Synthetic Images Impact in Parking Spot Classification using Neural Networks

Erick E. Cardoso
eec18@inf.ufpr.br
Universidade Federal do Paraná
Curitiba, PR, Brazil

Paulo Lisboa de Almeida
paulorla@ufpr.br
Universidade Federal do Paraná
Department of Informatics (DInf)
Curitiba, PR, Brazil

Abstract

Urban growth and the increasing number of vehicles have intensified the demand for efficient parking management solutions. In this context, machine learning-based image monitoring systems have gained prominence due to their low cost and ease of installation compared to traditional methods, such as physical sensors. These systems achieve an average accuracy of 95% in cross-validation scenarios using well-known datasets like PKLot and CNRPark-EXT. However, despite the availability of extensive datasets, challenges remain regarding the accessibility and diversity of training data. This is especially critical when aiming to improve the accuracy of generalist models or specialize them for specific scenarios, where each application requires a substantial effort to collect, segment, and label new images for optimal performance. This study proposes the use of synthetic images, generated with the Unity 5 graphics engine and the Unity Perception package, to complement or replace real data in training parking classification models. A synthetic image generation protocol was developed to reduce costs compared to the collection, segmentation, and labeling of real images. The images generated through this protocol are referred to as low-fidelity due to their lower quality and reduced capacity to simulate specific environments. Using MobileNetV3 and transfer learning, experiments were conducted in three scenarios: total replacement of real data, supplementation of diverse datasets, and specialization for specific scenarios. The results showed that synthetic images could improve model generalization by up to 2% in datasets with limited real data (e.g., CNRPark-EXT). However, synthetic images alone could not fully replace real data due to their limited fidelity in replicating real-world conditions, reinforcing the need for combinations with real data or more realistic synthetic data for better results.

Keywords

COMPUTER VISION, SYNTHETIC DATA, DEEP LEARNING, PARKING SPOT CLASSIFICATION

1 Introduction

In recent years, the rapid growth of cities and the increasing number of vehicles have raised the challenge of efficiently parking cars on public roads and large parking spaces, emphasizing the need for effective management of these areas. The search for innovative solutions to optimize their use, offering convenience to drivers and reducing traffic congestion, has been the focus of extensive research over the past decade [1, 2]. In this context, image monitoring solutions that use computer vision and machine learning methods are commonly chosen due to their low cost and ease of deployment [1, 3–5], compared to other management techniques, such as those based on sensors.

In computer vision and machine learning-based methods, images captured from a high and wide perspective are used to monitor large parking areas. Each parking space is identified and segmented from the image, and machine learning-based classification models are employed to classify the segmented spaces as occupied or empty. Figure 1 shows an example of parking space classification.

Figure 1: Example of parking spaces segmented images classification, taken from PKLot [6].

Parking space classification models require a significant amount of high-quality, diverse, and representative real-world images for training. Collecting such images is a complex and costly task, demanding time, financial resources, and effort. In this context, datasets such as PKLot [6] and CNRPark-EXT [4] were created to provide images for training and validating models in the parking space classification problem. Experiments with these and other datasets have shown an average accuracy of 95% for classification in cross-validation scenarios [1, 5] and 97% when fine-tuning a generalist model for a specific scenario with a small number of labeled data from the scenario [5]. Nonetheless, the difficulty and effort required to obtain and label data persist, as this collection process must be repeated for each new scenario where the best possible model accuracy is desired. Given the continuously evolving nature of the problem, the need for new datasets for training and validation remains a challenge [1].

One way to address the lack of data is through the use of synthetic data [7, 8]. This work proposes the use of synthetic images for training parking space classification models as a way to mitigate the challenges associated with real data. Synthetic data are artificially and algorithmically generated, resembling real data, although not derived from direct observations or real-world collections. This approach can overcome some challenges of real data collection, allowing the creation of large datasets with greater diversity at a reduced cost and in less time, with the potential to match or even surpass results obtained with real data [9].

In Tobin et al. [8], it is shown that randomized, non-realistic synthetic images with low similarity to the original scenario can achieve results comparable to real images when used in deep learning models for object recognition, requiring significantly less effort to generate. Therefore, this work proposes the use of such synthetic data generation methods to evaluate the impact of synthetic images on parking classification models. The images generated through this method will be referred to as low-fidelity synthetic images. The following research questions have been formulated to guide the study:

- P1 - How does the combination of real and low-fidelity synthetic images affect the generalization of parking space classification models?
- P2 - How do low-fidelity synthetic images perform when applied to fine-tuning parking space classification models for a specific target scenario?
- P3 - Is it possible to surpass the accuracy of models trained with real images by training models exclusively with low-fidelity synthetic images?

For generating low-fidelity synthetic images, the Unity 5 graphics engine will be used in conjunction with the Unity Perception package [10], chosen for its flexibility in creating customized environments and its capability to generate large volumes of automatically labeled data, significantly reducing the cost of collection and annotation. A Convolutional Neural Network (CNN) will be adopted for training using transfer learning, an efficient technique in scenarios with limited data. The MobileNetV3 [11], in its *large* version pre-trained on ImageNet [12], was selected as the base model due to its proven results in parking space classification in related studies, demonstrating a balance between computational cost and accuracy, making it suitable for handling various simulated scenarios. The PKLot and CNRPark-EXT datasets will be used both as a reference for generating synthetic data and for training and validating the models, serving as a baseline for evaluating the impact of synthetic images.

The remainder of this work is organized as follows: Section 2 discusses the related works and state of the art, covering the main approaches to parking space classification, available datasets, and the use of synthetic images in machine learning problems. In Section 3, the problem and proposal of this work is detailed, including the process of generating low-fidelity synthetic images using Unity Perception and the protocol for training classification models. Section 4 describes the experiments conducted, presenting the training scenarios and the results obtained with different data configurations. Finally, Section 5 presents the conclusions, discussing limitations and suggesting directions for future work.

2 Related Works

One of the main limitations of machine learning-based classification models is dealing with the lack of data. A viable solution to this issue is the use of synthetic data [7, 8]. This section discusses related works that focus on machine learning methods for parking space classification and synthetic data generation techniques.

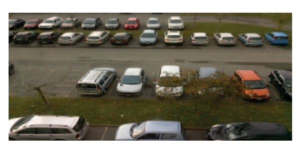
The PKLot dataset [6] is a widely used resource for parking space detection and classification research, containing 12.417 annotated images captured under various weather conditions and yielding approximately 695.900 segmented parking space images. Its diversity supports robust machine learning models development [1, 5, 6]. Figure 2 shows some examples taken from PKLot dataset. Similarly, the CNRPark-EXT dataset [4], with 4.287 images from nine parking environments and around 150.000 labeled segments, provides varied lighting, weather, and viewing angles, making it representative of real-world challenges.

(a) UFPR04

(b) UFPR05

(c) PUCPR

Figure 2: The PKLot dataset contains three scenarios: a) UFPR04, b) UFPR05 e c) PUCPR.



(a)

(b)

Figure 3: The CNRPark-EXT dataset includes different cameras installed in the same parking environment.

The study by Almeida et al. [6] used LPQ and LBP features with SVM classifiers, achieving approximately 99.5% accuracy within the same parking lot. However, cross-validation across different subsets showed a significant drop to around 85%, highlighting the method's limited generalization to varying environments. In Amato et al. [4], alongside the creation of the CNRPark-EXT dataset, deep learning models were introduced. The authors proposed mAlexNet, a lightweight Convolutional Neural Network (CNN) designed for parking space classification, reporting accuracy between 93% and 98% for cross-validation and within-dataset tests, respectively.

Similarly, authors of Hochuli et al. [5], Grbić and Koch [13], Dhuri et al. [14], Nurullayev and Lee [15], Hochuli et al. [16], Alves et al. [17] proposed deep learning approaches to tackle parking space classification. For instance, Nurullayev and Lee [15] introduced CarNet, a method based on Dilated CNNs, which skips certain pixels in the convolutional kernel. Dilated CNNs demonstrated greater robustness and generalization compared to previous methods like mAlexNet [4], achieving accuracies between 94% and 98%.

Dhuri et al. [14] proposed a real-time parking space occupancy detection system based on VGG16, trained with CNRPark-EXT [4] combined with a private dataset, achieving up to 93.4% accuracy. Grbić and Koch [13] utilized ResNet34 for parking space classification, where spaces were pre-located and segmented into squares using a method that identifies vehicles in image series and applies clustering in a bird's-eye view. The model was extensively evaluated on PKLot and CNRPark-EXT, achieving accuracies between 92% and 99%.

In Hochuli et al. [16], the impact of segmentation types (rotated rectangles, bounding boxes, and polygons) on results was analyzed, alongside the number of images needed to fine-tune a model for specific scenarios. Using a pre-trained CNN with three convolutional layers, the experiments showed over 99% accuracy in specific scenarios (same subset for training and testing), with the best results obtained using rotated rectangles. Additionally, it was found that with just 1,000 images, a pre-trained model could be fine-tuned to achieve an average accuracy of 97% in scenario changes.

Considering the time and effort required for image collection, segmentation, and labeling, Hochuli et al. [5] tested various models and techniques to identify the best-performing model in a cross-validation scenario. It was concluded that MobileNetV3 [11] is suitable for parking space classification across different scenarios, achieving an average accuracy of 95% without requiring fine-tuning, making it ideal for cases with limited data collection capabilities.

Table 1 summarizes the results obtained in the cited works that uses only real data, providing information about the machine learning classifier used, the minimum and maximum reported accuracies, and the datasets employed. Direct comparisons of the results are challenging, as the authors used different methods and validation approaches.

Table 1: Summary of the number of segmented synthetic images

Authors	Classifier	Accuracy	Validation Type	Used Datasets
Almeida et al. [6]	SVM	84% to 99,5%	Same Dataset	PKLot
Amato et al. [4]	mAlexNet	93% to 98%	Cross-Dataset & Same Dataset	PKLot + CNRPark-EXT
Nurullayev and Lee [15]	CarNet	94% to 98%	Cross-Dataset & Same Dataset	PKLot + CNRPark-EXT
Dhuri et al. [14]	VGG16	87% to 95%	Same Dataset	CNRPark + Private Dataset
Hochuli et al. [16]	Custom CNN	89% to 97%	Same Dataset	PKLot
Hochuli et al. [5]	MobileNetV3	82% to 95%	Cross-Dataset	PKLot + CNRPark-EXT
Grbić and Koch [13]	ResNet34	92% to 99%	Cross-Dataset & Same Dataset	PKLot + CNRPark-EXT

Despite these results, the problem of parking space classification still requires effords to collect and label images from the target parking lot, if an accuracy close to 99% is desired [16]. Also, To improve the generalization of models, a larger and more diverse amount of data is essential [1]. Given this, synthetic images can be used to increase the variability of the training data, so it is not necessary to train models with samples from the target parking lot.

Synthetic data is artificially and algorithmically generated, playing a crucial role in augmenting existing datasets or compensating for the lack of real data. It increases sample diversity and quantity, particularly in scenarios with limited datasets, while also helping simulate rare events, preserve data privacy, reduce dataset bias, and create testing scenarios to ensure model robustness.

Tools like Unity Perception [10] and NDDS [18] facilitate synthetic image generation. Techniques like domain randomization [8] introduce deliberate randomness in synthetic environments, varying parameters like textures, lighting, and shapes to enhance model generalization. For instance, Tobin et al. [8] showed that models trained exclusively on synthetic data with domain randomization achieved high accuracy in real-world object grasping tasks. Tremblay et al. [9] demonstrated that combining photorealistic and non-photorealistic synthetic images can effectively bridge the Reality Gap, achieving performance comparable to models trained on real data.

In the context of real-time parking space management and classification, the authors of Tschentscher et al. [19] present a 3D virtual parking simulation environment developed using Unreal Engine to evaluate video-based parking guidance systems. This environment allows for the simulation of varied conditions, such as weather, lighting, and obstructions like moving vehicles, providing highly customizable data for model training and validation. Additionally, a virtual camera was designed to generate realistic images, incorporating physical constraints such as blur, noise, and depth of field.

Furthermore, in Horn and Houben [20], the system proposed in Tschentscher et al. [19] was used to train classifiers like SVM and kNN exclusively with synthetic data. Two sequences of real images, sequence A and sequence B, were extracted from a private dataset and used for model validation. Average accuracies of 79.24% and 91.66% were achieved in validation with sequences A and B, respectively, demonstrating that tasks like parking space classification can be addressed without the need for real data, albeit with slightly lower accuracy compared to models trained on real data. This approach highlights the potential of simulated environments to reduce costs and overcome data collection challenges associated with parking space classification problems.

3 Problem Statement

As discussed in Section 2, the authors of Hochuli et al. [5] showed that it is possible to train a generalized parking space classification model capable of achieving an average accuracy of 95% in cross-validation across datasets, without the need for fine-tuning for specific scenarios. However, even better results—an average accuracy of 97%—can be achieved by fine-tuning a generalized model with a small amount of labeled data from specific scenarios [16].

Thus, it is still necessary to collect and annotate images from the target parking lot to achieve high accuracies, above 95%. This requirement poses scalability challenges, as gathering new data samples for each scenario is both costly and impractical for large-scale applications. A potential way to address this issue is through the use of synthetic data for training, specifically synthetic images. These can generate diverse weather and lighting conditions, and millions of labeled images can be produced in minutes. Such images can be used to replace or complement real ones [7–9, 21].

Also in Section 2, Horn and Houben [20] reported average accuracies of 79.24% and 91.66% in parking space classification for specific scenarios using synthetic images generated in a 3D simulation environment highly faithful to the real-world scenario, replicating realistic conditions of lighting, cameras, and weather. However, constructing such a high-fidelity simulated environment is more labor-intensive than collecting and labeling real images for a specific scenario, making it challenging to apply this method on a large scale. Moreover, the accuracy improvements do not sufficiently justify this approach compared to the collection and segmentation of real scenario-specific images.

In contrast, Tobin et al. [8] showed that randomized, low-fidelity synthetic images can achieve results comparable to real images when used in deep learning models for object recognition, requiring considerably less effort to generate. Thus, the goal of this work is to employ this method of synthetic data generation to evaluate the impact of synthetic images on parking space classification models and answer the research questions defined in Section 1.

The following subsections provide a detailed description of the synthetic image generation process, as well as the training and architecture of the classification models used in this study.

3.1 Synthetic Data Generation

The simulation tool used is the Unity 5 game engine, along with the Unity Perception package [10]. This package provides various environment randomization tools and allows the creation of both high- and low-fidelity simulation environments. Images are automatically captured, and context information for labeling is automatically generated through the view of a camera that can be positioned and angled in any manner.

For the low-fidelity simulated environment in this work, a script was developed to randomize the parking spaces and spacing between them, while a set of textures for the ground and 3D models of cars and trees was selected. First, the camera position and the number of iterations are chosen. Each iteration generates an image with context information, as well as segmentation and labeling data for the parking spaces, indicating whether they are occupied or empty. The script follows these steps for each iteration:

- A texture is selected and applied to the ground of the environment.
- The script defines the rows of parking spaces, their positions, spacing between the parking spots, and their rotation. Every 100 iterations, the position of all spaces is rotated by 15 degrees.
- Cars are added to the scene. Each parking space is either occupied or empty, with a 50% probability for each.
- The position and number of trees are determined, and the trees are added randomly to the scene.
- The position and intensity of the light are adjusted, changing the lighting and weather conditions of the environment. This generates environments simulating sunny, rainy, and various times of the day, such as morning, afternoon, and night.
- The image is captured by the camera, and the context information is recorded.

Several image datasets were generated, each simulating the camera position and parameters of the sub-datasets available in CNRPark-EXT [4] and PKLot [6]. In total, 12 different image datasets were created. The selected number of iterations was 1.500, with each iteration generating a synthetic image. Thus, each dataset contains 1.500 complete parking lot images and an average of 20.000 segmented and labeled parking space images. The parking spaces are segmented in the form of rotated rectangles, as this format provides the best classification accuracy according to Hochuli et al. [16]. An example of the segmented spaces is shown in Figure 4. Table 2 shows the number of segmented images for each dataset.

Figure 4: Example of parking spaces segmented into rotated rectangles. The spaces are labeled as either Occupied or Empty.

Table 2: List of segmented synthetic image quantities

Cam	Empty	Occupied	Total
CNR-cam1	14.495	14.605	29.100
CNR-cam2	5.429	5.371	10.800
CNR-cam3	9.533	9.667	19.200
CNR-cam4	12.158	12.542	24.700
CNR-cam5	15.558	15.642	31.200
CNR-cam6	9.027	9.073	18.100
CNR-cam7	12.332	12.168	24.500
CNR-cam8	11.434	11.066	22.500
CNR-cam9	8.245	8.255	16.500
PKLot-UFPR04	18.713	18.787	37.500
PKLot-UFPR05	9.433	9.567	19.000
PKLot-PUCPR	35.994	36.106	72.100
All Sets	162.351	162.849	325.200

The open-source program Fspy [22] was used to collect the position and camera parameters of each sub-dataset. Figure 5 shows a comparison between the first and the thousandth iterations of the dataset with UFPR04 parameters.

Figure 5: Comparison between the first and the thousandth iteration of the dataset with camera parameters simulating the UFPR04 subset.

3.2 Classification Models Training

The transfer learning process was chosen for model training and experiments. This process uses a classification model pre-trained on a generalist dataset. The classification layer is removed, a new classification layer is added, and fine-tuning is performed with the available datasets. This method is straightforward, reduces computational costs compared to full training, and has demonstrated satisfactory accuracy in recent studies [1, 5, 17], making it the preferred training method for this work.

The base model used in this study is MobileNetv3 [11] in its Large version, pre-trained on the ImageNet dataset [12]. This model was selected due to its widespread use in similar experiments and the balance it offers between computational cost and accuracy [1, 5, 17]. The Large version was chosen through experimentation and delivered the best results, achieving accuracy levels comparable to other studies using MobileNetv3 for parking space classification [1, 5, 17]. The input format of the images was standardized to 128×128, and a preprocessing layer was included at the model's input to resize the images, centering and cutting to the standart input size. The classification layer (last layer) was replaced to set the output channel to 2 classes, and all layers were unfrozen, with 4.6 million trainable parameters.

Data augmentation techniques were applied to increase the variability of the training data. The following augmentation methods were used:

- Rotation with a rate of 0.2.
- Horizontal flipping of the image.
- Contrast adjustment with a rate of 0.4.

Each model was trained with a batch size of 32 for 15 epochs, using the Adam optimizer with a learning rate of 0.0001 and a scheduler to reduce the learning rate by 0.1 every 7 epochs. The model with the lowest loss rate across epochs was selected. The results of this work are an average of 5 executions. For classification, the decision threshold was determined based on the Equal Error Rate in the ROC space using the validatin datasets. These parameters were chosen through experimentation and yielded the best results, with accuracy as the primary comparison metric.

4 Experiments

The experiments were developed to evaluate the accuracy of parking occupancy classification models trained under different data scenarios and to assess the impact of low-fidelity synthetic images on model accuracy. The comparison includes scenarios where models are trained exclusively with real images, synthetic images, or a combination of both. The real image datasets used were CNRPark-EXT [4] and PKLot [6], while synthetic images were generated as described in section 3. The following training scenarios were defined, employing transfer learning:

- M1 - Models trained exclusively with real images: This represents the conventional setup and serves as the baseline for comparing other experiments.
- M2 - Models trained exclusively with synthetic images: This scenario evaluates the potential of low-fidelity synthetic images to entirely replace real images in training.
- M3 - Real images + All synthetic images: Models trained with a combination of real images and all synthetic images. This scenario explores the impact of using a large amount of synthetic data to augment the real dataset, potentially increasing model variability and robustness.
- M4 - Real images + Synthetic images (Target scenario): Models trained with real images and synthetic images generated specifically for a target scenario (camera). Each target scenario receives its own specialized model. This setup allows for comparison with the method proposed in Hochuli et al. [16], where adding a small number of specific scenario images significantly improved model accuracy on that target scenario.

The real image datasets (PKLot and CNRPark-EXT) were sorted by the day of capture and split into 70% for training and 30% for validation to avoid bias. For models trained exclusively with synthetic images, the data was randomly split into 70% for training and 30% for validation. In scenarios combining real and synthetic images, the synthetic images were added to the training subset of the real dataset (70%), either using all images (M3) or only scenario-specific ones (M4). A total of 17 models were trained and tested across the experimental scenarios.

The evaluation metric used was average accuracy, with a focus on comparing scenarios to understand the impact of synthetic images on model performance and generalization, relative to models trained with real images. Tables 3 and 4 present the accuracies obtained with the M1, M2, and M3 models, considering the CNRPark-Ext and PKLot as the basis of the experiments. The highest accuracy in the PKLot basis experiments was achieved by the model trained exclusively with real images (M1), with an average accuracy of 95.29%. In contrast, the model trained with a combination of real images and a general set of synthetic images (M3) did not exhibit a significant change in accuracy on the PKLot dataset but showed a substantial improvement in the CNRPark-EXT training scenario, achieving accuracy of 96,49%, an increase of approximately 2% comparing with the model trained exclusively with CNRPark-EXT real images. Meanwhile, the model trained exclusively with synthetic images (M2) showed significantly lower accuracy in both scenarios.

Table 3: Accuracies obtained by testing the M1, M2, and M3 model types with CNRPark-EXT.

Train Set	Test CNRPark-EXT
PKLot (M1)	95,29%
Synthetic Images Only (M2)	85,96%
PKLot + Synthetic Images CNRPark-EXT (M3)	95,05%

Table 4: Accuracies obtained by testing the M1, M2, and M3 model types with PKLot.

Train Set	Test PKLot
CNRPark-EXT (M1)	94,57%
Synthetic Images Only (M2)	85,44%
CNRPark-EXT + Synthetic Images PKLot (M3)	96,49%

Tables 5 and 6 present the results of the M4 models, which were trained using a combination of the real image training set and synthetic images created to simulate a target scenario. For each test, a separate M4 model was used, trained with the combination of the basis dataset and the synthetic images simulation that target scenario. In most cases, the results of the M4 models did not show a significant improvement compared to the M1 models (trained only with real images). However, a considerable increase in accuracy was observed for the PKLot model when using synthetic data simulating camera 9, as well as an improvement in the CNRPark-EXT model when augmented with synthetic PUCPR images.

4.1 Analysis of Results

The chosen base model and transfer learning method demonstrated good generalization capability and achieved results close to and comparable with other works in the state of the art [5] [2] [13] addressing the same problem. Thus, the results obtained with M1 models provide a strong baseline for evaluating the performance of low-fidelity synthetic images. From this, the following conclusions can be drawn:

P1 - How does the combination of real and low-fidelity synthetic images affect the generalization of parking space classification models? The M3 model for PKLot, compared to M1, showed no improvement, with results remaining close to those achieved by M1. In contrast, the M3 model for CNRPark-EXT achieved a considerable accuracy increase compared to its respective M1. This behavior in PKLot can be attributed to the large number of real images in the training set, which helps the model converge to a stable solution, minimizing the impact of additional synthetic images. For CNRPark-EXT, which contains fewer real images, the addition of synthetic images more significantly improved model accuracy and generalization.

P2 - How do low-fidelity synthetic images perform when applied to fine-tuning parking space classification models for a specific target scenario? The results from M4 models indicate that low-fidelity synthetic images may not be the best choice for fine-tuning a generic model to a specific scenario. Most tests showed lower average accuracy, with improvements in only 3 out of 12 cases. Compared to the experiments in Hochuli et al. [16], real images remain a better option for fine-tuning, despite requiring more effort to collect than generating low-fidelity synthetic images.

These findings suggest that the impact of low-fidelity synthetic images on parking space classification model performance is minimal. They are not an ideal choice for fine-tuning models to specific scenarios and only improve generalization when the quantity of available real images for training is relatively small compared to the synthetic images. In such cases, adding synthetic images introduces greater variability in training features, aiding model classification.

P3 - Is it possible to surpass the accuracy of models trained with real images by training models exclusively with low-fidelity synthetic images? No. The accuracies obtained with M2 models were significantly lower than those of M1 models. This may be due to low-fidelity synthetic images lacking realistic features necessary to solve the parking space classification problem, such as faithfully reproducing image quality and details of weather and lighting conditions. A better approach for creating a model trained exclusively with synthetic images that can compete with the state of the art might be the one described in Tremblay et al. [9], which involves combining highly randomized synthetic images with photorealistic synthetic images.

The limitations of the developed protocol for generating synthetic images and the quality of the images themselves must be considered. The lack of variability in weather conditions, lighting, textures, and context in the images likely impacted the experiment. Although the protocol was designed to reduce human effort in collecting and generating parking space images, a more realistic scenario could potentially improve models for both generalized and scenario-specific cases whithout increasing the human effort.

5 Conclusion

This study investigated the impact of using low-fidelity synthetic images in combination with real images to train parking occupancy classification models, exploring their viability as a complement and maybe an substitute for real data. The proposal addressed two main questions: the impact of synthetic images on model accuracy using a comination of real images and synthetic ones and the feasibility of surpassing the performance of models trained exclusively on real data.

Table 5: Accuracies obtained with models trained using combination of real images and synthetic images from specific scenarios from CNRPark-EXT.

Train Set	cam1	cam2	cam3	cam4	cam5	cam6	cam7	cam8	cam9	Mean
PKLot (M1)	94,35%	97,17%	92,71%	96,51%	95,23%	94,64%	95,2%	97%	95,38%	95,39%
PKLot + Synthetic Scenario (M4)	93%	97,53%	91,8%	96,08%	94,84%	94,34%	94,2%	96,8%	97,05%	95,07%

Table 6: Accuracies obtained with models trained using combination of real images and synthetic images from specific scenarios from PKLot.

Train Set	UFPR04	UFPR05	PUCPR	Mean
CNRPark-EXT (M1)	91,88%	95,76%	94,83%	94,16%
CNRPark-EXT + Synthetic Scenario (M4)	90,98%	94,81%	98,05%	94,61%

The experiments demonstrated that low-fidelity synthetic images can complement real datasets in contexts with limited data, as evidenced by the CNRPark-EXT dataset, where the model trained with the combination of CNRPark-EXT and synthetic images showed an accuracy increase compared to the model trained exclusively with real images. However, their effectiveness diminished in scenarios with a larger availability of real data, such as in the PKLot dataset. Additionally, the models that used synthetic data for specialization to a target scenario, performed worse than models trained solely on real images, reinforcing that low-fidelity synthetic images are better suited for generalization rather than specialization.

On the other hand, models trained exclusively with synthetic images, showed significantly lower accuracy than the ones trained with only real images. This indicates that, despite the potential for cost and effort reduction in data generation, the low fidelity of synthetic data limits their ability to adequately represent real-world conditions, such as weather variations, textures, and lighting.

Based on the results obtained, it is concluded that low-fidelity synthetic images play an important role as a complement in scenarios with scarce data but are insufficient to replace real data. Future work could explore more advanced data generation approaches, such as combining photorealistic and highly randomized images, as well as improved domain randomization methods.

This study contributes to the field of computer vision by highlighting the limits and possibilities of using synthetic data in parking space classification, paving the way for further research into the use of more sophisticated generation techniques to overcome the challenges identified.

References

[1] Paulo Ricardo Lisboa de Almeida, Jeovane Honório Alves, Rafael Stubs Parpinelli, and Jean Paul Barddal. A systematic review on computer vision-based parking lot management applied on public datasets. *Expert Systems with Applications*, 198:116731, February 2022. URL https://www.sciencedirect.com/science/article/pii/S0957417422002032.

[2] Vijay Paidi, Hasan Fleyeh, Johan Håkansson, and Roger G. Nyberg. Smart parking sensors, technologies and applications for open parking lots: a review. *IET Intelligent Transport Systems*, 12(8):735–741, 2018. doi: https://doi.org/10.1049/iet-its.2017.0406. URL https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-its.2017.0406.

[3] Paulo R.L. Almeida, Luiz S. Oliveira, Eunelson Silva, Alceu Britto, and Alessandro Koerich. Parking space detection using textural descriptors. In *2013 IEEE International Conference on Systems, Man, and Cybernetics*, Manchester, UK, October 2013.

[4] Giuseppe Amato, Fabio Carrara, Fabrizio Falchi, Claudio Gennaro, Carlo Meghini, and Claudio Vairo. Deep learning for decentralized parking lot occupancy detection. *Expert Systems with Applications*, 72:327–334, October 2017. URL https://www.sciencedirect.com/science/article/pii/S095741741630598X.

[5] Andre G. Hochuli, AJean Paul Barddal, Paulo R. L. de Almeida, Gillian Cezar Palhano, and Leonardo Matheus Mendes. Deep single models vs. ensembles: Insights for a fast deployment of parking monitoring systems. In *International Conference on Machine Learning and Applications (ICMLA)*, pages 1379–1384, Florida, USA, December 2023.

[6] Paulo R.L. Almeida, Luiz S. Oliveira, Alceu S. Britto Jr., Eunelson J. Silva Jr., and Alessandro L. Koerich. Pklot - a robust dataset for parking lot classification. *Expert Systems with Applications*, 42(11):4937–4949, February 2015. URL https://www.sciencedirect.com/science/article/pii/S0957417415001086.

[7] Hadi Keivan Ekbatani, Oriol Pujol, and Santi Segui. Synthetic data generation for deep learning in counting pedestrians. In *International Conference on Pattern Recognition Applications and Methods (ICPRAM*, pages 318–323, Porto, Portugal, January 2017.

[8] Josh Tobin, Rachel Fong, Alex Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel. Domain randomization for transferring deep neural networks from simulation to the real world, 2017.

[9] Jonathan Tremblay, Thang To, Balakumar Sundaralingam, Yu Xiang, Dieter Fox, and Stan Birchfield. Deep object pose estimation for semantic robotic grasping of household objects. In *Conference on Robot Learning (CoRL)*, pages 307–316, Zurich - Switzerland, Setember 2018.

[10] Steve Borkman, Adam Crespi, Saurav Dhakad, Sujoy Ganguly, Jonathan Hogins, You-Cyuan Jhang, Mohsen Kamalzadeh, Bowen Li, Steven Leal, Pete Parisi, Cesar Romero, Wesley Smith, Alex Thaman, Samuel Warren, and Nupur Yadav. Unity perception: Generate synthetic data for computer vision. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Vancouver, BC, Canada, Setember 2017.

[11] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, Quoc V. Le, and Hartwig Adam. Searching for mobilenetv3, 2019.

[12] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009. doi: 10.1109/CVPR.2009.5206848.

[13] Ratko Grbić and Brando Koch. Automatic vision-based parking slot detection and occupancy classification. *Expert Systems with Applications*, 225:120147, April 2023. URL https://www.sciencedirect.com/science/article/pii/S0957417423006498.

[14] Vighnesh Dhuri, Afzal Khan, Yash Kamtekar, Dhananjay Patel, and Ichhanshu Jaiswal. Real-time parking lot occupancy detection system with vgg16 deep neural network using decentralized processing for public, private parking facilities. *2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*, pages 1–8, 2021. URL https://api.semanticscholar.org/CorpusID:234952384.

[15] Sherzod Nurullayev and Sang-Woong Lee. Generalized parking occupancy analysis based on dilated convolutional neural network. *Sensors*, 19(2), 2019. ISSN 1424-8220. doi: 10.3390/s19020277. URL https://www.mdpi.com/1424-8220/19/2/277.

[16] Andre G. Hochuli, Alceu S. Britto Jr, Paulo R. L. de Almeida, and Fabio M. C. Cagni Williams B. S. Alves. Evaluation of different annotation strategies for deployment of parking spaces classification systems. In *International Joint Conference oN Neural Networks (IJCNN)*, pages 1–8, Padua, Italy, July 2022.

[17] Paulo L. Alves, Andre G. Hochuli, Luiz E. de Oliveira, and Paulo R. L. de Almeida. Optimizing parking space classification: Distilling ensembles into lightweight

classifiers. In *International Conference on Machine Learning and Applications (ICMLA)*, pages 1379–1384, Florida, USA, December 2023.

[18] T. To, J. Tremblay, D. McKay, Y. Yamaguchi, K. Leung, A. Balanon, J. Cheng, , and S. Birch-field. Ndds: Nvidia deep learning dataset synthesizer,. https://github.com/NVIDIA/Dataset_Synthesizer., 2018. Acessado em 07/12/2023.

[19] Marc Tschentscher, Ben Pruß, and Daniela Horn. A simulated car-park environment for the evaluation of video-based on-site parking guidance systems. In *2017 IEEE Intelligent Vehicles Symposium (IV)*, pages 1571–1576, 2017. doi: 10.1109/IVS.2017.7995933.

[20] Daniela Horn and Sebastian Houben. Evaluation of synthetic video data in machine learning approaches for parking space classification. In *2018 IEEE Intelligent Vehicles Symposium (IV)*, pages 2157–2162, Changshu, China, October 2018. doi: 10.1109/IVS.2018.8500453.

[21] Stefan Hinterstoisser, Olivier Pauly, Hauke Heibel, Martina Marek, and Martin Bokeloh. An annotation saved is an annotation earned: Using fully synthetic training for object instance detection. *CoRR*, abs/1902.09967, 2019. URL http://arxiv.org/abs/1902.09967.

[22] Stuffmatic. fspy - open source still image camera matching. https://fspy.io/, 2023. Acessado em 12/12/2023.