	809.4	8495.6	533.8	0	1383
mandíbula	2498.6	5445.6	3836	0	2656.8
patela	0	0	167	0	0
rádio	0	26.2	0	8.6	3.4
costelas	1080.4	2988.6	5548.4	33.8	481
sacro	4313.6	4256.4	1084.6	903.4	2364.6
escápula	8638	13598.4	3198.4	155	5274.4
esterno	0	0	0.2	0	0
tíbia	1665.4	678.6	1322	18575.4	1423.8
ulna	14.8	0	0	4.6	1
vértebras	9467.4	17446.6	21827.6	2616.2	9279.4

Resultado da segmentação das redes U-Net, DenseNet, ResNet, DeepLab e FCN juntamente com o *ground truth* para a bacia feminina.

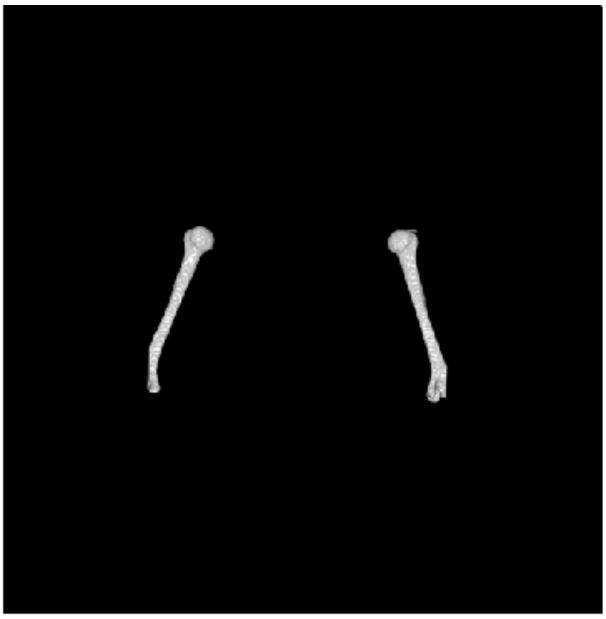


8.8 Úmero

Os ossos do úmero são os mais longos da parte dos membros superiores e têm a função de dar sustentabilidade ao antebraço e de ligar o cotovelo ao ombro.



Úmero - Corpo Masculino

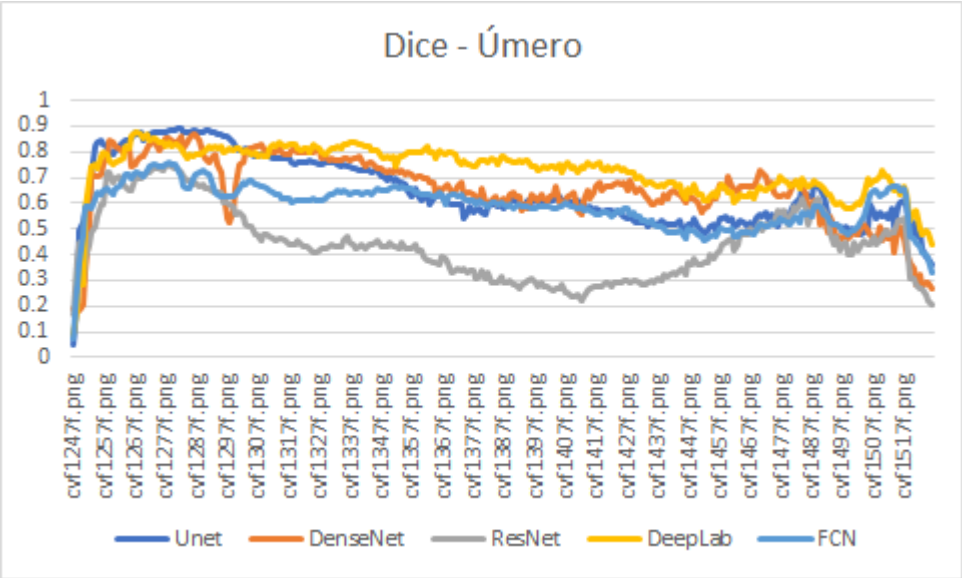


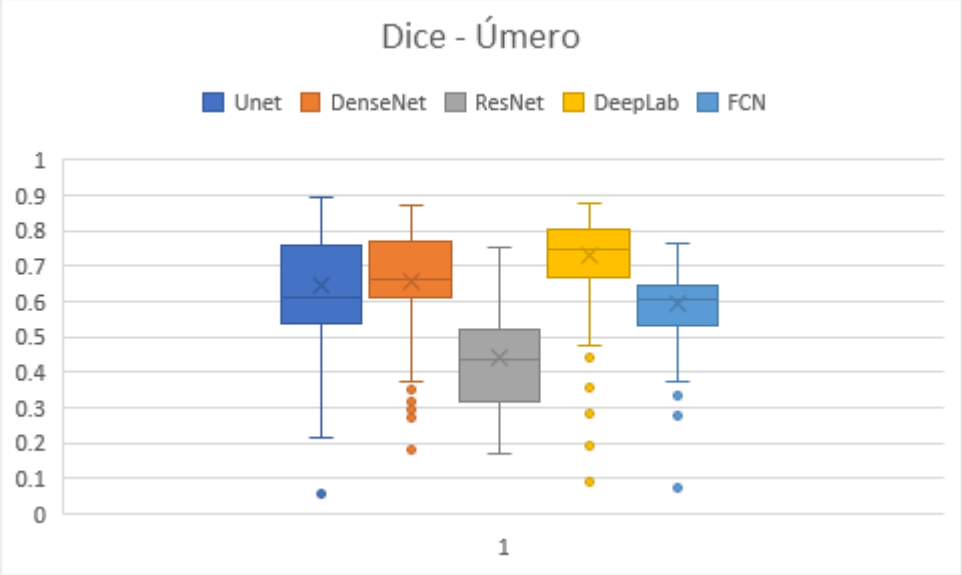
Úmero - Corpo Feminino

Quadro estatístico contendo a contagem de tomografias, soma, média, variância e desvio padrão para os coeficientes Dice na segmentação do úmero.

Dice					
Rede	*Contagem*	*Soma*	*Média*	*Variância*	*Desvio Padrão*
U-Net	280	180.672	0.6453	0.0185	∓0.1358
DenseNet	280	184.103	0.6575	0.0177	∓0.1328
ResNet	280	123.922	0.4426	0.0190	∓0.1376
DeepLab	280	203.987	0.7285	0.0104	∓0.1020
FCN	280	165.543	0.5912	0.0074	∓0.0856

Desempenho das redes *deep learning* U-Net, DenseNet, ResNet, DeepLab e FCN para os coeficientes Dice na segmentação do úmero do corpo feminino.

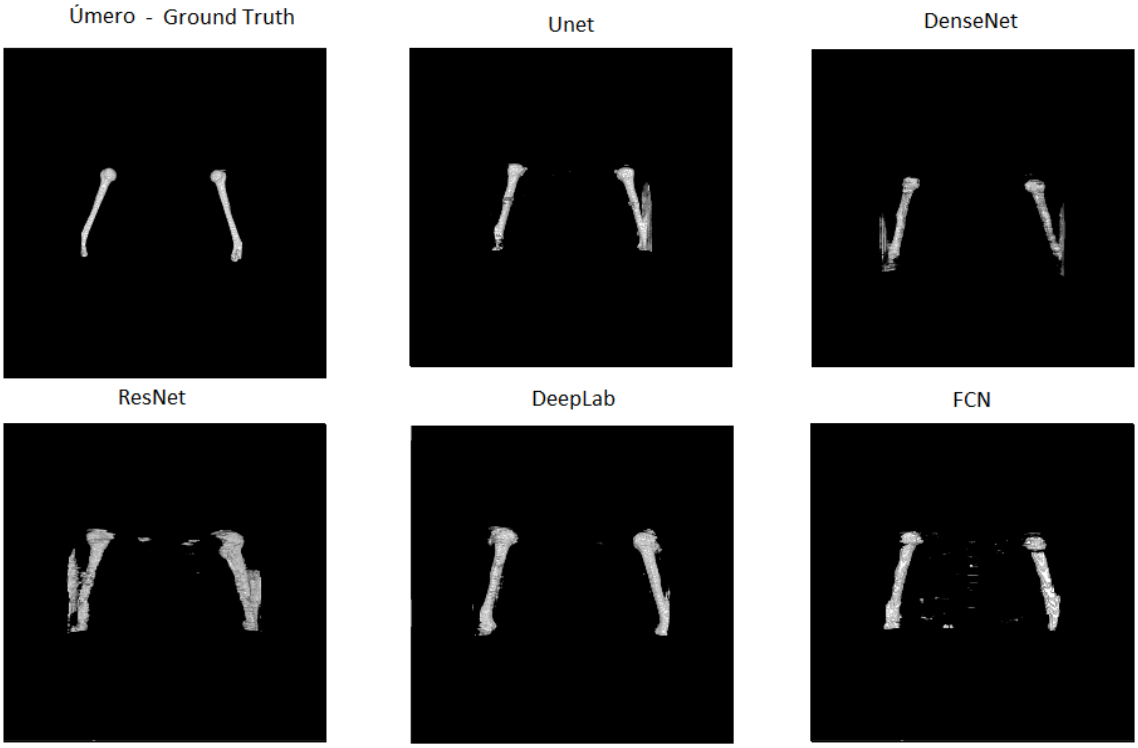




Quadro com as linhas da *matriz de confusão* para a segmentação do úmero do corpo feminino do VHP.

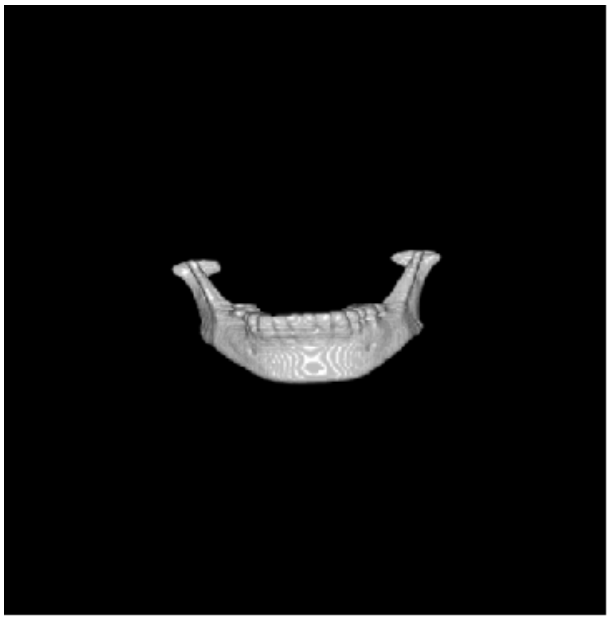
Matriz Confusão	U-Net	DenseNet	ResNet	DeepLab	FCN
fundo	324664.8	193270	729139.4	215399.4	404693
clavícula	1807.8	851.8	4.6	0	1306.2
crânio	26.4	12.8	537.4	0	539.2
pés	3117	6629	2771.2	0.2	50.6
fêmur	17050.8	9973.6	1872.8	17818.6	17824
fíbula	478.6	178.8	5.2	114.2	603.2
mãos	145	381.2	284.2	0	825
bacia	773	10.6	12.2	88.6	2052
úmero	**297910**	**245862.2**	**276318.6**	**241570.6**	**296097**
mandíbula	0	0	433.8	0	526
patela	0	0	33	25	7
rádio	14177.8	9209.2	14956.2	15306	18451.4
costelas	257.8	49.2	635.8	104.6	262
sacro	3	0	0	0	0
escápula	6417.4	1546.6	2704	11.4	4332.2
esterno	0	0	0	0	0
tíbia	538	12130.4	55.8	4328.4	10535.8
ulna	8284	5461.4	6757.4	3159	7286.8
vértebras	22.6	0	0	0	2016.6

Resultado da segmentação das redes U-Net, DenseNet, ResNet, DeepLab e FCN juntamente com o *ground truth* para o úmero.

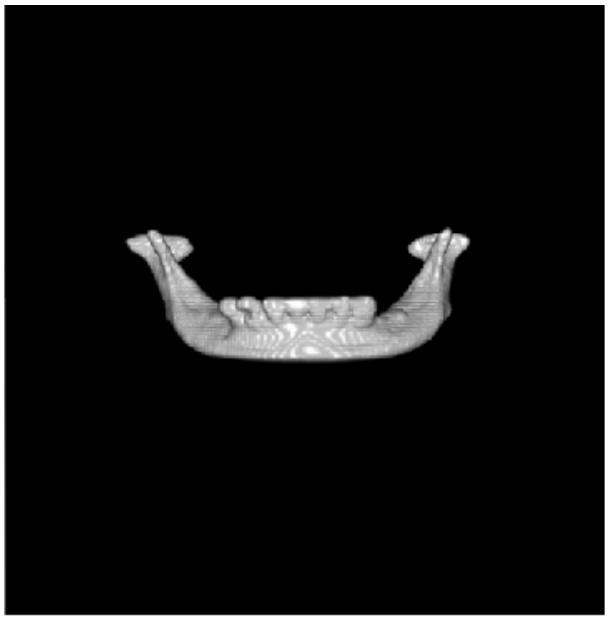


8.9 Mandíbula

A mandíbula é um osso do crânio tendo como função ajudar na mastigação, trituração e deglutição de alimentos.



Mandíbula - Corpo Masculino

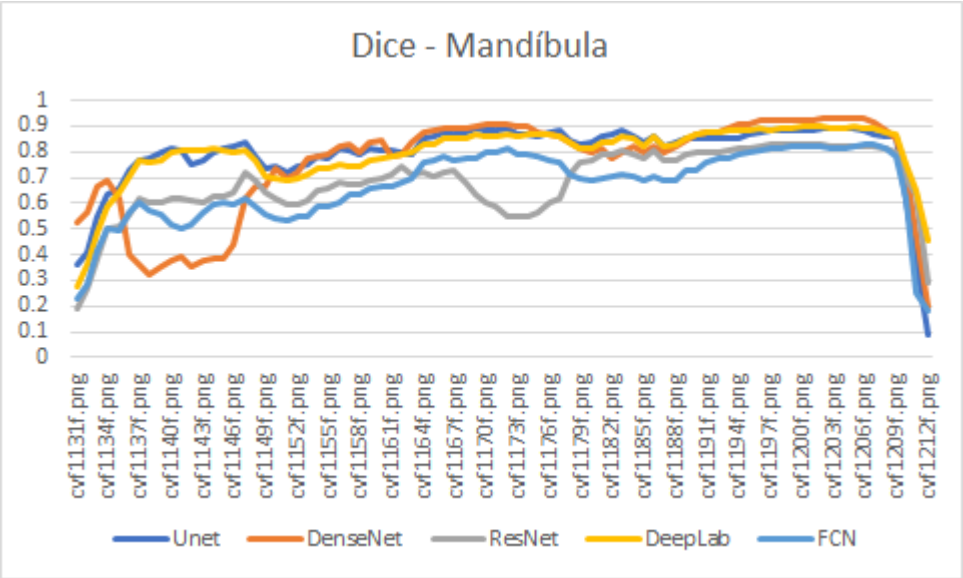


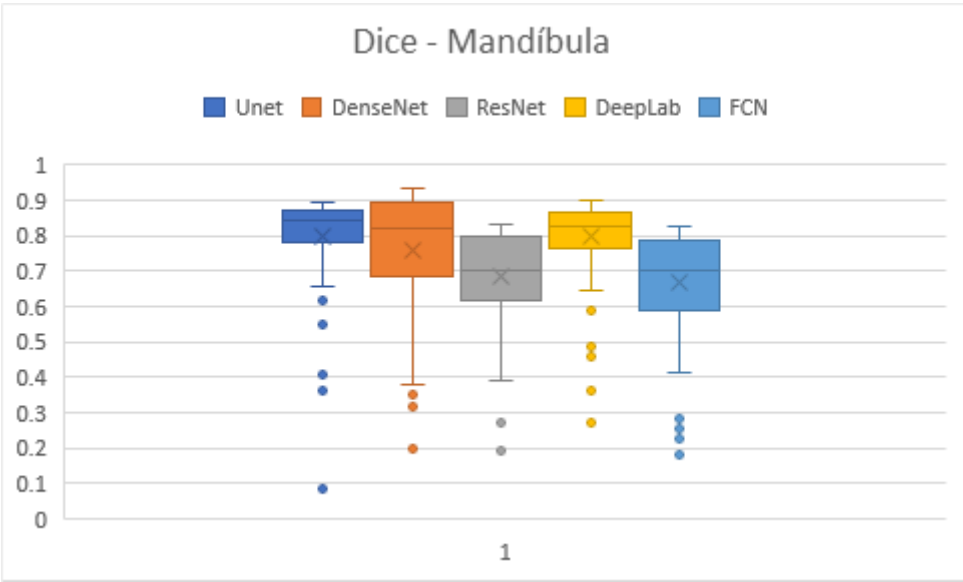
Mandíbula - Corpo Feminino

Quadro estatístico contendo a contagem de tomografias, soma, média, variância e desvio padrão para os coeficientes Dice na segmentação da mandíbula.

Dice					
Rede	*Contagem*	*Soma*	*Média*	*Variância*	*Desvio Padrão*
U-Net	82	65.529	0.7991	0.0178	∓0.1325
DenseNet	82	61.983	0.7559	0.0362	∓0.1891
ResNet	82	56.127	0.6845	0.0172	∓0.1303
DeepLab	82	65.282	0.7961	0.0134	∓0.1148
FCN	82	54.902	0.6695	0.0207	∓0.1429

Desempenho das redes *deep learning* U-Net, DenseNet, ResNet, DeepLab e FCN para os coeficientes Dice na segmentação da mandíbula do corpo feminino.

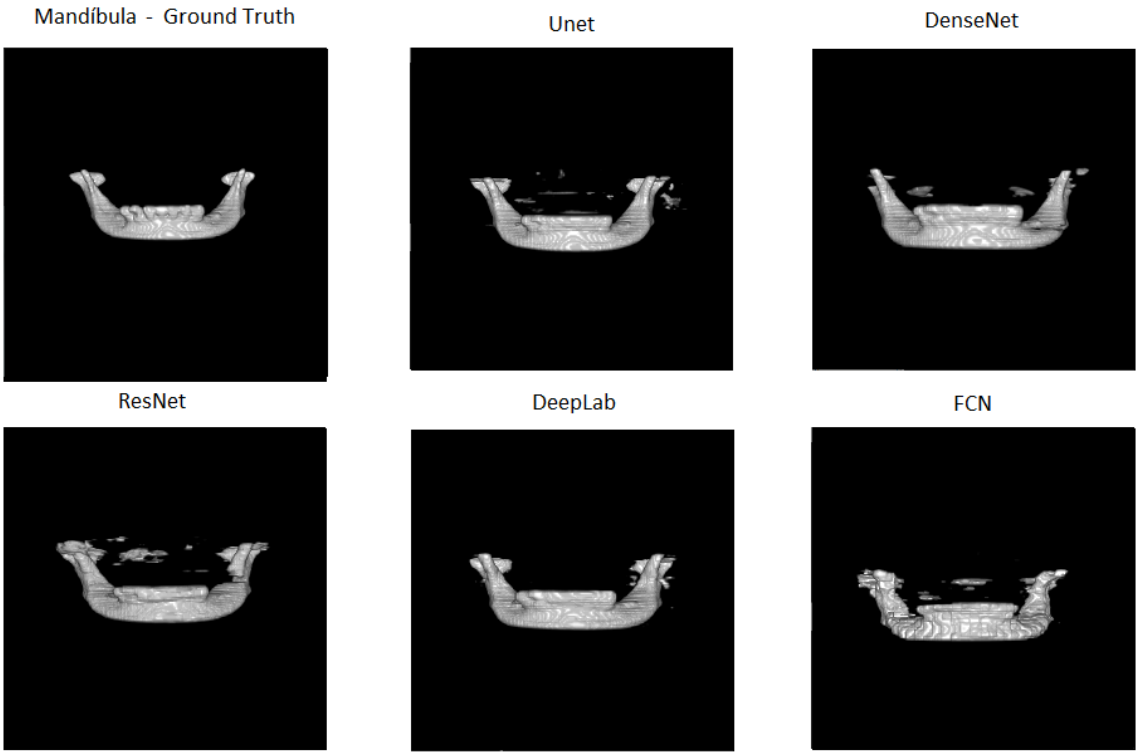




Quadro com as linhas da *matriz de confusão* para a segmentação da mandíbula do corpo feminino do VHP.

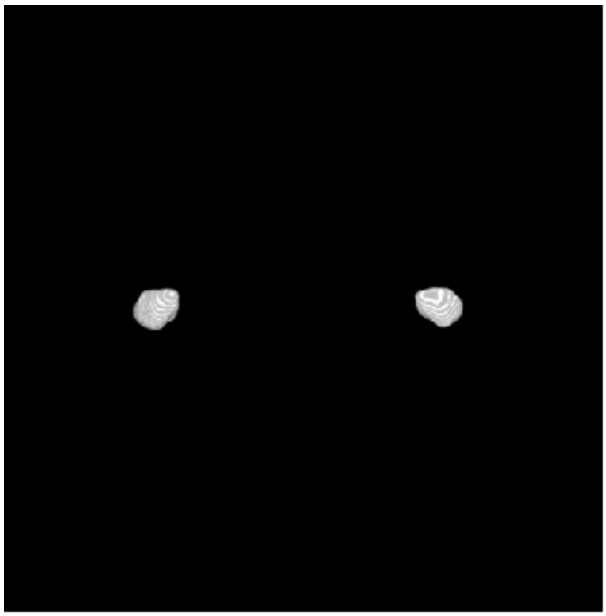
Matriz Confusão	U-Net	DenseNet	ResNet	DeepLab	FCN
fundo	92414.2	38467.4	160763	75166.2	150774
clavícula	0	0	0	0	0
crânio	51470.8	19637.8	12186.6	2292.6	31412.2
pés	114.8	1922.8	3092	10639.6	6997.8
fêmur	431.4	650.4	17095.2	0	16741.8
fíbula	398.4	43.8	6067	0	135
mãos	0	0	3.6	124	231.8
bacia	27	18	0	0	1501.6
úmero	0	0	15.8	0	1419.4
mandíbula	**204490.4**	**180247.4**	**198638.8**	**170318.2**	**193893.2**
patela	0	4.6	1.4	0	2493.6
rádio	0	12.4	18	0	0
costelas	0	0.2	215.4	0	14.4
sacro	0	0	0	0	0
escápula	0	0	0	0	3.6
esterno	0	0	0	0	0
tíbia	352.4	7030.4	1244.4	347	11.2
ulna	0	0	0	0	0
vértebras	108.8	432.2	52.2	5.8	152.6

Resultado da segmentação das redes U-Net, DenseNet, ResNet, DeepLab e FCN juntamente com o *ground truth* para a mandíbula feminina.

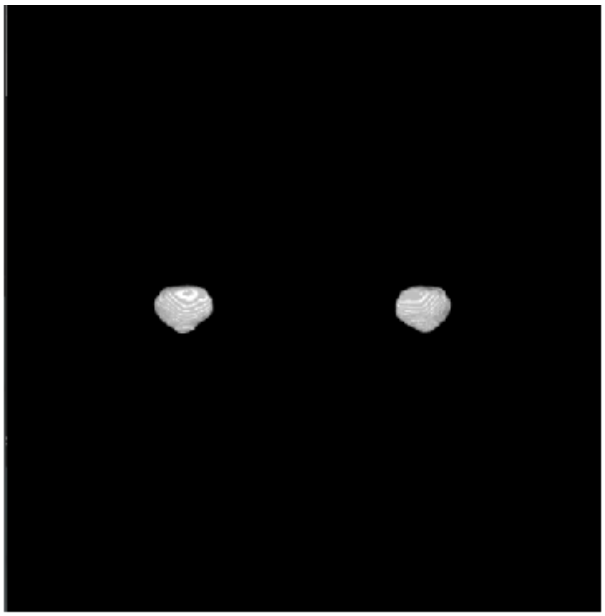


8.10 Patela

A patela, ou rótula, é um osso pequeno cuja finalidade é conectar os músculos do quadril e da coxa com a perna.



Patela - Corpo Masculino

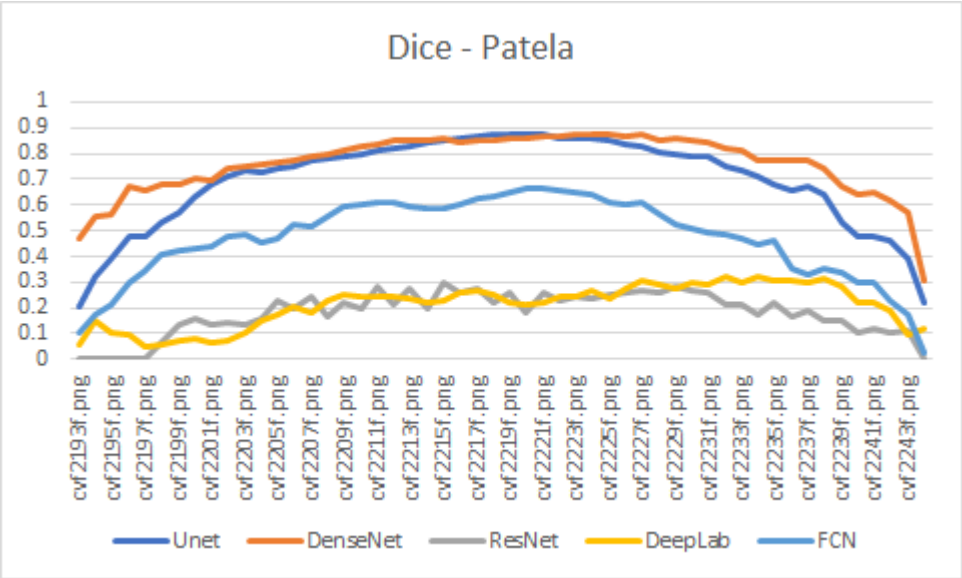


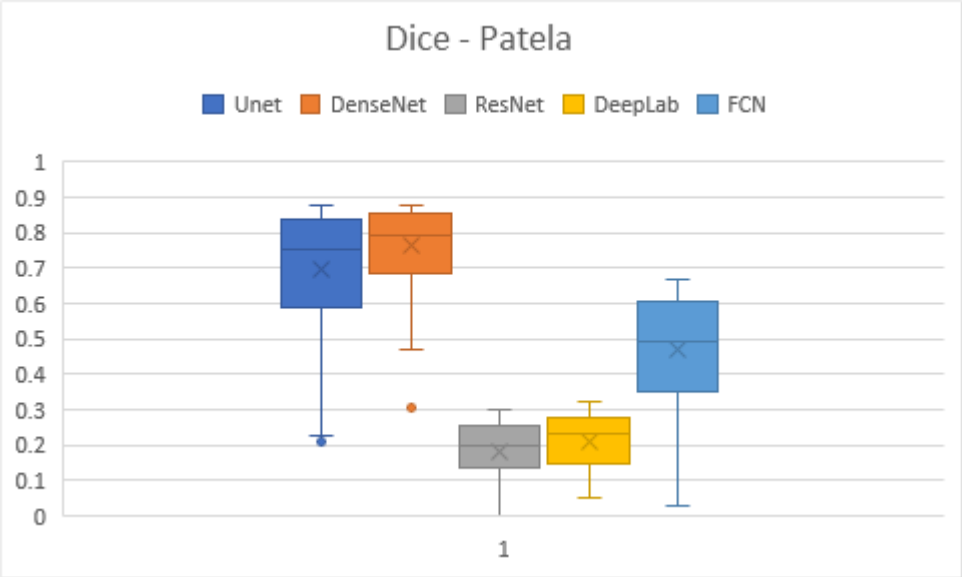
Patela - Corpo Feminino

Quadro estatístico contendo a contagem de tomografias, soma, média, variância e desvio padrão para os coeficientes Dice na segmentação da patela.

Dice					
Rede	*Contagem*	*Soma*	*Média*	*Variância*	*Desvio Padrão*
U-Net	52	36.147	0.6951	0.0315	∓0.1758
DenseNet	52	39.637	0.7623	0.0141	∓0.1176
ResNet	52	9.349	0.1798	0.0073	∓0.0845
DeepLab	52	10.949	0.2106	0.0069	∓0.0823
FCN	52	24.461	0.4704	0.0251	∓0.1568

Desempenho das redes *deep learning* U-Net, DenseNet, ResNet, DeepLab e FCN para os coeficientes Dice na segmentação da patela do corpo feminino.

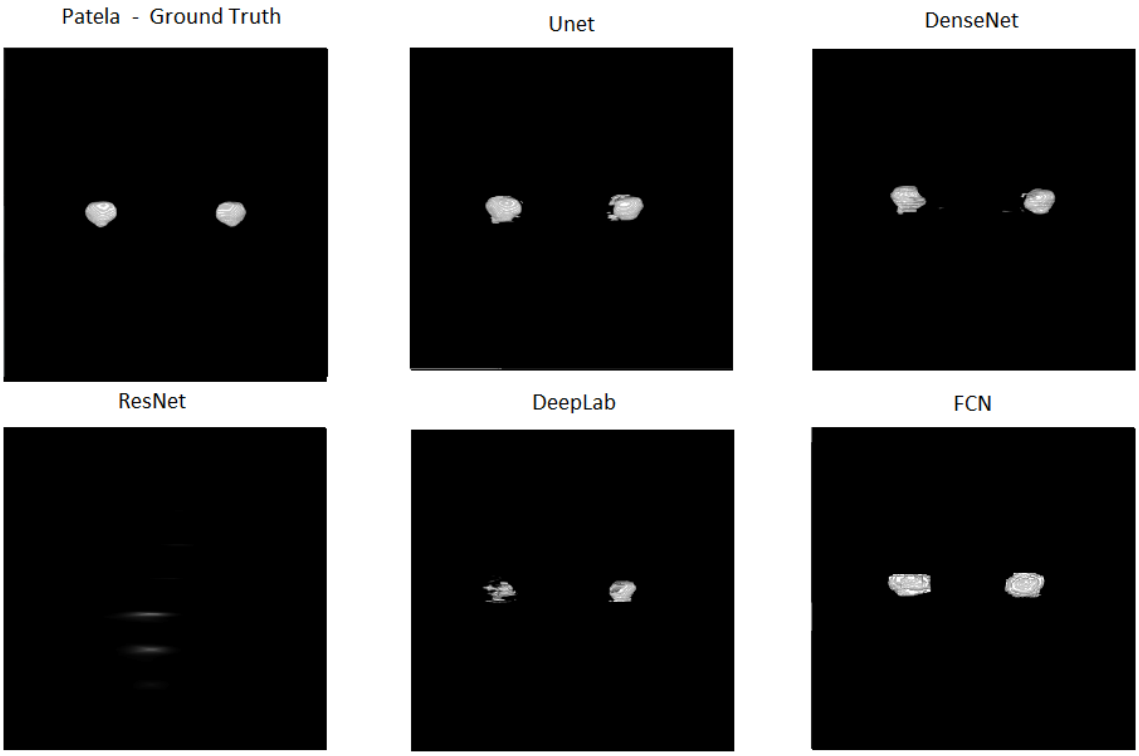




Quadro com as linhas da *matriz de confusão* para a segmentação da patela do corpo feminino do VHP.

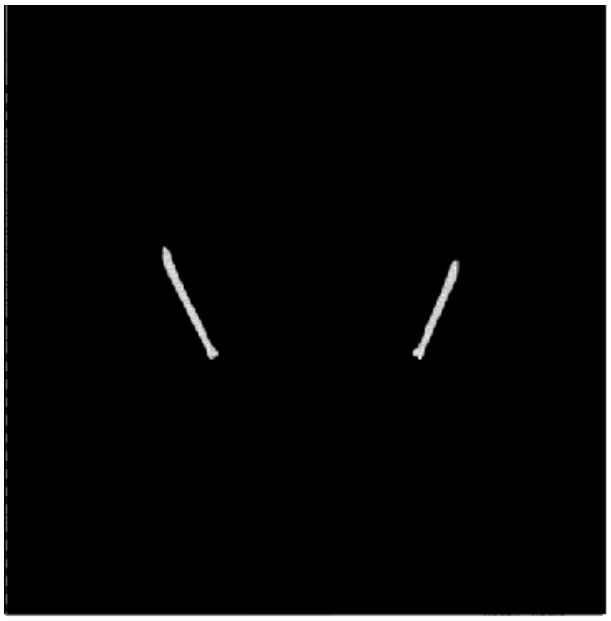
Matriz Confusão	U-Net	DenseNet	ResNet	DeepLab	FCN
fundo	92403.2	47085.8	19676	2203	88180.6
clavícula	0	0	0	0	0
crânio	9	0	21.6	0	1319.8
pés	29959.4	57094.6	1616.6	344	7831
fêmur	13.4	321.6	90.2	0	6300.4
fíbula	0	68	4.6	0	0
mãos	39.2	0	53.4	67	4864.4
bacia	0	0	0	0	21.8
úmero	0	0	0	0	0
mandíbula	3.4	16.6	7	0	0
patela	**88393.6**	**78654.8**	**13372**	**6551**	**60425.2**
rádio	0	0	0	0	0.2
costelas	0	0	0	0	17.8
sacro	0	0	0	0	0
escápula	0	0	0	0	0
esterno	0	0	0	0	0
tíbia	534.4	2516	627.6	0	446.4
ulna	0	0	8	0	376
vértebras	0	86.4	0	0	0

Resultado da segmentação das redes U-Net, DenseNet, ResNet, DeepLab e FCN juntamente com o *ground truth* para a patela feminina.

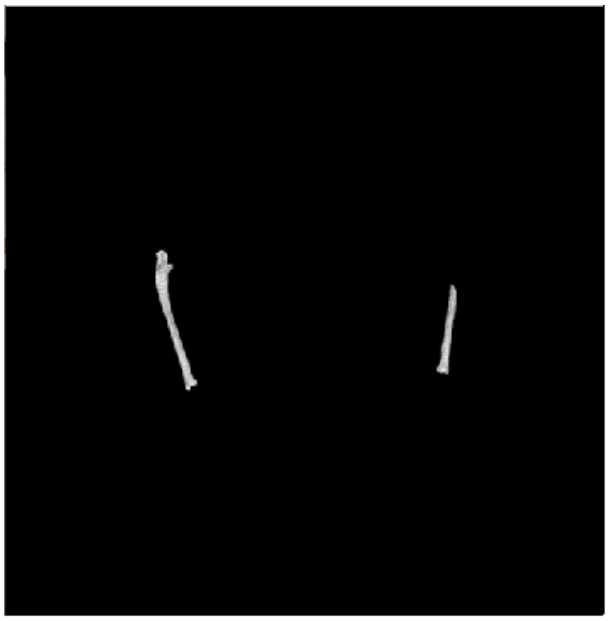


8.11 Rádio

O rádio é um osso do antebraço com a função de ajudar na articulação juntamente com a ulna.



Rádio - Corpo Masculino

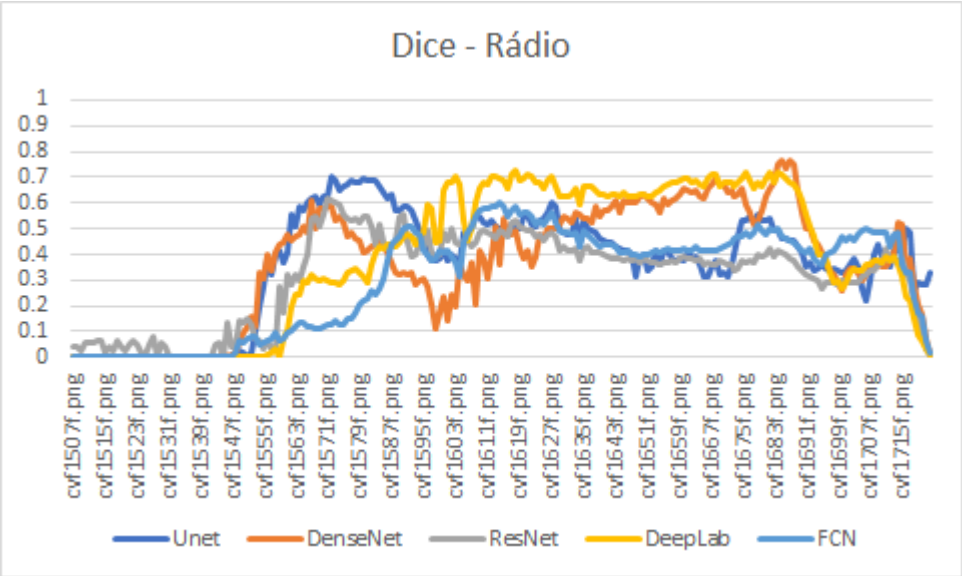


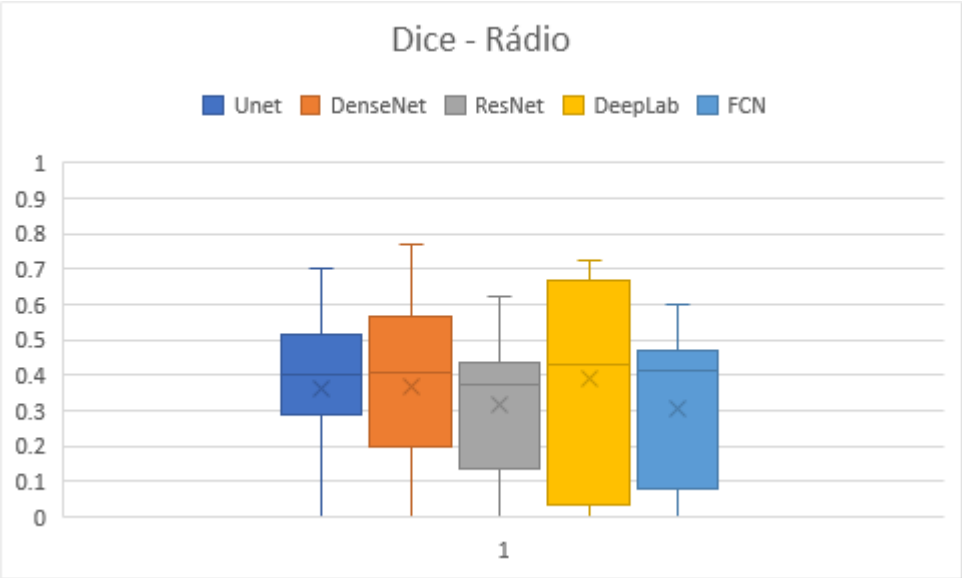
Rádio - Corpo Feminino

Quadro estatístico contendo a contagem de tomografias, soma, média, variância e desvio padrão para os coeficientes Dice na segmentação do rádio.

Dice					
Rede	*Contagem*	*Soma*	*Média*	*Variância*	*Desvio Padrão*
U-Net	216	78.202	0.3620	0.0458	∓0.2134
DenseNet	216	79.923	0.3700	0.0531	∓0.2300
ResNet	216	68.029	0.3149	0.0309	∓0.1754
DeepLab	216	84.805	0.3926	0.0753	∓0.2738
FCN	216	66.052	0.3058	0.0416	∓0.2036

Desempenho das redes *deep learning* U-Net, DenseNet, ResNet, DeepLab e FCN para os coeficientes Dice na segmentação do rádio do corpo feminino.

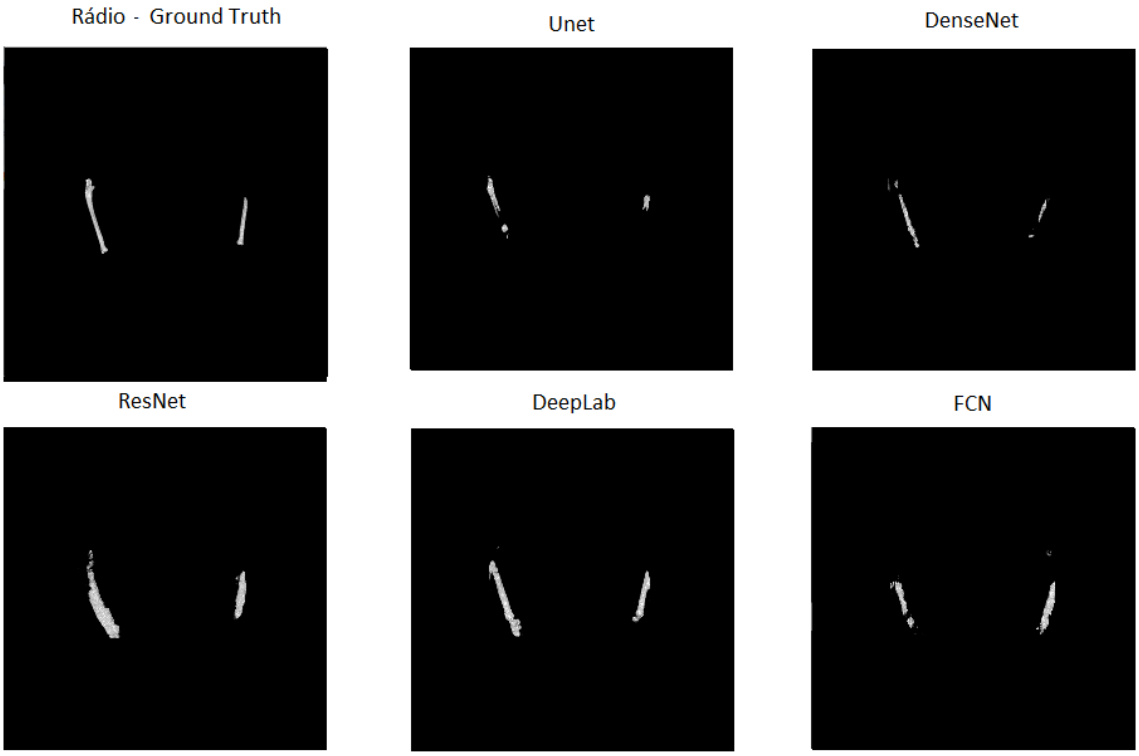




Quadro com as linhas da *matriz de confusão* para a segmentação do rádio do corpo feminino do VHP.

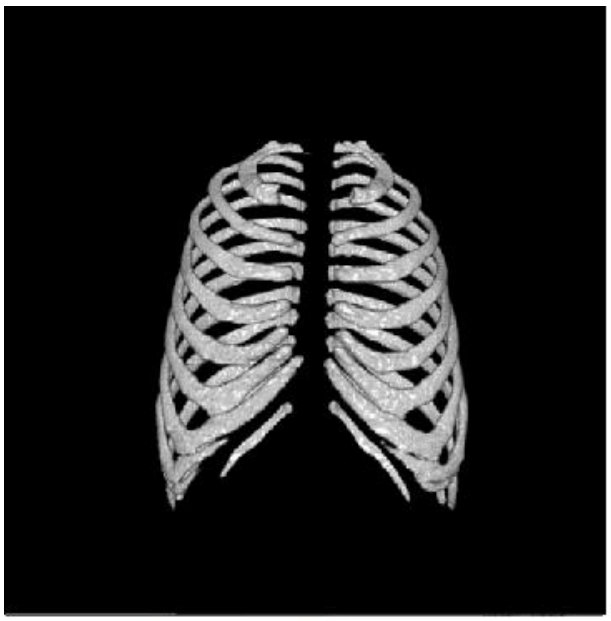
Matriz Confusão	U-Net	DenseNet	ResNet	DeepLab	FCN
fundo	35720.8	21457	173118.8	79535	88809.6
clavícula	0	0	0	0	0
crânio	0	0	27.2	0.2	3.4
pés	2068.2	5404.6	2691.4	92.4	2622.6
fêmur	0	0	2.8	27	120.8
fíbula	25.6	1.8	339	128.6	573.8
mãos	4262.8	8183.6	30629.8	43446.2	17144.8
bacia	0	0	0	0	0
úmero	0	0.2	369.6	0	43
mandíbula	0	0	0	0	0
patela	0	0	17.6	0	0
rádio	**28371**	**27443.4**	**44551.6**	**32246**	**34548.2**
costelas	0	0	0	0	0
sacro	0	0	0	0	0
escápula	0	0	0	0	0
esterno	0	0	0	0	0
tíbia	0	0	14.8	0	40
ulna	234.8	736.6	736.8	110.4	1089.8
vértebras	0	0	0	0	0

Resultado da segmentação das redes U-Net, DenseNet, ResNet, DeepLab e FCN juntamente com o *ground truth* para o rádio feminino.

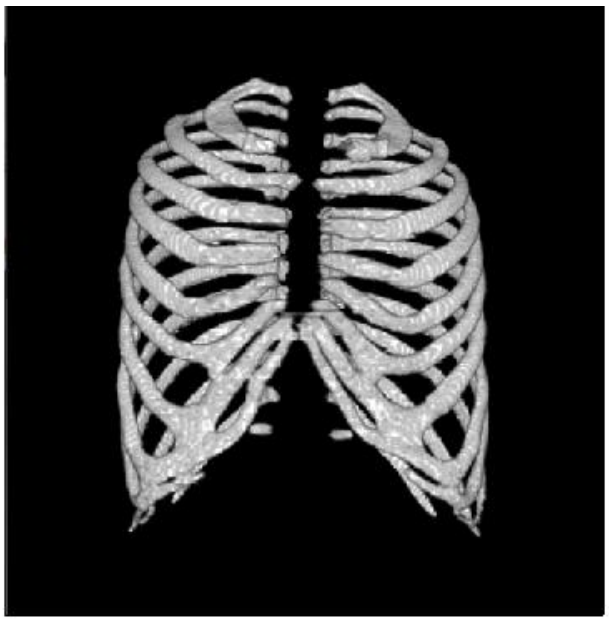


8.12 Costelas

As costelas são ossos alongados em forma de arco com a funções de sustentação do corpo humano, além de fazer a proteção dos órgãos localizados no tórax, como coração e pulmões.



Costelas - Corpo Masculino

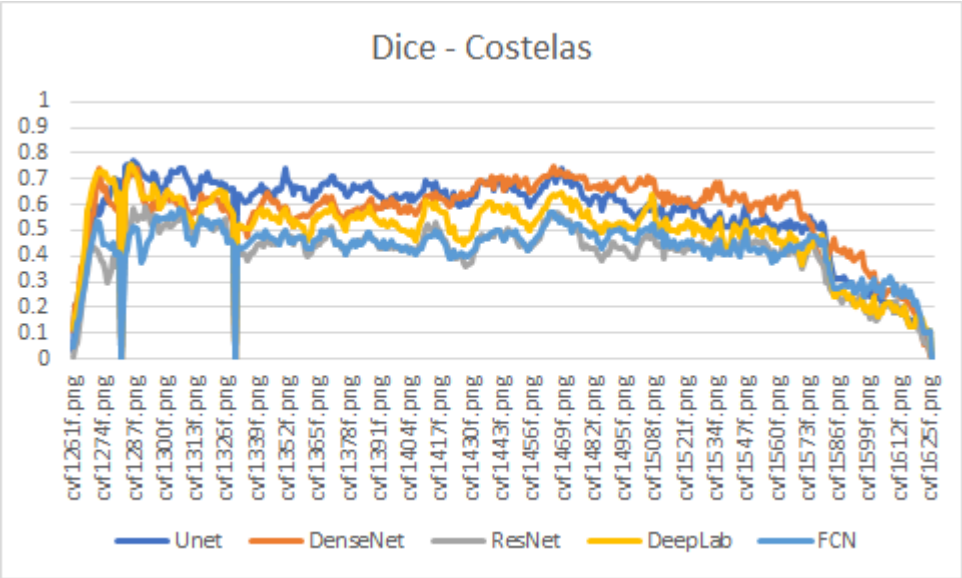


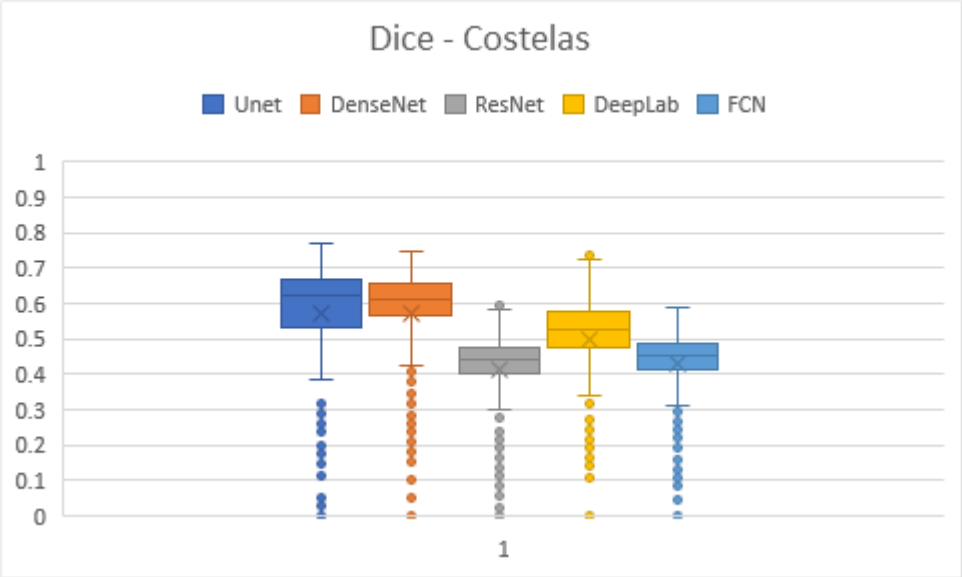
Costelas - Corpo Feminino

Quadro estatístico contendo a contagem de tomografias, soma, média, variância e desvio padrão para os coeficientes Dice na segmentação das costelas.

Dice					
Rede	*Contagem*	*Soma*	*Média*	*Variância*	*Desvio Padrão*
U-Net	366	208.135	0.5687	0.0258	∓0.1604
DenseNet	366	208.653	0.5701	0.0207	∓0.1438
ResNet	366	151.09	0.4128	0.0132	∓0.1149
DeepLab	366	182.419	0.4984	0.0203	∓0.1422
FCN	366	157.509	0.4304	0.0094	∓0.0969

Desempenho das redes *deep learning* U-Net, DenseNet, ResNet, DeepLab e FCN para os coeficientes Dice na segmentação das costelas do corpo feminino.

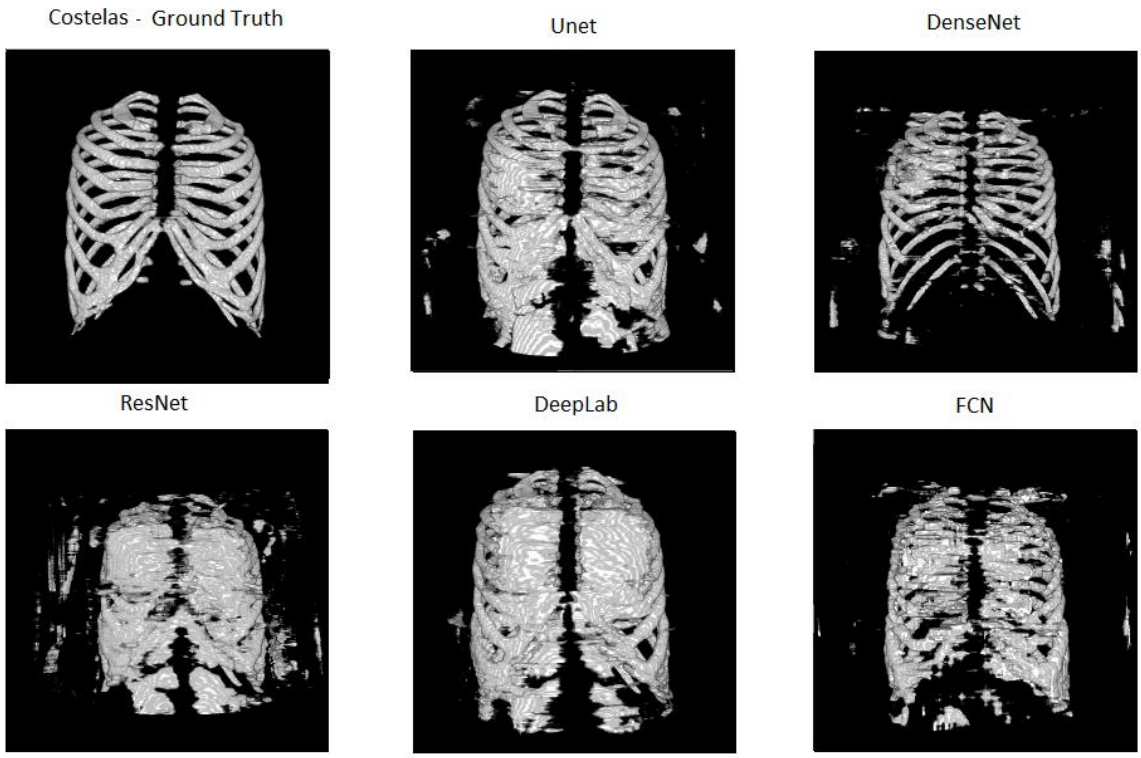




Quadro com as linhas da *matriz de confusão* para a segmentação das costelas do corpo feminino do VHP.

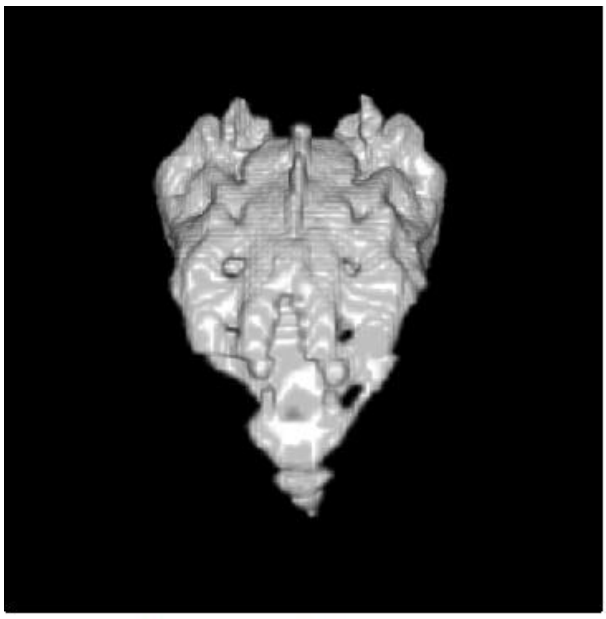
Matriz Confusão	U-Net	DenseNet	ResNet	DeepLab	FCN
fundo	920155.6	992699.8	1594542.2	996673.8	1235523.4
clavícula	659	4158.2	1268	177.4	3075.8
crânio	7402	73976.8	41902	0	17122.6
pés	3361.4	34801.2	13026.4	1182.4	1250.2
fêmur	603.4	221.2	4594.2	412	323.4
fíbula	2379.4	1978.6	6829.2	1340.6	21.4
mãos	8538.6	35566	15323.6	4970.6	18171.8
bacia	1699.2	3777.8	7514.8	88	2480.6
úmero	3123.8	3779.2	7160.4	142.4	19.8
mandíbula	69.4	293.8	314	0	480.2
patela	0	1133.2	116.8	11.8	605.8
rádio	1683.6	14806.6	2468.8	755.2	272.8
costelas	**557087.8**	**498776.4**	**481738**	**453490.2**	**503788**
sacro	1344.2	1220.6	4	0	75.6
escápula	5262.6	14666.4	7454	1340.6	4825.8
esterno	530.4	217.8	501.8	126	1612.8
tíbia	34.8	237.8	2085	484	783
ulna	272.4	2355.2	304.6	31.4	302.8
vértebras	14551.2	41200.6	42444.4	9459.6	15984.8

Resultado da segmentação das redes U-Net, DenseNet, ResNet, DeepLab e FCN juntamente com o *ground truth* para as costelas femininas.

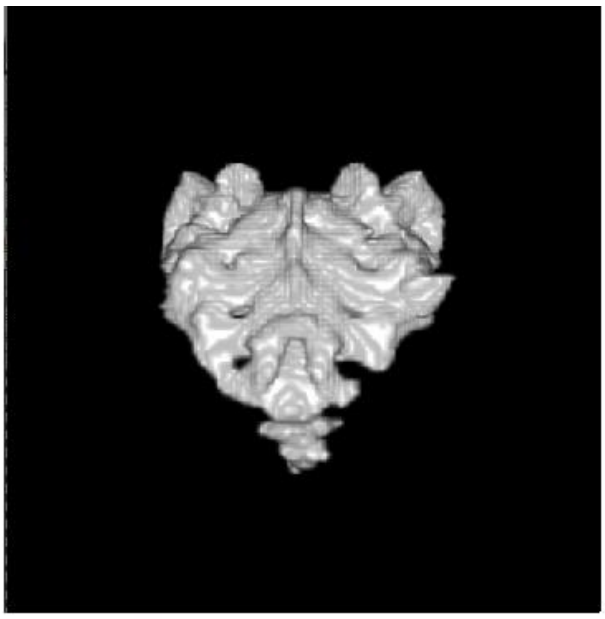


8.13 Sacro

O osso sacro é um osso curto, em formato triangular, localizado na base da coluna vertebral e tem como função transmitir toda a força da cabeça, tronco e membros superiores para os membros inferiores.



Sacro - Corpo Masculino

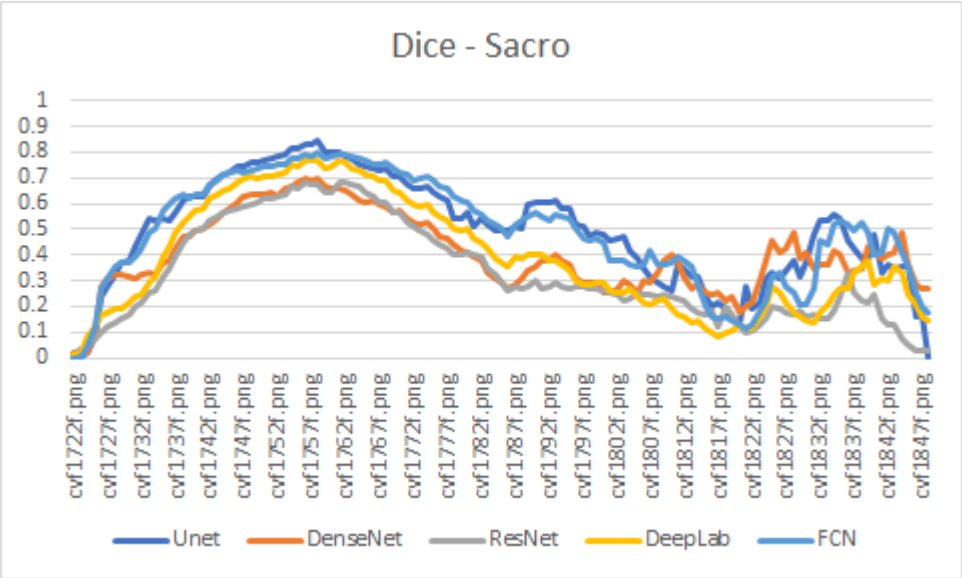


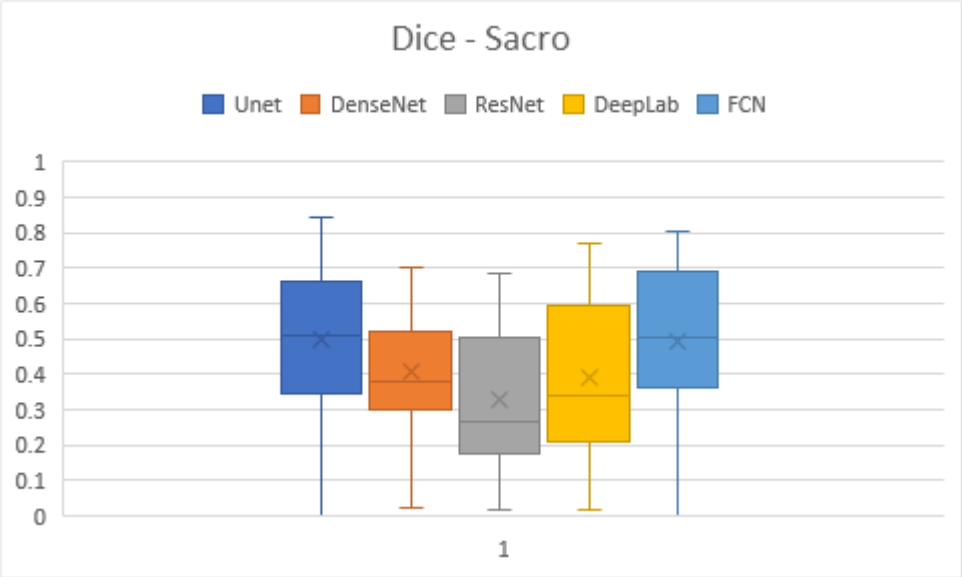
Sacro - Corpo Feminino

Quadro estatístico contendo a contagem de tomografias, soma, média, variância e desvio padrão para os coeficientes Dice na segmentação do sacro.

Dice					
Rede	*Contagem*	*Soma*	*Média*	*Variância*	*Desvio Padrão*
U-Net	127	62.954	0.4957	0.0430	∓0.2064
DenseNet	127	51.673	0.4069	0.0227	∓0.1500
ResNet	127	41.828	0.3294	0.0379	∓0.1938
DeepLab	127	49.395	0.3889	0.0478	∓0.2177
FCN	127	62.622	0.4931	0.0432	∓0.2071

Desempenho das redes *deep learning* U-Net, DenseNet, ResNet, DeepLab e FCN para os coeficientes Dice na segmentação do sacro do corpo feminino.

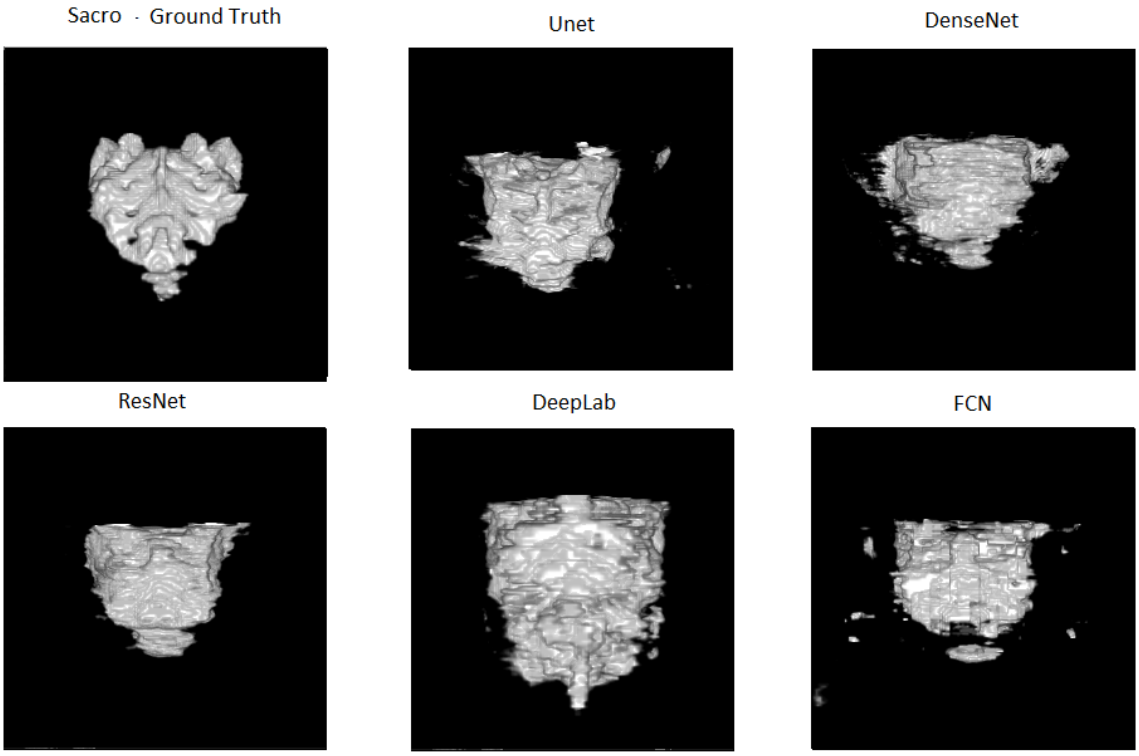




Quadro com as linhas da *matriz de confusão* para a segmentação do sacro do corpo feminino do VHP.

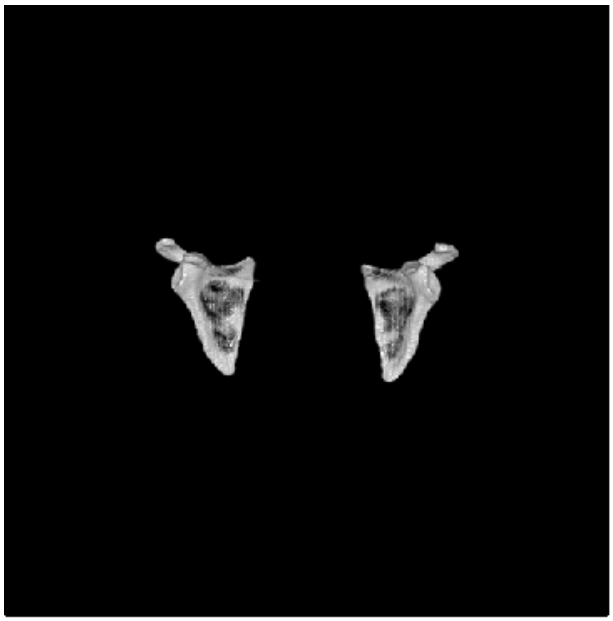
Matriz Confusão	U-Net	DenseNet	ResNet	DeepLab	FCN
fundo	254872.2	601758.6	767084.4	398385.4	291693.6
clavícula	132.4	11.2	0	0	25.8
crânio	261.2	108.2	527	0	2526.6
pés	65.8	1137.8	464.8	271.8	0
fêmur	26.2	78.8	0	6601.2	439.4
fíbula	0	0	3.8	0	0
mãos	0	0	0	0	415.6
bacia	2516.2	6004.4	19018.8	1208.2	5961.2
úmero	27.4	3197.8	0	0	1.8
mandíbula	0	0	0.2	0	21.4
patela	0	0	0	0	0
rádio	0	17.6	0	0	0
costelas	1478.6	3170.6	7788.4	538	10540.2
sacro	**233175.2**	**235864.8**	**241207.6**	**194141.8**	**234682.8**
escápula	1590.8	554	505.8	0.4	2684.4
esterno	0	0	0	0	0.6
tíbia	0.2	0	597.6	446	141.6
ulna	1	0	0	0	91.8
vértebras	11489.6	17380.8	63226.6	28892.2	26589

Resultado da segmentação das redes U-Net, DenseNet, ResNet, DeepLab e FCN juntamente com o *ground truth* para o sacro feminino.

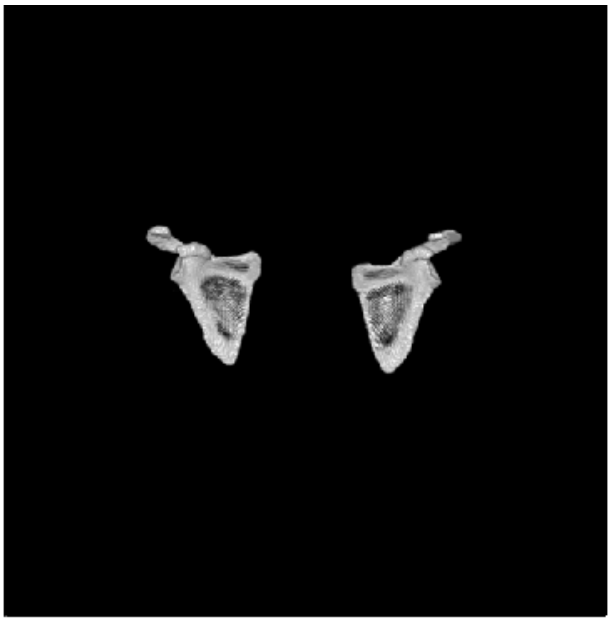


8.14 Escápula

A escápula é um osso achatado, triangular e fino, cuja função, em conjunto com os músculos, é auxiliar na estabilidade e movimentação dos ombros.



Escápula - Corpo Masculino

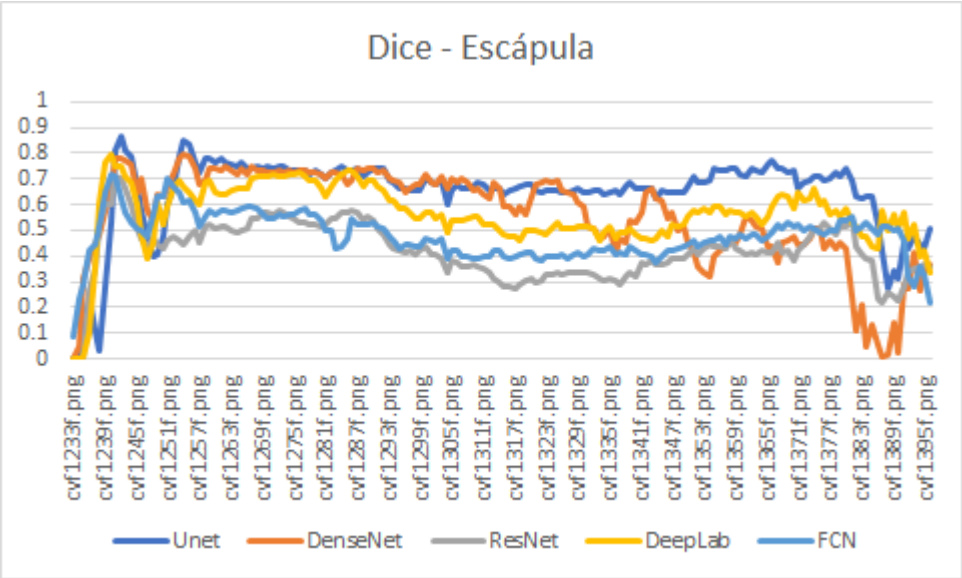


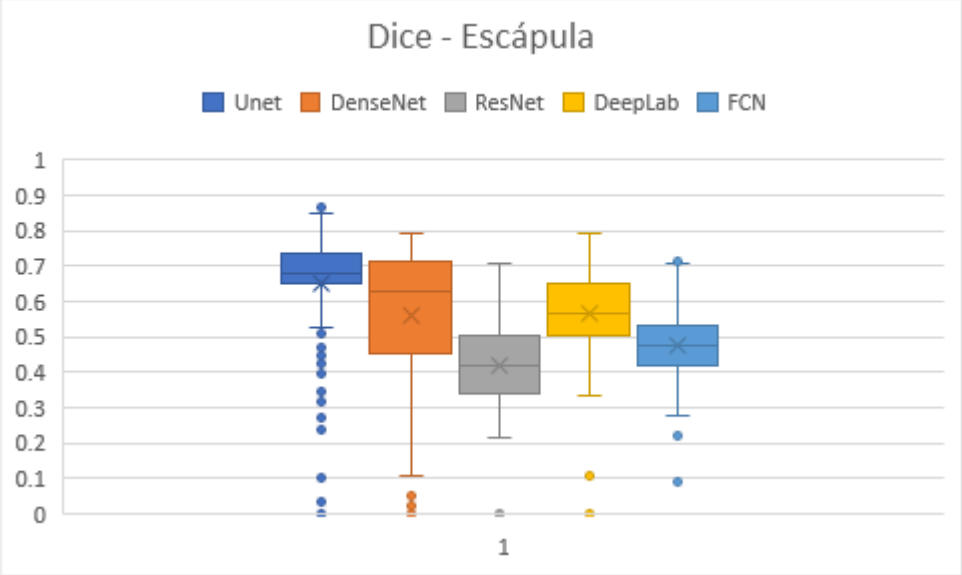
Escápula - Corpo Feminino

Quadro estatístico contendo a contagem de tomografias, soma, média, variância e desvio padrão para os coeficientes Dice na segmentação da escápula.

Dice					
Rede	*Contagem*	*Soma*	*Média*	*Variância*	*Desvio Padrão*
U-Net	164	106.427	0.6489	0.0227	∓0.1503
DenseNet	164	92.215	0.5623	0.0348	∓0.1859
ResNet	164	68.876	0.4200	0.0128	∓0.1129
DeepLab	164	92.439	0.5637	0.0154	∓0.1237
FCN	164	78.293	0.4774	0.0080	∓0.0892

Desempenho das redes *deep learning* U-Net, DenseNet, ResNet, DeepLab e FCN para os coeficientes Dice na segmentação da escápula do corpo feminino.

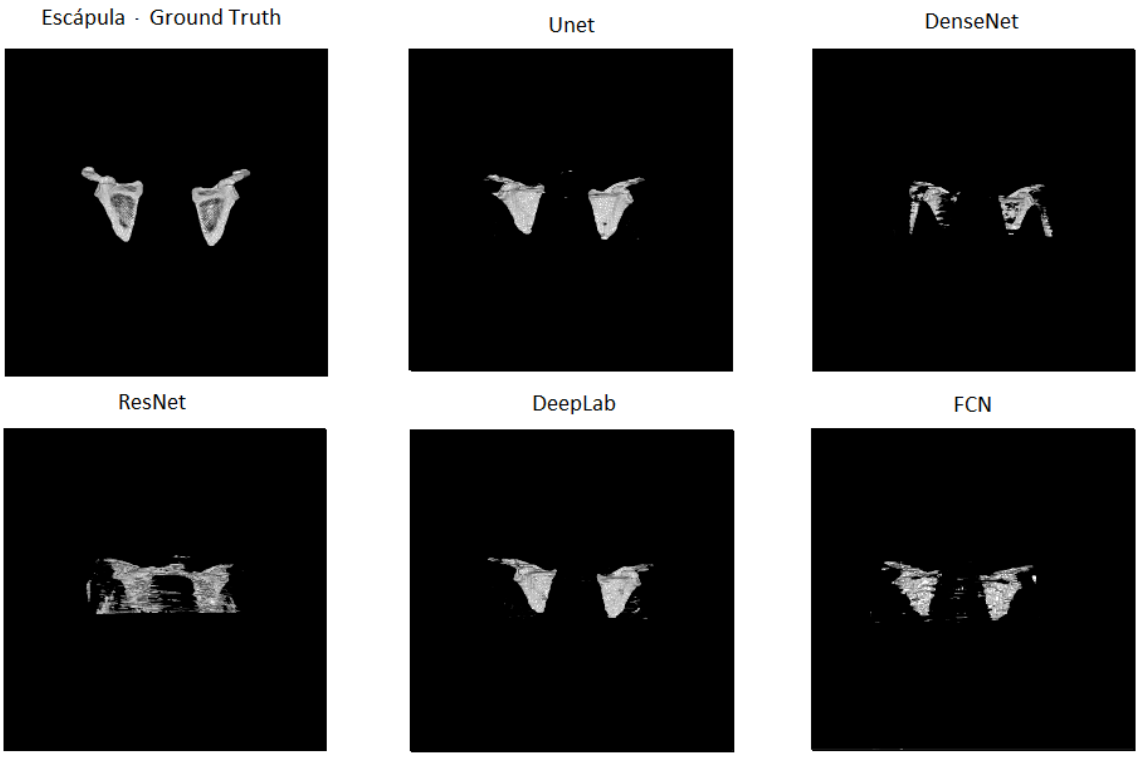




Quadro com as linhas da *matriz de confusão* para a segmentação da escápula do corpo feminino do VHP.

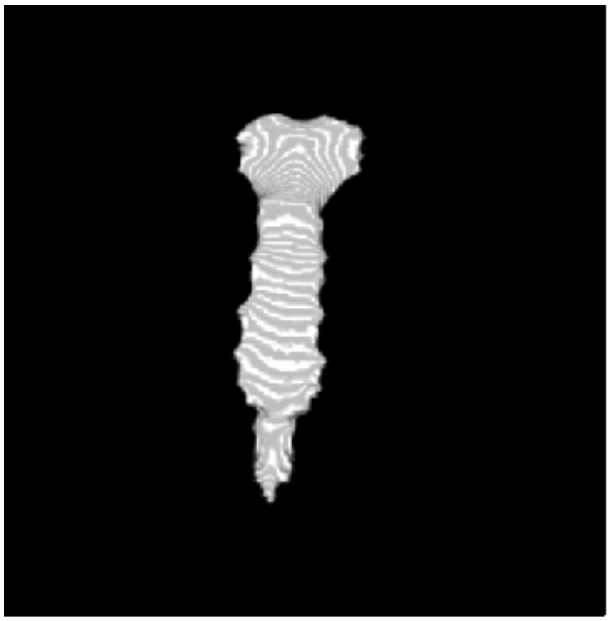
Matriz Confusão	U-Net	DenseNet	ResNet	DeepLab	FCN
fundo	106829	85578.2	414851.6	243202	273322.6
clavícula	229.6	486.4	672.4	0	4269
crânio	395.2	105.2	2474.2	0.6	4602
pés	15792.4	36861.8	41597	1135.2	2700.4
fêmur	1.2	3837.4	8743.8	44456.4	1262.4
fíbula	3035	51683.4	2769.4	15656.6	29
mãos	344	3428.4	2331.4	0	505.8
bacia	12177	935.8	894.8	234.4	14048
úmero	1908.6	22332.4	14838.8	4586.8	4317.2
mandíbula	0	152.8	159.8	0	3380.4
patela	0	0	28.2	11.8	0
rádio	224.6	10092.4	798.8	39.8	101
costelas	2105.4	1427	8369.6	4386.6	5859.4
sacro	504.4	0	196.2	0	3308
escápula	**139072**	**123041.2**	**138620.4**	**124406**	**139681.4**
esterno	0	0	0	0	22.4
tíbia	201.2	3897.2	3963.2	5657.6	950.6
ulna	785.6	280.8	189.2	0	38.6
vértebras	823.8	89	5309	172.6	4526.6

Resultado da segmentação das redes U-Net, DenseNet, ResNet, DeepLab e FCN juntamente com o *ground truth* para a escápula feminina.



8.15 Esterno

O osso esterno é um osso chato e longo, no formato de “T”, localizado na região do peito, cuja função é proteger os órgãos torácicos.



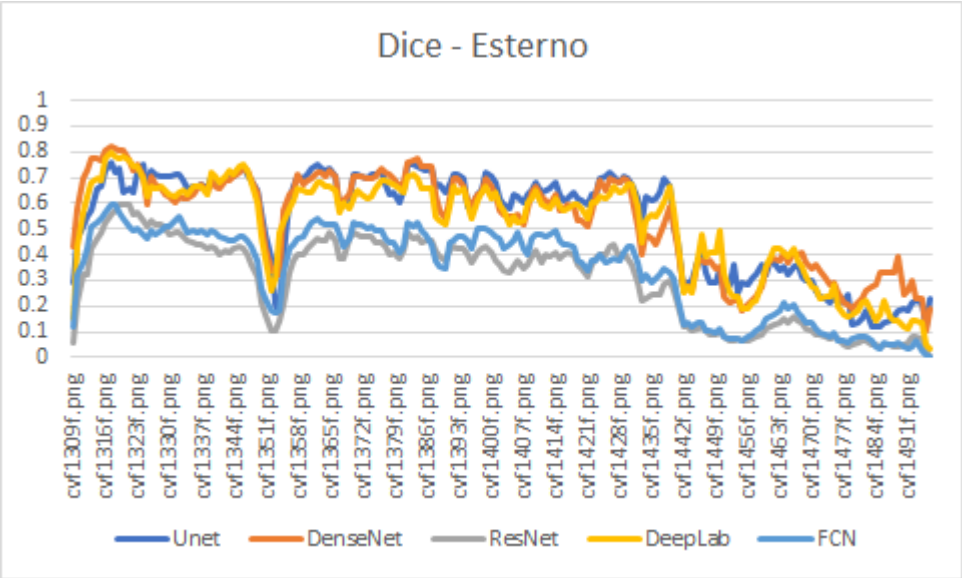
Esterno - Corpo Masculino

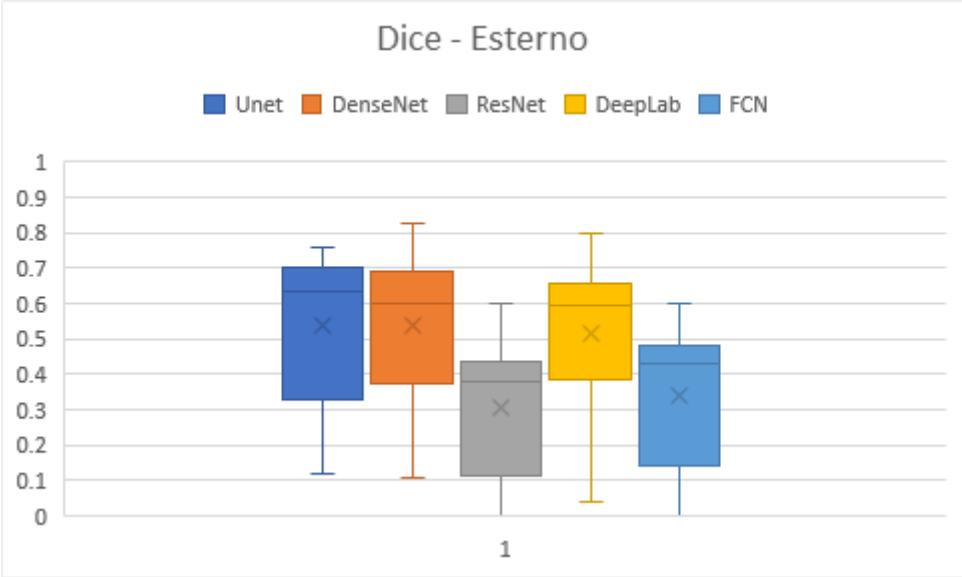
Esterno - Corpo Feminino

Quadro estatístico contendo a contagem de tomografias, soma, média, variância e desvio padrão para os coeficientes Dice na segmentação do esterno.

Dice					
Rede	*Contagem*	*Soma*	*Média*	*Variância*	*Desvio Padrão*
U-Net	188	100.672	0.5355	0.0406	∓0.2010
DenseNet	188	101.416	0.5394	0.0334	∓0.1824
ResNet	188	57.746	0.3072	0.0278	∓0.1664
DeepLab	188	96.855	0.5152	0.0374	∓0.1929
FCN	188	64.019	0.3405	0.0309	∓0.1752

Desempenho das redes *deep learning* U-Net, DenseNet, ResNet, DeepLab e FCN para os coeficientes Dice na segmentação do esterno do corpo feminino.

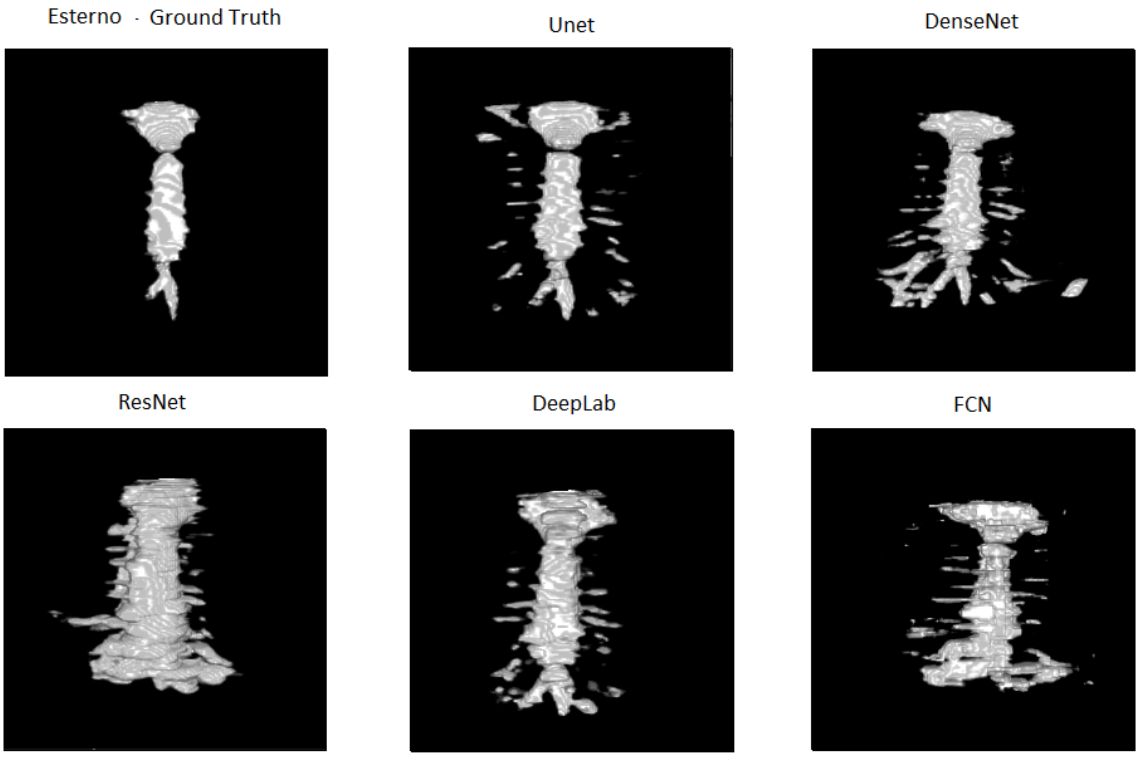




Quadro com as linhas da *matriz de confusão* para a segmentação do esterno do corpo feminino do VHP.

Matriz Confusão	U-Net	DenseNet	ResNet	DeepLab	FCN
fundo	44641.6	59729.6	234697	50297.4	141772.2
clavícula	581.4	2732.4	3476	199	1941.8
crânio	9.4	2045.4	190.4	0	2155.4
pés	0	42.6	0	33.2	6
fêmur	0	0	6.8	1135.8	116.2
fíbula	0	0	0	0	5
mãos	0.2	1349.8	223.2	5.4	1459.4
bacia	9.4	1545	77.2	0.4	21.6
úmero	0	14.6	0	0	3.4
mandíbula	0	931.8	0	0	0.4
patela	0	15.4	0	0	148
rádio	0	0.6	0	0	27.4
costelas	18619.8	24553.2	40447.6	12928.6	38036.4
sacro	0	0	0	0	0.4
escápula	0	0	0.4	0	16.4
esterno	**49015.4**	**52024.8**	**50496.6**	**40960**	**46993.6**
tíbia	0	26.8	0	877.4	15.8
ulna	0	0	0	0	30.2
vértebras	4.2	4326.8	0	0	31

Resultado da segmentação das redes U-Net, DenseNet, ResNet, DeepLab e FCN juntamente com o *ground truth* para o esterno feminino.



8.16 Tíbia

Os ossos da tíbia são longos, localizados entre os pés e o joelho, com a função de facilitar a sustentação do corpo, além de ajudar na movimentação.

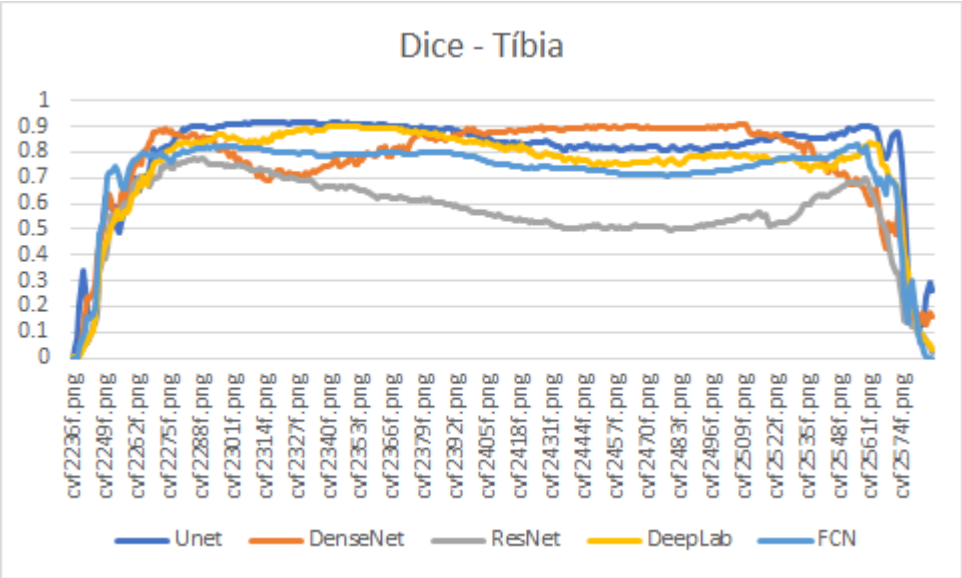
Tíbia - Corpo Masculino

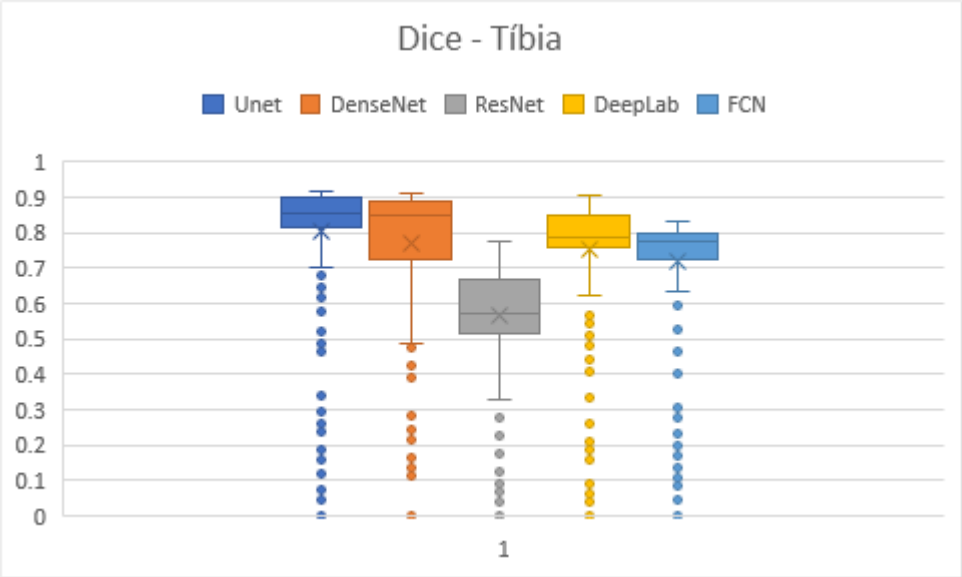
Tíbia - Corpo Feminino

Quadro estatístico contendo a contagem de tomografias, soma, média, variância e desvio padrão para os coeficientes Dice na segmentação da tíbia.

Dice					
Rede	*Contagem*	*Soma*	*Média*	*Variância*	*Desvio Padrão*
U-Net	351	282.607	0.8051	0.0312	∓0.1763
DenseNet	351	269.657	0.7683	0.0352	∓0.1872
ResNet	351	198.233	0.5648	0.0251	∓0.1581
DeepLab	351	263.83	0.7517	0.0360	∓0.1895
FCN	351	252.556	0.7195	0.0288	∓0.1696

Desempenho das redes *deep learning* U-Net, DenseNet, ResNet, DeepLab e FCN para os coeficientes Dice na segmentação da tíbia do corpo feminino.

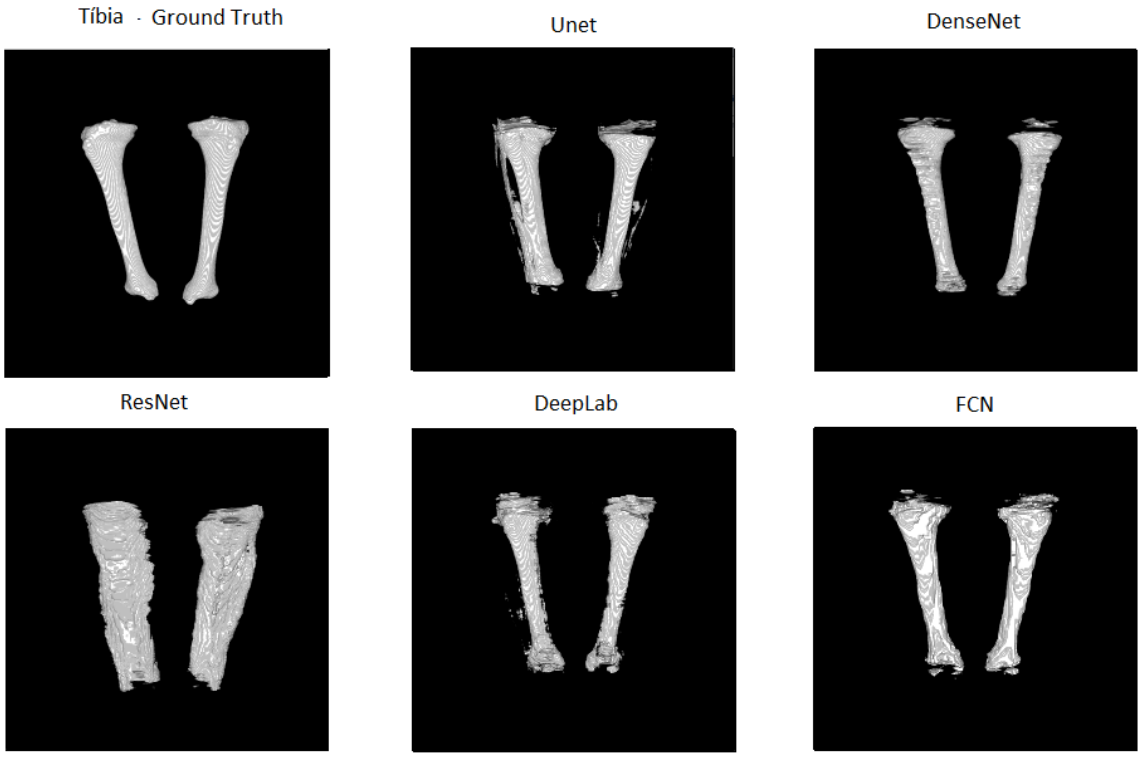




Quadro com as linhas da *matriz de confusão* para a segmentação da tíbia do corpo feminino do VHP.

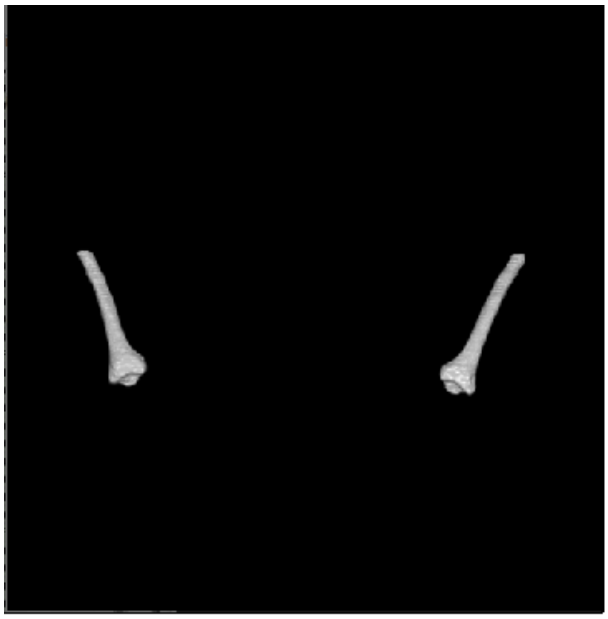
Matriz Confusão	U-Net	DenseNet	ResNet	DeepLab	FCN
fundo	288432	170043.6	1456584.8	445221.4	572629
clavícula	0	0	0.6	0	1.2
crânio	10010.6	40.8	1609.4	1827	879
pés	15161.2	44816.6	30325.6	97160.2	7760
fêmur	2157.6	98751.4	112518.8	79589.2	95317.8
fíbula	949	31.4	16500.6	8007.6	7816.2
mãos	48.2	0	714.4	13.2	1687.4
bacia	0	0	0	0	0
úmero	0	0.2	0	110.2	12.4
mandíbula	1.6	32.4	31.8	9.6	78.4
patela	197.4	3585.2	27706.4	45012	6450.6
rádio	0	0	0	0	0
costelas	0	0	0	0.6	64
sacro	0	0	0	0	0
escápula	0.2	0	0	0	0
esterno	0	0	0	0	0
tíbia	**933223.6**	**799234**	**897553.4**	**689956.8**	**958814.2**
ulna	187	0.6	1	0.2	158.4
vértebras	0	0	0	55	71.2

Resultado da segmentação das redes U-Net, DenseNet, ResNet, DeepLab e FCN juntamente com o *ground truth* para a tíbia feminina.

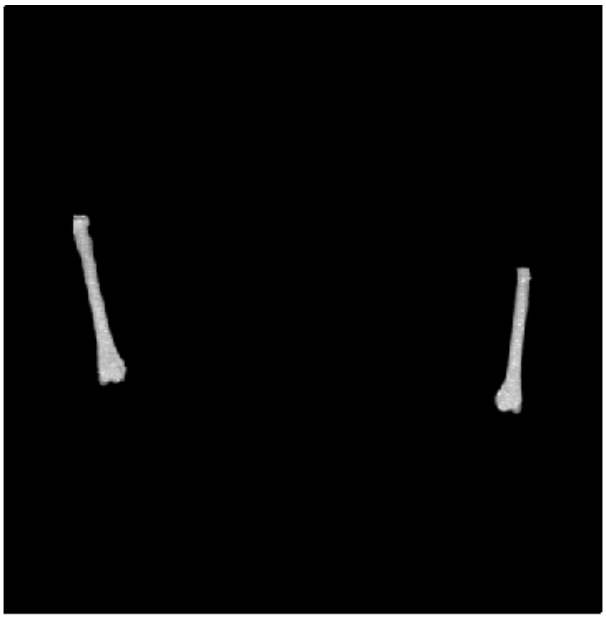


8.17 Ulna

Os ossos da ulna são longos e estão localizados no antebraço tendo como função ajudar na movimentação dos membros superiores.



Ulna - Corpo Masculino

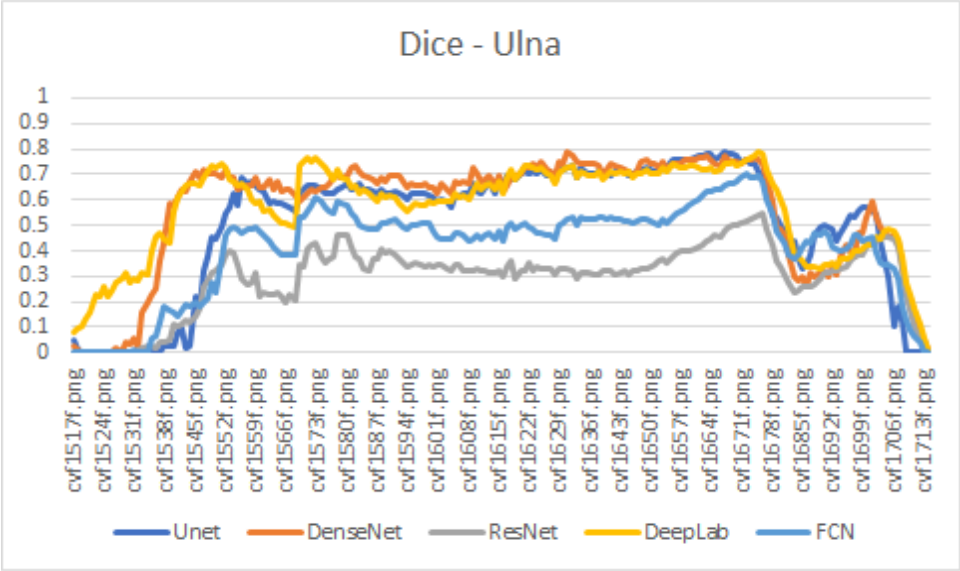


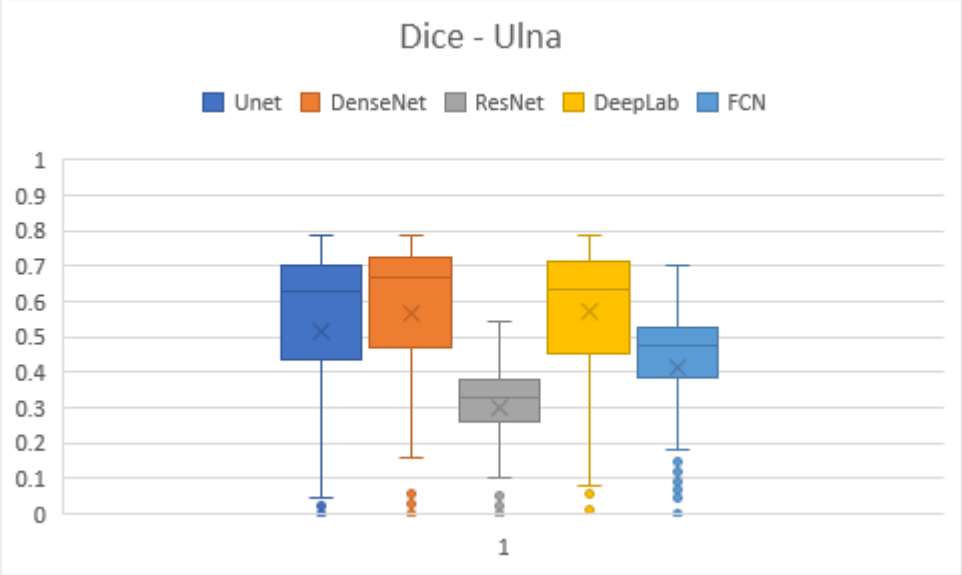
Ulna - Corpo Feminino

Quadro estatístico contendo a contagem de tomografias, soma, média, variância e desvio padrão para os coeficientes Dice na segmentação da ulna.

Dice					
Rede	*Contagem*	*Soma*	*Média*	*Variância*	*Desvio Padrão*
U-Net	198	101.424	0.5122	0.0673	∓0.2588
DenseNet	198	112.127	0.5663	0.0532	∓0.2301
ResNet	198	58.957	0.2978	0.0188	∓0.1366
DeepLab	198	113.018	0.5708	0.0326	∓0.1802
FCN	198	81.825	0.4133	0.0357	∓0.1885

Desempenho das redes *deep learning* U-Net, DenseNet, ResNet, DeepLab e FCN para os coeficientes Dice na segmentação da ulna do corpo feminino.

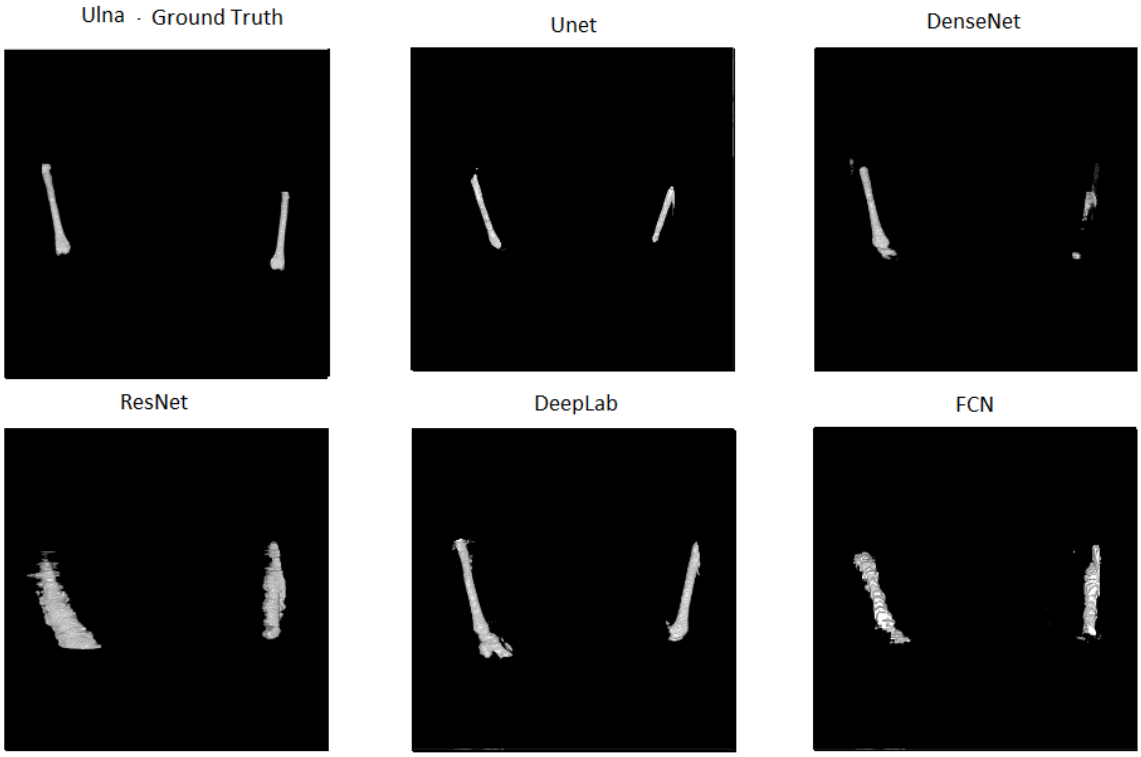




Quadro com as linhas da *matriz de confusão* para a segmentação da ulna do corpo feminino do VHP.

Matriz Confusão	U-Net	DenseNet	ResNet	DeepLab	FCN
fundo	73444	68024.2	394787.2	103864	146485.4
clavícula	0	0	0	0	0.4
crânio	0	22.2	1201.6	0	99.8
pés	10646.2	11346.6	25796	0.6	11539.4
fêmur	142.4	4.6	13.8	1.2	2889
fíbula	0	0	89.6	0	53.2
mãos	11677.4	19247.6	59757.2	51064	21009.4
bacia	0	0	0	0	0
úmero	2131	222.4	1200	1636	24.8
mandíbula	0	0	15.8	0	285.2
patela	0	120.2	1486	0.6	166.6
rádio	1100	319.6	937	3643.8	437.6
costelas	0	0	0	0	12.8
sacro	0	0	0	0	0
escápula	0	0	0	1.6	0
esterno	0	0	0	0	0.6
tíbia	4.4	760.8	2015	0	1152.4
ulna	**60407.8**	**60540**	**69495.8**	**59346**	**64635.6**
vértebras	0	0	0	0	0

Resultado da segmentação das redes U-Net, DenseNet, ResNet, DeepLab e FCN juntamente com o *ground truth* para a ulna feminina.



8.18 Vértebras

As vértebras são um conjunto de ossos que compõe a coluna vertebral formando o eixo de sustentação do corpo. Têm como função dar sustentabilidade e flexibilidade à movimentação do tronco, além de proteger a medula espinhal.

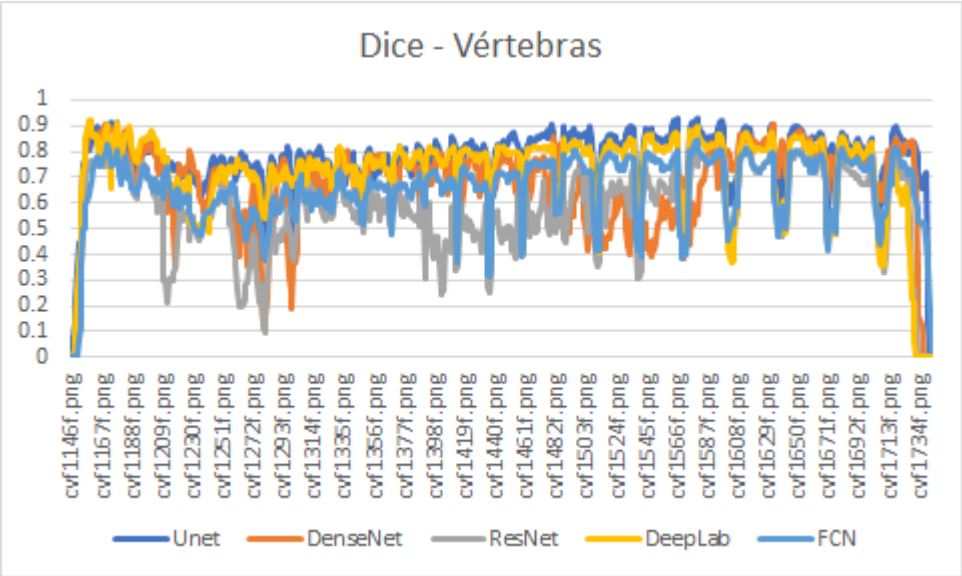
Vértebras - Corpo Masculino

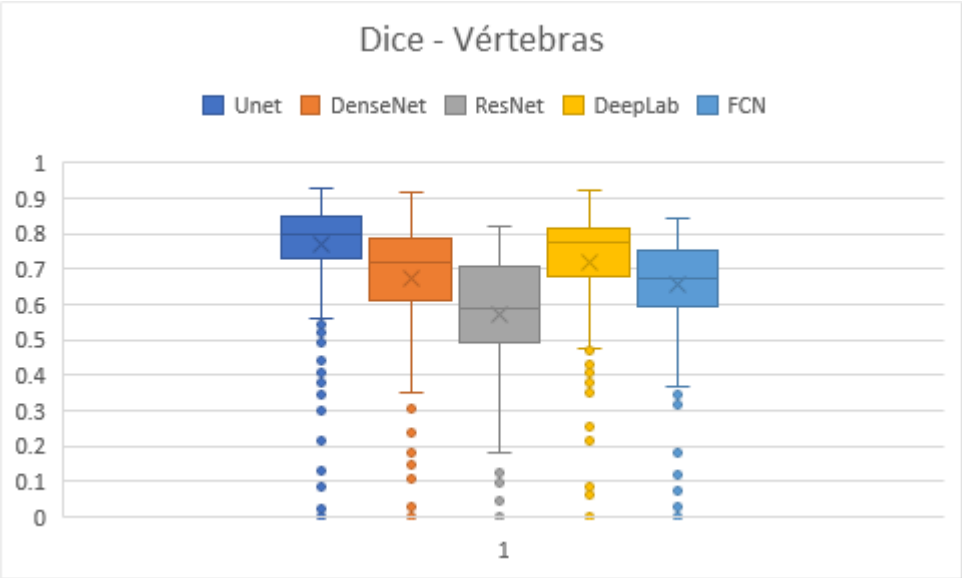
Vértebras - Corpo Feminino

Quadro estatístico contendo a contagem de tomografias, soma, média, variância e desvio padrão para os coeficientes Dice na segmentação das vértebras.

Dice					
Rede	*Contagem*	*Soma*	*Média*	*Variância*	*Desvio Padrão*
U-Net	598	461.039	0.7710	0.0150	∓0.1225
DenseNet	598	403.405	0.6746	0.0289	∓0.1700
ResNet	598	342.46	0.5727	0.0288	∓0.1695
DeepLab	598	428.528	0.7166	0.0290	∓0.1702
FCN	598	392.47	0.6563	0.0168	∓0.1294

Desempenho das redes *deep learning* U-Net, DenseNet, ResNet, DeepLab e FCN para os coeficientes Dice na segmentação das vértebras do corpo feminino.





Quadro com as linhas da *matriz de confusão* para a segmentação das vértebras do corpo feminino do VHP.

Matriz Confusão	U-Net	DenseNet	ResNet	DeepLab	FCN
fundo	541157	370847.6	795611	542740.4	818883
clavícula	0	0	0	1.2	41.6
crânio	5806.6	8231.6	9396	4107.6	4018.4
pés	2973.4	18150.4	545.6	34057.2	687
fêmur	8.6	0.2	19.2	4483.8	4391.6
fíbula	0	292.4	391.2	5028	27.4
mãos	24.4	0.2	18.6	0	2366.4
bacia	228	347.6	131.2	0	1781.6
úmero	0	32	0	0	33.6
mandíbula	0	1471	0	6.2	735.6
patela	0	0	0	9.8	0
rádio	0	7	0	0	46.4
costelas	8748.8	14568.2	23581.6	6299	23030.6
sacro	1337.4	1071.4	818.6	43.4	1795.8
escápula	1112.2	1796	355.6	0	3594.2
esterno	0	0	0	0	0
tíbia	0	0	693.8	18307	517.4
ulna	0	0	0	0	48.8
vértebras	**1032789.2**	**862786.4**	**881319.4**	**815781.8**	**1001499.4**

Resultado da segmentação das redes U-Net, DenseNet, ResNet, DeepLab e FCN juntamente com o ground truth para as vértebras femininas.