A Generative Domain Adaptation Scheme for Swift Deployment of Parking Monitoring Systems

Antonio Michel Ferreira dos Santos[1][0009−0000−8528−7079], Paulo Ricardo Lisboa de Almeida[2][0000−0002−5502−7839], Jean Paul Barddal[1][0000−0001−9928−854X], and André Gustavo Hochuli[1][0000−0003−0190−547X]

[1] Programa de Pós-Graduação em Informática (PPGIa), Pontíficia Universidade Católica do Paraná, Curitiba, Brazil
`{antonio.fsantos,jean.barddal,aghochuli}@ppgia.pucpr.br`
[2] Departamento de Informática (DINF), Universidade Federal do Paraná, Curitiba, Brazil
`paulorla@ufpr.br`

Abstract. Deep learning models have demonstrated remarkable accuracy in distinguishing between empty and occupied parking spaces when large amounts of annotated training data are available from the target environment. However, in real-world deployments, the major bottleneck lies in the labor-intensive annotation process required whenever a new scenario arises, or retraining is needed due to changes in the camera setup, often driven by maintenance, repositioning, or environmental conditions. This paper addresses this challenge by proposing a generative domain adaptation scheme designed to reduce annotation requirements and accelerate deployment significantly. Instead of relying on extensive labeled datasets and computationally expensive model retraining, our method synthesizes new training samples based on a small subset of instances from the target domain. In particular, by combining generative augmentation with a lightweight convolutional network for inference, our approach achieves a favorable balance between annotation cost, computational efficiency, and accuracy. These results highlight the method's potential as a cost-effective and rapidly deployable solution for real-world parking lot monitoring. Under a cross-dataset evaluation protocol, we highlight that our approach achieves competitive accuracy (close to 97%) using as few as 256 labeled samples, thus substantially reducing human annotation effort without sacrificing classification performance.

Keywords: Parking Lot Monitoring · Parking Space Classification · Domain Adaptation · Generative Models

1 Introduction

Camera-based parking lot monitoring systems are being integrated into smart cities due to deep learning techniques achieving near-human performance [6], with recent studies demonstrating over 99% accuracy in determining parking spot occupancy [1, 2, 4, 8]. However, this performance relies on extensive target-annotated data, which is time-consuming and demands high computational costs for model training, creating deployment bottlenecks. A fast deployment solution is essential for real-world applications, as maintenance tasks like camera repositioning or short-term events impose time constraints.

Deep Convolutional Generative Adversarial Networks (DCGANs) have proven effective in generating synthetic images for data-constrained scenarios such as medical imaging, facial expression recognition, and edge applications [7]. An important aspect is that most works utilize as many available target-annotated samples as possible to fine-tune the model, without considering the data collection and annotation costs. On the other hand, their application in domain adaptation, i.e., training a model in one domain and fine-tuning it in a similar one, is still emerging [9]. In the context of parking lot monitoring systems, employing a generative model to reduce the need for annotated data represents a promising yet underexplored approach [3, 11, 12].

The main contribution of this paper is to evaluate what is the minimum number of target-annotated data required for effective domain-specific adaptation of DCGANs in parking space classification, ensuring that synthetic data provides a reliable representation for training classification models. This approach mitigates the manual annotation time bottleneck, as model training can be automated on cloud-based servers.

To answer this question in the context of parking lot monitoring, we propose a few-shot generative domain adaptation pipeline that includes i) DCGANs to minimize the amount of target annotated data - *as much as possible* - and ii) a convolutional classification model. The generative and classification models will be pre-trained on a public, robust, and fully annotated parking lot dataset and then adapted to the target domain using the least amount of samples possible.

To guide our study, we introduced the following research questions:

- RQ1: What is the minimal number of labeled target samples needed for effective domain adaptation of generative models in parking lot scenarios?
- RQ2: Can synthetic data, generated by the DCGANs, provide reliable representations to adapt the classification models to the target parking lot?
- RQ3: How does the proposed framework compare to state-of-the-art supervised methods in balancing classification accuracy and annotation effort?

To address these questions, we conducted a two-way cross-dataset experiment involving 12 different cameras, each camera simulating a new deployment environment. Our findings suggest that DCGANs achieve competitive accuracy while significantly reducing the need for labeled data, making fast annotation feasible. The computational overhead introduced by generative models into the pipeline is discussed in Section 4.

The remainder of this work is structured as follows: Section 2 reviews state-of-the-art parking lot approaches regarding cross-dataset scenarios. The proposed approach is discussed in Section 3. Section 4 details the experimental protocol and results. Finally, Section 5 summarizes the findings.

2 Related Works

In this section we discuss related works on image-based parking lot classification. Next, we bring forward existing works that rely on DCGANs to generate synthetic data in various applications.

2.1 Image-based Parking Lot Classification

Several parking lot classification approaches are available in the state-of-the-art literature, with a comprehensive review provided in [1]. While many of these methods demonstrate promising results, only some tackle the challenge within a domain adaptation protocol, mainly when limited or no samples are available from the target scenario. Furthermore, we focus on approaches that utilize public datasets, as this enables comparability and reproducibility.

In light of this, Almeida et al. [2] proposed the PKLot dataset, which contains approximately 700,000 labeled samples collected from two parking lots and three camera views. Using texture-based features and SVM models, they reported over 99% accuracy with extensive training samples from the target domain. However, the authors also highlighted generalization issues, achieving about 90% accuracy without domain-specific adaptation to the target parking lot.

Amato et al. [4] introduced the CNPark-EXT dataset, which includes nearly 160,000 annotated samples collected from 9 camera views in a single parking lot. The camera shifts capture various lighting conditions, shadows, and partial occlusions caused by obstacles. They reported an average accuracy of 88.5% with no domain adaptation. In this vein, Nurullayev and Lee [14] proposed a deep learning model incorporating dilated convolutions and skipping pixels in the convolution kernel. This approach showed promising results, achieving an average accuracy of 96.5% in a similar protocol.

In a recent study, Almeida et al. [1] demonstrated that state-of-the-art methodologies achieve an average accuracy of 92% when trained with no samples from the target parking lot. The work of Hochuli et al. [12] further supports this finding, using datasets including the PKLot and CNRPark-EXT datasets. Their approach develops a global model capable of accurately classifying images from new parking lots, achieving an average accuracy of 92.8%.

Regarding parking space annotation, state-of-the-art approaches typically use rotated rectangles or polygons for precise demarcation. Hochuli et al. [11] showed that by employing bounding boxes for annotation — though easier and less precise — can yield promising results due to the contextual information from neighboring areas. They also showed that reducing to 1,000 target-labeled samples for domain-specific adaptation can achieve 97% accuracy with a custom deep convolutional network. However, it is worth noting that their work was limited to the PKLot dataset.

2.2 Synthesizing Data using DCGANs

DCGANs have been applied to address the lack of annotated data and class imbalance across various computer vision tasks, including facial expression recognition [18] and medical image analysis [17]. A comprehensive review of these techniques and their applications can be found in [7]. Traditionally, the use of DCGANs involves training the model on a source domain to generate synthetic samples, thereby enhancing the generalization of classifier that is trained using both original and synthetically data.

An important aspect is that many generative strategies aim to enhance target domain data representation, often without regard to the quantity of available data. In contrast, we focus on minimizing the number of annotated samples required to expedite and automate the deployment pipeline. With this in mind, we explore using DCGANs for domain adaptation in parking lot scenarios with minimal annotated data.

3 Proposed Method

This work postulates that annotation effort can be minimized by employing a DCGAN to enhance target-domain representation. To address our research questions (RQs) outlined in Section 1, we propose a generative domain adaptation framework depicted in Figure 1.

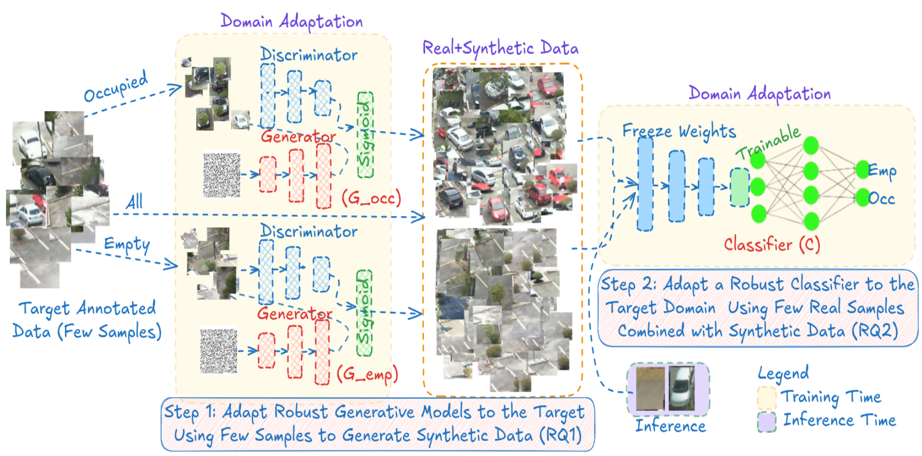


Fig. 1: The proposed deployment scheme involves performing domain adaptation of robust generative and classifier models to: i) augment the limited annotated data from the target domain and ii) adapt the classification model using a combined set of synthetic and real samples.

Given a small set of target-labeled images, the approach utilizes domain adaptation of two task-specific generative networks to synthesize images of empty and occupied parking spaces, referred to as G_{emp} and G_{occ}, respectively. The real and generated images are combined to adapt the classification model (C) to the target domain. The decision to employ two generative models is based on the fact that class-specific DCGANs require less annotated data than conditional DCGANs (cGANs) or variational autoencoders (VAEs).

It is important to remark that the generative networks G_{emp} and G_{occ} are optimized only during the domain adaptation training time, providing a robust dataset to fine-tune the classifier C. During inference, only the classifier C is used. Next, Section 3.1 details the datasets utilized, while Section 3.2 presents the model architectures.

3.1 Datasets

To the best of our knowledge, the most comprehensive parking lot datasets available in the state-of-the-art are the PKLot [2] and CNRPark-EXT [4]. These datasets include three different parking lot areas with diverse camera positions.

PKLot. The PKLot [2] dataset includes images captured over approximately three months, with a time interval of 5 minutes between each image, resulting in a total of 12,417 images and about 700,000 annotated samples divided into three deployment scenarios: UFPR04, UFPR05, and PUCPR. Image examples extracted from the PKLot dataset are provided in Figure 2.

(a) UFPR04 (b) UFPR05 (c) PUCPR

Fig. 2: The three deployment scenarios from PKLot.

CNRPark-EXT. The CNRPark-EXT [4] contains approximately 160,000 annotated parking spaces collected from nine cameras across a single parking lot. This dataset presents specific challenges, including solar light reflections, raindrops on the camera lens, and partially occluded parking spaces due to trees or lamp posts. An example of these challenges is depicted in Figure 3.

(a) Camera 1 (b) Camera 4 (c) Camera 9

Fig. 3: Three out of nine scenarios from the CNRPark-EXT.

Table 1 briefly summarizes the properties of the datasets. For a thorough description of the datasets and their applications, refer to [1, 2, 4].

Table 1: Summary parking lots datasets used in this work

Dataset	Images	Spots	Days	Cameras	Park. Lots
PKLot	12,417	695,851	99	3	2
CNRPark-EXT	4,278	157,549	23	9	1

Although using only two datasets might seem limited, we address this by combining them to simulate real-world deployment scenarios through a cross-dataset protocol. This strategy is widely recognized as a benchmark in the literature [1, 4, 11, 12, 14], ensuring a fair comparison with state-of-the-art methods. Details on the experimental protocol adopted are given in Section 4.

3.2 Model Archictectures

The generative models G_{emp} and G_{occ}, along with the classifier C shown in Figure 1, are convolutional-based networks with an input-layer of shape [128,128,3]. The classification model, C, comprises the well-known MobileNetV3-Large [13] architecture to perform feature extraction, followed by average pooling to reduce the feature map dimensions to a flattened shape of size 2048. As in [12], to perform classification, we incorporate dense layers with 512, 256, 128, and 64 neurons along a softmax activation function, as shown in Figure 4. During the training of the classifier C, the convolutional backbone is initialized with the ImageNet weights [16]. We kept all convolutional weights frozen except for the last convolutional layer. This transfer-learning strategy preserves the knowledge encoded in the pre-trained weights while allowing adjustment to the specific features of the target dataset [19].

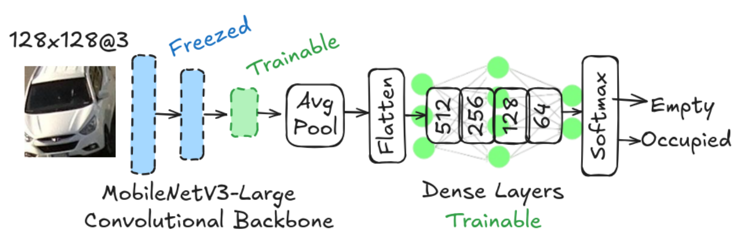


Fig. 4: The classifier model C architecture, consisting of a MobileNet backbone followed by four trainable dense layers

The decision to adopt MobileNetV3-Large as the backbone is supported by Hochuli et al. [12], which demonstrated that this architecture achieves a balance between generalization and computational costs in classifying parking spots in a cross-dataset scenario.

The generative models G_{emp} and G_{occ} use a discriminator with four convolutional layers to estimate the prior probabilities of real versus synthetic samples, reducing array dimensions by half at each layer. Conversely, the generator network consists of four deconvolutional layers, progressively upscaling the input random noise from [8,8,1024] to an output image of [128,128,3]. The architectural design was initially based in [15] and refined based on insights from [7], using the Frechet Inception Distance (FID) [10] to model selection. Figure 5 illustrates the final architecture.

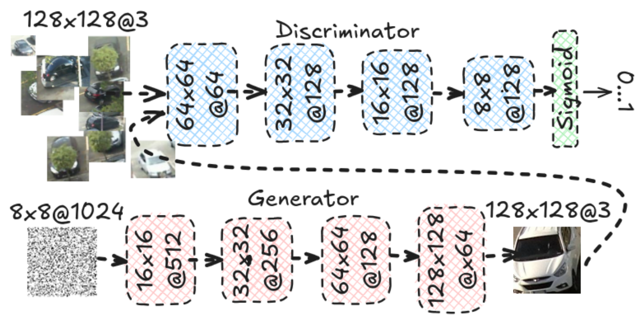


Fig. 5: The DCGAN architecture (G) includes a generator and a discriminator, each with four convolutional layers. The generator synthesizes samples, while the discriminator estimates the posterior probability.

The Adam optimizer with backpropagation was used to train all networks using mini-batches of 32 instances. The learning rate is set to 10^{-3} to expedite convergence and then reduced over time to refine the weights. Finally, the cross-entropy was used as a loss function, and early stopping was employed for regularization.[1]

4 Experiments

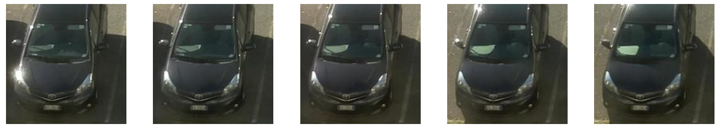
In this section, we assess our proposed approach considering a series of experiments following well-established protocols in the state-of-the-art [1, 4, 11, 12, 14], utilizing the datasets discussed in Section 3.1.

First, we detail the training protocol for the base models in Section 4.1. Subsequently, in Section 4.2, we outline the deployment protocol. Finally, Section 4.3 presents the results, offering insights into the strengths and limitations of the proposed approach.

4.1 Base Models Training Protocol

To train the classification model C on the source domain dataset, we allocated 50% of the days for training, 20% for validation, and 30% for testing. Training and validation were balanced using random undersampling of the majority class. This day-based holdout approach, discussed in [2], reduces training bias by preventing subsequent camera shots with minimal changes from appearing in both the training and testing sets, as illustrated in Figure 6.

[1] The trained models will be made available upon request



(a) Shot 1 (b) Shot 2 (c) Shot 3 (d) Shot 4 (e) Shot 5

Fig. 6: Sequential camera shots with minimal changes

To train the generative base models G_{emp} and G_{occ}, we randomly selected 9,000 class samples from the source domain: 3,000 samples per camera in the PKLOT dataset and 1,000 samples per camera in the CNRPark-EXT dataset. This ensures a balanced representation across source scenarios, preventing model bias towards a specific camera in the source domain.

Figure 7 depicts the resulting synthetic samples on source scenarios. Although it exhibits marginal deformations, the rationale here is to provide diverse shapes and textures for representation learning [5] as the classification is unassociated with the object itself, i.e., car model. The impact of generated images on the recognition rate is discussed in Section 4.3.

(a) PKLot (b) CNRPark-EXT

Fig. 7: Qualitative comparison between real samples (top row) and synthetic samples (bottom row) produced by the generative base models trained on a) PKLot and b) CNRPark-EXT datasets.

4.2 Cross-Dataset Deployment Protocol

We implement the cross-dataset protocol utilized in state-of-the-art approaches [1, 4, 11, 12, 14], in which one dataset is used entirely as the source domain to construct the generative and classification models, referred to as the base models (Section 4.1). In contrast, each camera from the other dataset represents a unique target scenario where the proposed scheme should adapt the base models.

This strategy mimics twelve deployment scenarios, encompassing challenges such as perspective shifts, varying car positions, shadows, raindrops and sunlight reflections in the camera lens, partial occlusions, and diverse weather conditions. Figure 8 illustrates individual parking spots across several camera views to highlight this diversity.

To address RQ1, which investigates the minimum number of annotated real samples (R) required for adapting generative models to the target domain, we

Fig. 8: Examples individual parking spots across the camera views in the PKLot and CNRPark-EXT datasets.

conducted experiments using $R = 64, 128, 256, 512, 1024$ labeled target samples to fine-tune the base generative models G_{emp} and G_{occ}. The samples were collected chronologically, beginning with the first day of the target camera until the desired amount was reached. The days selected for annotation were excluded from the testing set. The domain-specific generative models, G_{emp} and G_{occ}, were then used to augment the target data up to 5,000 samples. The combined real and synthetic samples were used to fine-tune the classifier C.

The rationale of using the above quantities of data is based on the findings of [11], which suggest that a classifier fine-tuned with up to 1,000 human-annotated samples achieves performance levels of 97% on the target dataset, with optimal performance attained at 5,000 annotated samples.

4.3 Analysis

The results for the proposed scheme across twelve deployment scenarios, involving cross-evaluation of the CNRPark-EXT and PKLot datasets, are summarized in Table 2, with reported accuracies representing the average of five runs initialized with different seeds to avoid biased comparisons. The *Baseline* column reports the isolated performance of classifier C on the target scenario without fine-tuning. Subsequent columns present performance metrics with varying amounts of annotated data from the target domain, comparing two approaches: i) using only real data (denoted as "Real"), and ii) using a combination of real and synthetic samples that sum up to 5000 samples, referred to as "w/Gen.". For example, the column "64/4936 - w/Gen." represents the scenario in which 64 real and 4936 generated synthetic images have been used for training.

Domain Adaptation Using Only Real Data. In this experiment, we aim to evaluate the impact of using only real data, as this is the standard approach in the state of the art. In this case, the generative models are not incorporated into the pipeline, which means that the classifier C does not include synthetic samples for the representation learning.

Considering this, 64 target-annotated samples boosted the overall accuracy from 95.1% to 96.2%, on average. It is worth mentioning the UFPR05 deployment scenario from PKLot, where the accuracy improves from 91.3% to 95.0%. However, doubling the annotated data does not yield proportional accuracy gains. With 128 real samples, the average accuracy reaches 96.8%. Beyond this

Table 2: The average accuracy over five runs for the proposed domain adaptation approach across different target scenarios.

Domain Adaptation from PKLot to CNRPark-EXT Target Scenarios

Scen.	Baseline	64 / 4936		128 / 4872		256 / 4744		512 / 4488		1024 / 3976	
	No Train	Real	w/Gen.	Real	w/Gen.	Real	w/Gen.	Real	w/Gen.	Real	w/Gen.
Cam-1	92.3 ±1.3	92.3 ±1.3	94.1 ±0.1	94.2 ±0.6	94.3 ±0.2	94.1 ±0.7	94.6 ±0.1	95.0 ±0.6	95.8 ±0.2	94.9 ±1.0	96.5 ±0.2
Cam-2	96.5 ±1.2	97.9 ±0.5	97.3 ±0.9	97.8 ±0.4	98.3 ±0.2	98.1 ±0.2	98.6 ±0.3	98.4 ±0.3	98.8 ±0.2	98.7 ±0.3	99.1 ±0.2
Cam-3	96.6 ±0.7	97.1 ±0.5	97.5 ±0.4	97.6 ±0.3	97.8 ±0.2	97.6 ±0.4	98.2 ±0.2	97.8 ±0.4	98.4 ±0.1	98.3 ±0.4	98.7 ±0.2
Cam-4	97.5 ±0.1	97.7 ±0.2	97.7 ±0.2	97.6 ±0.2	97.9 ±0.2	97.7 ±0.3	98.1 ±0.1	98.0 ±0.1	98.2 ±0.1	98.0 ±0.5	98.5 ±0.1
Cam-5	96.3 ±0.4	97.1 ±0.4	97.2 ±0.1	97.1 ±0.5	97.5 ±0.1	97.2 ±0.4	97.7 ±0.1	97.6 ±0.4	97.9 ±0.1	97.6 ±0.2	98.1 ±0.1
Cam-6	95.2 ±0.3	95.8 ±0.9	95.3 ±0.3	96.2 ±0.4	96.2 ±0.2	96.8 ±0.3	96.6 ±0.2	96.7 ±0.4	97.0 ±0.1	97.0 ±0.7	97.4 ±0.2
Cam-7	96.4 ±0.4	96.7 ±0.4	96.6 ±0.3	97.0 ±0.3	97.2 ±0.2	97.0 ±0.3	97.4 ±0.1	97.2 ±0.2	97.7 ±0.1	97.4 ±0.1	97.9 ±0.1
Cam-8	95.1 ±0.3	96.4 ±0.2	96.0 ±0.2	96.7 ±0.3	96.7 ±0.2	97.0 ±0.4	97.1 ±0.2	96.9 ±0.4	97.7 ±0.2	97.6 ±0.3	98.1 ±0.3
Cam-9	94.8 ±0.6	95.5 ±0.6	96.2 ±0.3	95.7 ±0.4	96.8 ±0.4	96.1 ±0.4	97.2 ±0.2	96.5 ±0.3	97.5 ±0.1	97.1 ±0.4	97.9 ±0.1
Average	95.6 ±1.5	96.3 ±1.7	96.4 ±1.2	96.7 ±1.2	97.0 ±1.2	96.7 ±1.2	97.3 ±1.2	97.1 ±1.0	97.7 ±0.9	97.4 ±1.1	98.0 ±0.8

Domain Adaptation from CNRPark-EXT to PKLoT Target Scenarios

Scen.	Baseline	64 / 4936		128 / 4872		256 / 4744		512 / 4488		1024 / 3976	
	No Train	Real	w/Gen.	Real	w/Gen.	Real	w/Gen.	Real	w/Gen.	Real	w/Gen.
UFPR04	93.6 ±2.1	96.5 ±2.6	96.9 ±0.6	97.9 ±0.7	97.8 ±0.2	98.2 ±0.6	98.2 ±0.2	98.4 ±0.3	98.8 ±0.1	98.5 ±0.2	99.1 ±0.1
UFPR05	91.3 ±2.1	95.0 ±2.6	97.1 ±0.6	97.2 ±0.7	98.3 ±0.2	97.6 ±0.6	98.7 ±0.2	98.2 ±0.3	99.0 ±0.1	98.9 ±0.2	99.3 ±0.1
PUCPR	95.9 ±1.1	96.5 ±0.9	97.1 ±0.3	96.9 ±0.8	97.4 ±0.2	97.4 ±0.2	97.8 ±0.1	97.5 ±0.4	98.0 ±0.1	97.8 ±0.2	98.1 ±0.1
Average	93.6 ±2.3	96.0 ±0.9	93.6 ±2.3	96.0 ±0.9	97.0 ±0.1	97.3 ±0.5	97.9 ±0.5	97.7 ±0.4	98.2 ±0.4	98.0 ±0.5	98.6 ±0.5
Overall	**95.1 ±1.9**	**96.2 ±1.5**	**96.8 ±1.2**	**96.8 ±1.5**	**97.0 ±1.3**	**97.1 ±1.5**	**97.3 ±1.2**	**97.4 ±1.0**	**97.7 ±0.9**	**97.6 ±1.0**	**97.9 ±1.0**

point, a *plateau* emerges, improving accuracy by only about 1 percentage points from 128 to 1024 annotated samples.

A minor improvement is observed in Cam-4 from CNRPark-EXT, with a performance gain of only 0.5% over the baseline. This limited improvement is due to the similarity in environmental setup between Cam-4 (Figure 3b) and the PKLot dataset scenarios, which means the base model generalization leaves little room for accuracy enhancement in this target scenario. In contrast, Cam-1 (Figure 3a) faces significant challenges due to its unique perspective, which is impacted by drastic angle differences and solar reflections, resulting in the poorest classification performance regardless of the amount of real data used.

This analysis offers a fair comparison of the approach proposed by Hochuli et al. [12], which employs MobileNetV3 within a target-free training framework, reporting a 92% accuracy on the same deployment protocol, closely aligning with our baseline rates.

Generative Domain Adaptation Scheme (RQ1/RQ2). Extending the analysis to evaluate the utilization of synthetic samples in the proposed scheme reveals that higher accuracies can be achieved with fewer annotated samples compared to scenarios where only real samples are used. For example, the UFPR05 camera achieved 97% accuracy by synthesizing data from just 64 real samples, marking a 2 percentage points improvement over using only real data and a notable 5.8 percentage points increase compared to the baseline. A similar behavior is observed in Cam-1 from CNRPark-EXT, with a gain of approximately 1.8 percentage points.

Fig. 9: The generated samples from the fine-tuned DCGAN, trained on target datasets ranging from 64 to 1,024 real images. The top rows display synthetic samples from Camera 2 (CNRPark-EXT dataset), while the bottom two rows feature samples from UFPR05 of the PKLot dataset.

On average, including the synthetic samples for the classification model domain adaptation always resulted in better accuracies when compared with the model adapted using only real images, considering all amounts of annotated samples tested. These findings are supported by qualitative observations in Figure

9, which show that the generative models yield reliable synthetic images for all numbers of real images used for its domain adaptation.

Another perspective is illustrated in Figure 10. As observed, with a quarter the amount of annotated samples, i.e., 256, the proposed approach can reach competitive results compared with a system adapted using 1,024 real samples. This trend is consistent across different amounts of real data, demonstrating that synthetic data is feasible to learn representation, thus supporting fasting annotation to swift deployment.

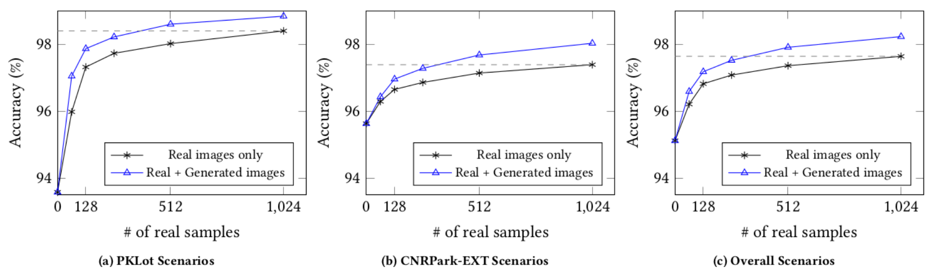


Fig. 10: Performance comparison between real images and the proposed real and generated approach using: a) PKLot as test set, b) CNRPark-EXT as test set, and c) the average across both datasets.

State-of-Art Comparison (RQ3). Table 3 compares our approach with other state-of-the-art methods, considering the use of a cross-dataset protocol, adaptation to the target scenario, and whether the amount of annotated target data was limited or extensive.

Table 3: State-of-the-Art Comparison

Author	**Approach**	**Cross Dataset**	**Limited Target Annot. Data**	**Target Training**	**Accuracy**
Almeida et al. [2]	LBP + SVM	No	No	Yes	~99%
Hochuli et al. [11]	Custom CNN	No	1000	Yes	~97%
Nurullayev et al. [14]	Custom CNN	Yes	No	Yes	~96%
Almeida et al. [1]	Survey	Yes	No	No	~92%
Hochuli et al. [12]	MobileNetV3	Yes	No	No	~92%
Amato et al. [4]	Custom CNN	Yes	No	Yes	~88%
Ours	**Custom GANs and MobileNetV3**	**Yes**	**256** **1024**	**Yes**	**~97%** **~98%**

Considering overall results, by using 256 annotated samples, our proposed approach lags only by two percentage points compared to the traditional method proposed by Almeida et al. [2], which relies on a substantial annotated dataset (>100K samples per target). Compared to Hochuli et al. [11], the proposed approach reached the same accuracy using only a quarter of the annotated data. Additionally, when annotating 1,024 samples is feasible, accuracy improved from

97% to 98%, which is a notable enhancement given the upper limit of 99% achieved in Almeida et al. [2] with an impractical number of annotated samples.

It is also noteworthy that, while both Almeida et al. [2], and Hochuli et al. [11] evaluate their methods only within the PKLot scenarios, our approach is more realistic once it spans twelve different deployment scenarios using a cross-dataset protocol well-established in the literature, as discussed in Section 4.2.

Computational Cost. One might argue that introducing Generative Adversarial Networks (GANs) could significantly increase computational costs. While this is partially true, using only 256 samples significantly reduces the training burden. Moreover, the fine-tuning process can be automated and executed on an AI server infrastructure, minimizing demands on edge devices, which perform only the inference. Additionally, the training overhead is easier to handle than annotating thousands of target samples, which is a contribution of this work in enabling a fast deployment pipeline, requiring only a quarter of the annotated data compared to the method outlined in [11, 12].

Another key aspect is that inference requires no augmentation, meaning it is performed solely by the classifier C, as shown in Figure 1. Consequently, no additional computational overhead are introduced during inference time, keeping costs comparable to state-of-the-art methods. The number of computed parameters for each forward pass during training and inference is provided in Table 4. While the training process incurs an overhead of 35.4 million parameters due to the inclusion of two generative models (G_{occ} and G_{emp}) for data augmentation and domain adaptation, the inference stage retains only 3.6 million parameters, which is consistent with the original MobileNet architecture. The result is a lightweight model considering modern deep learning approaches.

Table 4: Training and Inference Costs Based on Computed Parameters

Model	G_{occ}	G_{emp}	**Classifier** C **(MobileNet + FCs)**	**Total**
Training Params	17.7 M	17.7 M	3.6 M	39 M
Inference Params	-	-		3.6 M

5 Conclusion

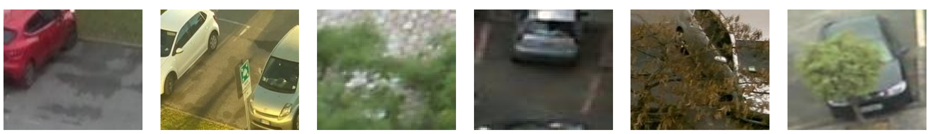
This work evaluated a deployment framework utilizing DCGANs to alleviate the data annotation burden in parking management systems. We assessed the framework through an established cross-dataset protocol in the literature that simulates twelve real-world deployment scenarios.

The results presented in Section 4 offer significant insights. Our scheme's use of synthetic samples yields performance competitive with using real samples only. Our approach achieves a 75% reduction in the manual annotation effort compared to a state-of-the-art method [11], which represents a significant decrease in potential annotation costs. This contribution addresses our research questions RQ1 and RQ2, demonstrating that synthetic data is a feasible solution for adapting models to a target-domain parking lot and contributes to time-efficient deployment.

The computational cost associated with adapting GANs is alleviated by the significant reduction to 256 annotated samples, being faster than annotating a thousand samples to train large models. Conversely, the inference pipeline does not incur additional computational costs, maintaining a comparable cost to state-of-the-art methods.

On the other hand, in scenarios where manual effort and deployment time are not constraints, traditional approaches (RQ3) trained on large datasets still achieve better accuracy, reaching nearly 99%, which is a 1% improvement over our proposed method when trained with 1,024 annotated samples. Even though our proposal does not overcome the state-of-the-art accuracy rates, we highlight that the amount of annotated data largely decreases, a significant aspect for real-world Parking Lot Monitoring systems.

A critical question arises from the performance plateau observed despite increasing the number of real samples from 64 to 1,024, as shown in Tables 3 and 4. Figure 11 highlights that occlusions caused by traffic signs, trees, lighting posts, or neighboring parked cars can lead to misclassifications when these conditions are underrepresented in the training set. Future work should incorporate these challenging cases into the target sample selection to improve sample diversity.



(a) (b) (c) (d) (e) (f)

Fig. 11: Misclassification: (a), (b), and (c) are empty spots misclassified as occupied; (d), (e), and (f) are occupied spots misclassified as empty.

Acknowledgments. The authors would like to thank CAPES-COFECUB (Grant 88887.987125/2024-00), Fundação Araucária (Grant 188/2025) and CNPQ (Grant 444192/2024-7).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Almeida, P., Alves, J., Parpinelli, R., Barddal, J.: A systematic review on computer vision-based parking lot management applied on public datasets. Expert Systems With Applications **198**, 116731 (2022)
2. Almeida, P.R., Oliveira, L.S., Britto Jr, A.S., Silva Jr, E.J., Koerich, A.L.: Pklot–a robust dataset for parking lot classification. Expert Systems with Applications **42**(11), 4937–4949 (2015)
3. Alves, P.L., Hochuli, A., de Oliveira, L.E., de Almeida, P.L.: Optimizing parking space classification: Distilling ensembles into lightweight classifiers. In: 2024 International Conference on Machine Learning and Applications (ICMLA). pp. 1016–1020 (2024)

4. Amato, G., Carrara, F., Falchi, F., Gennaro, C., Meghini, C., Vairo, C.: Deep learning for decentralized parking lot occupancy detection. Expert Systems with Applications **72**, 327–334 (2017)
5. Baker, N., Lu, H., Erlikhman, G., Kellman, P.J.: Deep convolutional networks do not classify based on global object shape. PLoS computational biology **14**(12), e1006613 (2018)
6. Bhattacharya, S., Somayaji, S.R.K., Gadekallu, T.R., Alazab, M., Maddikunta, P.K.R.: A review on deep learning for future smart cities. Internet Technology Letters **5**(1), e187 (2022)
7. Dash, A., Ye, J., Wang, G.: A review of generative adversarial networks (gans) and its applications in a wide variety of disciplines: From medical to remote sensing. IEEE Access **12**, 18330–18357 (2024)
8. Grbić, R., Koch, B.: Automatic vision-based parking slot detection and occupancy classification. Expert Systems with Applications **225**, 120147 (2023)
9. HassanPour Zonoozi, M., Seydi, V.: A survey on adversarial domain adaptation. Neural Processing Letters **55**(3), 2429–2469 (2023)
10. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. p. 6629–6640 (2017)
11. Hochuli, A.G., Britto, A.S., Almeida, P.R., Alves, et al.: Evaluation of different annotation strategies for deployment of parking spaces classification systems. In: 2022 IJCNN. pp. 1–8. IEEE (2022)
12. Hochuli, A.G., Barddal, J.P., et al.: Deep single models vs. ensembles: Insights for a fast deployment of parking monitoring systems. In: 2023 ICMLA. pp. 1379–1384 (2023)
13. Howard, A., et al.: Searching for mobilenetv3. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 1314–1324 (2019)
14. Nurullayev, S., Lee, S.W.: Generalized parking occupancy analysis based on dilated convolutional neural network. Sensors **19**(2), 277 (2019)
15. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. In: Bengio, Y., LeCun, Y. (eds.) 4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings (2016)
16. Russakovsky, O., et al.: ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision (IJCV) **115**(3), 211–252 (2015)
17. Xun, S., et al.: Generative adversarial networks in medical image segmentation: A review. Computers in Biology and Medicine **140**, 105063 (2022)
18. Yang, H., Zhu, K., Huang, D., Li, H., Wang, Y., Chen, L.: Intensity enhancement via gan for multimodal face expression recognition. Neurocomputing **454**, 124–134 (2021)
19. Yosinski, J., Clune, J., Bengio, Y., Lipson, H.: How transferable are features in deep neural networks? Advances in neural information processing systems **27** (2014)