Parking Spot Segmentation Using Deep Learning Techniques

Marcin Garnowski
Gdańsk University of Technology,
Faculty of Electronics, Telecommunications
and Informatics, Multimedia Systems
Department
Gdańsk, Poland *s172578@student.pg.edu.pl*

Jakub Śmigielski
Gdańsk University of Technology,
Faculty of Electronics, Telecommunications
and Informatics, Multimedia Systems
Department
Gdańsk, Poland *s184742@student.pg.edu.pl*

Szymon Zaporowski
Gdańsk University of Technology,
Faculty of Electronics, Telecommunications
and Informatics, Multimedia Systems
Department
Gdańsk, Poland *szyzapor@pg.edu.pl*

Abstract

Existing computer vision models are developing rapidly and obtaining successful results, but expensive and time-consuming solutions are still used in many cases and applications. Currently, proposed parking management solutions based on computer vision and artificial intelligence are not sufficiently automated and general. They often assume simplifications and specific conditions. This project provides an overview of state-of-the-art in occupancy detection methods and describes the original solution separated into modules. The main advantages are versatility, lack of requirements for the parking structure, high resistance to external conditions, and openness to extension with additional functionalities.
Keywords: Detection, segmentation, computer vision, parking lots occupancy, parking management system.

1. Introduction

1.1. Object detection

Object detection is an important part of the field of computer vision. The goal is to identify and localize all instances of a given object. The greater difficulty compared to other computer vision tasks, such as image classification, lies in the number of objects, distortions, and obstacles. With the development of machine learning, solutions using artificial neural networks began to be created. One of the first solutions was the Region-based Convolutional Neural Network (RCNN) [11], followed by FastRCNN [10] and FasterRCNN [19] belonging to the Two Stage Detectors category. The first processing generates proposals for locations and classes and the next processing is used to refine them and make final predictions. A solution belonging to One Stage Detectors is You Only Look Once (YOLO) [18]. The main idea of this algorithm is to divide the image into $S \times S$ grid and then, in each cell, detect objects whose centers fall within that cell. This approach achieves high accuracy in real-time.

1.2. Segmentation

Significant development followed the introduction of convolutional neural networks, leading to an approach called semantic segmentation. Based on the extracted features, each pixel is classified. The next step in development was instance segmentation, which distinguished between individual objects of the same classes. The breakthrough was transformer-based image segmentation. It gains an advantage over the CNN-based approach in the case of complex images by analyzing the entire image at once. An example is the Segment Anything Model (SAM) [16]. It allows automatic segmentation of the entire image or individual objects specified by a bounding box or point. The problem, however, is that SAM's output is class-agnostic.

1.3. Parking systems

The most common method of detecting parking space occupancy is to use various types of sensors [21]. These solutions have two main problems: price and the requirement to predefine parking spaces, which is not always possible. Another solution that is devoid of these drawbacks is to use computer vision methods for occupancy detection based on camera images, which can cost less than a single sensor. Among several approaches to this method, detection-based systems have demonstrated the best performance in parking space identification.

These systems typically rely on publicly available datasets with annotated parking spaces, such as CARPK [12], PKLot [9], and CNRPark+EXT [2], to train neural networks with the possibility of fine-tuning or retraining using custom datasets. While these datasets contain thousands of annotated images under various lighting and weather conditions, they assume static, clearly defined parking spots and simplify the task to classification of cropped image regions as either vacant or occupied. This limits generalization to more dynamic and unstructured environments, such as those with unmarked or irregular parking layouts.

Moreover, image annotation and incorporating diverse weather and lighting conditions, such as rain or snow, pose challenges due to the time required to capture and annotate data across different seasons. Additionally, training requires significant computational resources, and even minor modifications often necessitate retraining the entire model, further adding to the complexity and expense. The accuracy of detection-based systems is heavily influenced by camera placement, with a top view being ideal to maximize visibility and minimize occlusions. There is a clear need for alternative datasets and methodologies that support segmentation-based occupancy detection without relying on predefined parking spot layouts. Automated and adaptable solutions in this context could be a major breakthrough for real-world parking occupancy detection.

2. Related Work

In [2], a method was introduced for parking space occupancy detection using the mAlexNet model, a variant based on AlexNet. The authors used their own CNRPark dataset for training. Images from 9 cameras were divided into smaller chunks and then classified into free or occupied space. Each camera was connected to a Raspberry Pi, and classification results were sent to an external server. A key advantage of this decentralized architecture is the distribution of computational load across multiple devices, reducing the server load, which is often a system bottleneck. The reported processing and transmission time was approximately 15 seconds.

Image fragment classification has become a standard approach for parking space occupancy detection. In subsequent years, variations using different types of convolutional networks achieved classification accuracies of around 96-99%. Additionally, enhancements in system components introduced new functionalities. In [23], 2D detection was extended to 3D detection as described in [7], enabling more precise localization of vehicles and improved robustness against partial occlusions. The authors also implemented a Multinomial Logistic Regression Model that considered three adjacent parking spaces simultaneously when determining occupancy status. This approach increased system resilience to classification errors, particularly under challenging conditions such as nighttime operations. The output of this system includes not only the identification of occupied parking spaces but also the detection of improperly parked vehicles. Such information can be further utilized to highlight potential traffic safety hazards.

Another extension of the parking occupancy system expands the classification beyond the binary labels to include categories such as "not very occupied" and "very occupied," as described in [5]. This improvement mitigates errors where a parking space is incorrectly classified as a fully occupied space, but smaller vehicles are able to park there. Instead of relying on datasets annotated with real-world images, the authors created a synthetic dataset with models of vehicles and manually constructed parking lots. Edge detection was used to segment parking spaces. This setup enabled controlled verification and error analysis. Later adaptation to real-world datasets showed only a slight drop in classification performance.

Beyond user-oriented improvements, several techniques were also proposed to enhance the overall system performance. In the context of large environments with visual obstructions, the deployment of multiple cameras offers significant advantages. The study presented in [3] provides a comprehensive analysis of three algorithms designed for detecting common features between camera images. The comparative results assist in selecting an appropriate method for effective image merging. Another integrated and systematic approach is presented in [14]. In this method, parking coordinates and space occupancy data are initially collected and normalized. Subsequently, the parameters of the Artificial Bee Colony (ABC) algorithm are optimized to facilitate the partitioning of available space into individual parking spots.

3. Methodology

3.1. Dataset

For this study, a custom validation dataset was created to support the model selection and evaluation. Images were remotely collected from two provided cameras pointed at the parking lot of the Gdańsk University of Technology. Positioned approximately 20 meters high inside a building, the cameras had a partially obstructed view due to window and wall elements. Using the custom tool, a total of 328 images were labeled, marking the exact outlines of the vehicles in the parking area. Each marked vehicle also includes vehicle type information and a unique ID to track them between cameras. The data was collected in October and November from 9 a.m. to 6 p.m., capturing periods of peak activity and diverse conditions such as sunlight, overcast, and snowfall. According to the initial concept, the cameras were supposed to provide a view of an unmarked parking. However, due to the unavailability of the intended data and the lack of suitable public resources, this approach could not be fully examined. For this reason, the data with marked parking spots was used, but treated as unmarked at each stage of the system development.

3.2. Evaluation

As part of the occupancy detection system conceptualization, the resulting dataset was used as a validation set for model selection. Utilization of detailed object masks instead of parking spots offered better comparison. The key consideration was whether to prioritize fast and accurate object detection or to use segmentation models capable of generating detailed masks. Three models were tested: YOLOv11 [15], SAM 2.1 [17] and FastSAM [22]. In addition, a pipeline that combines YOLO detections with subsequent SAM segmentation was proposed. Moreover, the parameters of YOLO have been optimized. The results are shown in Table 1.

Table 1. Comparison of models by weather conditions.

Model	**Sunny**		**Dark**		**Snow**	
	Precision	Recall	Precision	Recall	Precision	Recall
YOLOv11	**0.984**	0.591	**0.964**	0.681	**0.986**	0.553
SAM 2.1	0.675	0.770	0.556	0.582	0.688	0.691
FastSAM	0.343	**0.999**	0.353	**0.975**	0.390	**0.988**
YOLOv11 + SAM 2.1	0.913	0.983	0.903	0.971	0.898	0.972

3.3. Parking lot segmentation

One of the primary objectives is to design a system that can be applied to various types of parking lots, making it necessary to use a different solution than manually marking parking spaces. Assuming drivers follow traffic rules, allowed parking areas can be identified as locations where cars remain stationary for extended periods throughout the day. The following approach was adopted: initially, car detection was performed using YOLOv11, followed by segmentation using SAM 2.1. This process is repeated in images captured at arbitrary intervals across a sufficiently large and diverse dataset to ensure a comprehensive identification of potential parking spots. The resulting masks are assigned a transparency level depending on the data used and are layered onto a blank image, producing an effect similar to long exposure, as shown in Figure 1.

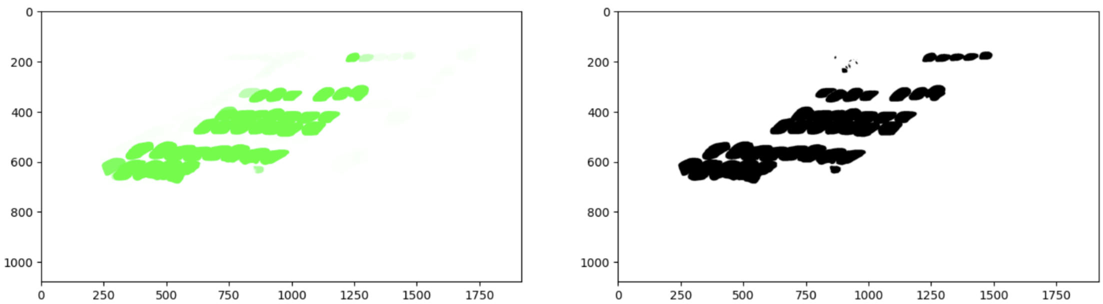

Fig. 1. Allowed area mask. On the left, layered masks with a transparency value of $\frac{1}{\text{number of images}}$. On the right, the mask after applying a threshold of 0.05.

Then, a transparency threshold is applied to effectively filter out single detections, such as passing vehicles or false positive detections. In cases where forbidden areas, such as sidewalks or green spaces, exist near parking spots, these regions may be partially covered by parked cars. SAM's capability to use additional input allows the model to identify restricted areas by manually highlighting them in the image and, similarly to cars, adding their corresponding masks with a specific transparency value. This approach offers significant flexibility in customizing the segmentation process, enabling the exclusion of specific parking spots or regions based on predefined criteria. As the last step, the forbidden area mask is subtracted from the allowed area mask, producing the final allowed parking area mask.

3.4. Transformation

To finally make effective use of the camera images, it is necessary to get rid of the perspective and transform them into a top projection. To be able to create a transformation matrix [1], accurate geographic coordinates are needed. For this purpose, the Geoportal was used, and points were selected that are additionally marked with symbols and fall within the field of view of the cameras. In order to simultaneously make the result of the transformation correspond to a known real size, the EPSG:2180 coordinate system is used. The corresponding points were then marked on the images using the custom marking tool. To create the transformations, the homography function from the opencv-python package was used. The resulting masks of allowed and forbidden areas were transformed and subtracted, as shown in Figure 2.

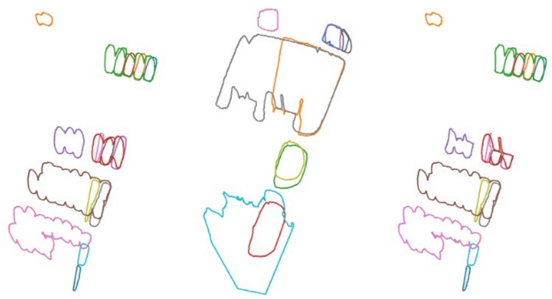

Fig. 2. The results of transformation of allowed, forbidden and final areas

Execution of this part in the way described above is the main non-automated part of the system. To automate the process of creating a bird's eye view transformation, ASIFT (Affine Scale-Invariant Feature Transform) [20] can be used. The algorithm searches both given images for similar features and connects points identified in this way. ASIFT enhances SIFT by simulating possible viewpoint changes, making it especially suitable for the case of satellite and camera images. This method was used in the system, creating a known size grid over the parking area [8]. The results presented by the authors are not satisfactory. They show evident errors, such as combining different points in the first image with a single point in the other one or creating clusters of points and vast areas without points, which can cause further errors. A parking lot is usually too uncharacteristic space to use ASIFT effectively. The best solution may be to keep this step unautomated and improve the density of the point grid. For example, by constructing a device that could be detected in camera images and measure its position using GPS.

3.5. Genetic packing algorithm

To create a fast method for determining the locations of individual spaces, a genetic algorithm approach was chosen for this purpose. There are well-known efficient solutions to similar problems, but they are greatly simplified compared to this issue: packing multiple polygons into a rectangle [13], packing into a convex polygon [4], [6]. Therefore, the following custom solution was designed and implemented. The population represents a set of grid arrangements of rectangles with the size of parking spaces with different values of the features: offset vector and rotation angle. The population G is denoted by:

$$G = \{(\vec{v}_i, a_i) \mid i = 1, 2, \ldots, n\}, \tag{1}$$

where $\vec{v}_i$ denotes the i-th candidate's offset vector and a_i the rotation angle. The fitness function is described by the formula:

$$f = \sum_{p \in \mathcal{P}} \left(\left(\frac{|p \cap A|}{|p|} \right)^2 + \left\lfloor \frac{|p \cap A|}{|p|} \cdot t \right\rfloor \right) \tag{2}$$

Here $\mathcal{P}$ denotes the set of rectangles, p the rectangle from the set $\mathcal{P}$, A the allowed area polygon and t the threshold. The genetic packing algorithm is as follows. An initial generation of size N is randomly generated, where $|\vec{v}_i| \in [0, \sqrt{2}]$, $a \in [0, \frac{\pi}{2}]$. The entire population is sorted by fitness function. Then a new population of the same size is created. Parents are selected from the sorted list with a probability of the i-th element of the list:

$$p_i = \begin{cases} \frac{N-i+1}{N(N+1)}, & \text{if } i \leq \frac{N}{2} \\ 0, & \text{if } i > \frac{N}{2} \end{cases} \tag{3}$$

Crossing is performed by copying the vector of the first parent and the rotation angle of the second parent. These features are then mutated, changing them by values: $|\vec{v}_i| \in [0, 0.25]$, $a_m \in [-\frac{\pi}{6}, \frac{\pi}{6}]$. This is repeated for m iterations. This approach was accepted on the basis of visual evaluation due to the lack of other solutions considered suitable.

4. Results and Discussion

The proposed segmentation system effectively handles parking lot analysis without the simplifications commonly employed in existing solutions. Deployed pipeline consistently delivered high precision and recall values (0.898-0.913 precision and 0.971-0.983 recall). While YOLOv11 alone demonstrated excellent precision (0.964-0.986), its recall was significantly lower (0.553-0.681), indicating that it missed many vehicles. Conversely, FastSAM achieved nearly perfect recall (0.975-0.999) but suffered from low precision (0.343-0.390), generating numerous false positives. The performance of the combined approach demonstrates that leveraging the strengths of both models yields more robust results across variable conditions.

The genetic packing algorithm proved effective for parking space allocation, with the selected parameters (N=100, m=20, t=0.8) providing an optimal balance between computational efficiency and accuracy. The transformation methodology utilizing geographical coordinates and homography successfully merged multiple camera views, creating a cohesive top-down representation of the parking area. This approach eliminated the need for cameras to be positioned directly overhead, making the system more flexible in real-world deployments.

A notable strength of described approach is the automated identification of allowed and forbidden parking areas through the transparency-based masking technique. By overlaying multiple vehicle masks captured over time and applying appropriate thresholds, the system effectively distinguished between legitimate parking areas and transient vehicle positions. This method eliminates the requirement for pre-marked parking spaces, significantly enhancing the system's versatility across different parking environments.

The complete system iteration, from image capture through detection, segmentation, transformation, and space allocation, requires approximately 30 seconds, which is acceptable for a real-time parking management application. However, this processing time could potentially be further optimized, particularly in scenarios involving distant objects that occupy small portions of the image. In such cases, the confidence parameters needed to be substantially different from those typically used in more common detection scenarios, requiring careful tuning.

The results demonstrate that the proposed approach offers a comprehensive solution to parking space segmentation without the limitations that characterize existing methods. By eliminating the need for predefined parking spaces and employing a combination of state-of-the-art detection and segmentation techniques, the system provides a flexible and robust foundation for practical parking management applications.

5. Conclusion and Future Work

This research presents a novel approach to parking space management through the integration of advanced computer vision techniques, eliminating the requirement for predefined parking space markings and model training on labeled datasets. The system successfully employs a dual-model architecture combining YOLOv11 and SAM 2.1 to achieve robust vehicle detection and segmentation across varying environmental conditions, while implementing a genetic algorithm-based allocation methodology for efficient parking space management.

One promising direction for optimization is the implementation of a cascaded architecture wherein YOLO detection outputs are utilized to constrain the operational domain of the segmentation model. By limiting segmentation processing to regions defined by bounding boxes, computational resources could be allocated more efficiently, potentially enabling the deployment of more sophisticated segmentation algorithms without prohibitive latency penalties. However, technical integration challenges currently impede seamless model fusion, necessitating further investigation into interoperability frameworks.

The distortion artifacts introduced during homographic transformation represent another area requiring refinement. Implementation of post-processing algorithms to correct for perspective distortion in segmented vehicle masks could substantially improve spatial accuracy and potentially eliminate the need for manual delineation of forbidden parking areas. Specifically, implementing non-linear transformation correction algorithms could preserve the geometric integrity of vehicle representations in the bird's-eye projection.

A particularly innovative direction for future work involves the development of adaptive parking space recognition. By leveraging temporal information from long-exposure composites, the system could autonomously identify structural changes in parking configurations and dynamically update the allowed parking area parameters. This self-calibrating mechanism would enhance the system's adaptability to evolving parking environments automatically.

The current implementation demonstrates promising results on a single parking facility dataset. To establish the generalizability of presented approach, extensive validation across diverse parking infrastructures is essential. This cross-validation should encompass variations in layout complexity, camera positioning constraints, and environmental conditions to comprehensively assess system robustness.

Additionally, exploring the integration of semantic understanding capabilities could enhance the system's functionality beyond binary occupancy detection. Classifying vehicles by type, identifying improper behaviors, and recognizing potential security concerns represent valuable extensions to the core technology that merits investigation in subsequent research iterations.

In conclusion, while the current system successfully addresses the fundamental challenges of markerless parking space management, numerous opportunities exist for refinement and functional expansion that could significantly enhance its practical utility in real-world deployments.

Acknowledgements

This work was partly supported by the Polish National Centre for Research and Development (NCBR) through the European Regional Development Fund titled "INFOLIGHT-Cloud-Based Lighting System for Smart Cities" under Grant POIR.04.01.04/2019.

References

[1] Abbas, A., Zisserman, A.: A geometric approach to obtain a bird's eye view from an image. arXiv preprint arXiv:1905.02231 (2019), iCCV Workshop "Geometry Meets Deep Learning" 2019

[2] 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, pp. 327–334 (2017)

[3] Asy'ari, M.F., Fatichah, C., Suciati, N.: Parking space availability detections from two overlapping cameras using yolov5 and image stitching methods. International Journal of Intelligent Engineering and Systems 16(4), pp. 276–288 (2023)

[4] Birgin, Martinez, N., Ronconi: Orthogonal packing of rectangular items within arbitrary convex regions by nonlinear optimization. Computers Operations Research (2005)

[5] Carrera García, J.M., Recas Piorno, J., Guijarro Mata-García, M.: Expert system design for vacant parking space location using automatic learning and artificial vision. Multimedia Tools and Applications 81, pp. 38661–38683 (2022)

[6] Chazelle: The polygon containment problem. Advances in Computing Research (1983)

[7] Chen, X., Kundu, K., Zhu, Y., Ma, H., Fidler, S., Urtasun, R.: 3d object proposals using stereo imagery for accurate object class detection. arXiv preprint arXiv:1608.07711 (2016)

[8] Dacayan, Ponte, H., Kwak: Utilizing a spatial grid for automated parking space vacancy detection. International Conference on Computational Science and Computational Intelligence (2023)

[9] De Almeida, P.R., Oliveira, L.S., Britto, A.S., Silva, E.J., Koerich, A.L.: PKLot – A robust dataset for parking lot classification. Expert Systems with Applications 42(11), pp. 4937–4949 (2015)

[10] Girshick: Fast r-cnn. Computer Vision and Pattern Recognition (2015)

[11] Girshick, Donahue, D., Malik: Region-based convolutional networks for accurate object detection and segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence (2015)

[12] Hsieh, L., Hsu: Drone-based object counting by spatially regularized regional proposal network. International Conference on Computer Vision (2017)

[13] Jakobs: On genetic algorithms for the packing of polygons. European Journal of Operational Research (1993)

[14] Janowski, A., Hüsrevoğlu, M., Renigier-Bilozor, M.: Sustainable parking space management using machine learning and swarm theory—the spark system. Applied Sciences 14(24), pp. 12076 (2024)

[15] Khanam, R., Hussain, M.: Yolov11: An overview of the key architectural enhancements. arXiv preprint arXiv:2410.17725 (2024)

[16] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollár, P., Girshick, R.: Segment anything. arXiv preprint arXiv:2304.02643 (2023)

[17] Ravi, N., Gabeur, V., Hu, Y.T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., Mintun, E., Pan, J., Alwala, K.V., Carion, N., Wu, C.Y., Girshick, R., Dollár, P., Feichtenhofer, C.: Sam 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714 (2024)

[18] Redmon, Divvala, G., Farhadi: You only look once: Unified, real-time object detection. Computer Vision and Pattern Recognition (2016)

[19] Ren, He, G., Sun: Faster r-cnn: Towards real-time object detection with region proposal networks. Computer Vision and Pattern Recognition (2016)

[20] Yu, Morel: Asift: An algorithm for fully affine invariant comparison. Image Processing On Line (2011)

[21] Zhang, Z., Tao, M., Yuan, H.: A parking occupancy detection algorithm based on amr sensor. IEEE Sensors Journal 15(2), pp. 1261–1269 (2015)

[22] Zhao, X., Ding, W., An, Y., Du, Y., Yu, T., Li, M., Tang, M., Wang, J.: Fast segment anything (2023)

[23] Zheng, Y., Zhang, W., Wu, X., Zhao, B.: 3d bounding box proposal for on-street parking space status sensing in real world conditions. Computer Modeling in Engineering & Sciences 119(3), pp. 559–576 (2019)