

Deep Learning Based Parking Vacancy Detection for Smart Cities

Qiangwen Xu
CIS Department
Fordham University
New York, NY
qxu47@fordham.edu

Mengyao Sun
CIS Department
Fordham University
New York, NY
msun54@fordham.edu

Bailing Fu
CIS Department
Fordham University
New York, NY
bfu6@fordham.edu

Yijun Zhao
CIS Department
Fordham University
New York, NY
yzhao11@fordham.edu

Abstract

Parking shortage is a major problem in modern cities. Drivers cruising in search of a parking space directly translate into frustration, traffic congestion, and excessive carbon emission. We introduce a simple and effective deep learning-based parking space notification (PSN) system to inform drivers of new parking availabilities and re-occupancy of the freed spaces. Our system is particularly designed to target areas with severe parking shortages (i.e., nearly all parking spaces are occupied), a situation that allows us to convert the problem of detecting parking vacancies into recognizing vehicles leaving from their stationary positions. Our PSN system capitalizes on a calibrated Mask R-CNN model and a unique adaptation of the IoU concept to track the changes of vehicle positions in a video stream. We evaluated PSN using videos from a CCTV camera installed at a private parking lot and publicly available YouTube videos. The PSN system successfully captured all new parking vacancies arising from leaving vehicles with no false positive detections. Prompt notification messages were sent to users via cloud messaging services.

1. Introduction

Cars have become an indispensable means of transportation for modern life. With the increased number of privately owned vehicles, parking space shortage has become a common predicament in metropolitan areas. When drivers cannot find a parking space, they will slowly cruise the city, generating large amounts of exhaust emissions and creating traffic congestion. Simple solutions, such as adding more parking spaces, are not feasible due to the high cost and limited supply of commercial real estate in cities. Thus, it is essential to develop intelligent methods to

help drivers quickly locate empty parking spaces.

Existing technologies for parking vacancy detection can be classified into sensor-based or image-based categories. The sensor-based methods require many sensor units, such as smart parking meters [1], to cover the entire parking areas. Because a sensor can typically monitor only one parking space, it is cumbersome and costly to install a complete sensor-based system in a large city. Nevertheless, high reliability and deployment readiness could make the technology a favored choice for small parking areas.

The alternative image-based detection technology requires only a few cameras to successfully monitor an entire parking area, which significantly reduces the cost of installation and maintenance. Additionally, surveillance cameras have already been installed in many public places. Thus, an image-based parking detection system can potentially leverage existing camera devices and video signals, which is financially and environmentally efficient. In the past decade, image-based approaches have further attracted a great amount of attention due to the unprecedented success of deep learning [2] algorithms in computer vision. In particular, deep neural networks have produced results comparable, and in some cases superior, to human experts [3–5] in various tasks such as object recognition [6], image captioning [7] and image restoration [8]. We conduct a brief survey on deep learning-based studies related to our study in Section 2.

In this paper, we present a parking space notification (PSN) system that informs drivers of parking locations freed by leaving vehicles. Our system monitors a video stream and generates alerts when new vacancies become available and when the freed spaces are reoccupied. PSN is most applicable to crowded metropolitan areas where one can safely assume nearly all parking spaces are taken. As a result, we can convert the problem of detecting parking vacancies into recognizing vehicles

leaving from their stationary positions.

A key component of the PSN system is an object localization model, (V-MRCNN, Section 3.3), which adopts the Mask R-CNN [9] architecture due to its flexibility and computational efficiency. V-MRCNN was trained to detect cars, trucks, and buses using the COCO dataset [10]. To track leaving vehicles, we utilize the IoU (Section 3.2) ratio between a car and its corresponding parking space. When any parking space becomes unoccupied, an SMS message will be sent to the users using cloud messaging services. We illustrate the details of our algorithm in Section 3.

The main contribution of our study is an effective method to convert parking vacancy detection into a car object detection task using a unique adaptation of the IoU concept. While IoU is a known concept, in our model it serves a purpose that is different from its conventional role of evaluating performance of an object detection model. Specifically, we measure the IoU ratio between the bounding boxes of two different objects (i.e., the vehicle and its corresponding parking space) to detect vehicles leaving or occupying parking locations. To our knowledge, there has not been any similar methodology in the literature.

The rest of the paper is organized as follows. We give a brief survey of related work in Section 2. We present our PSN system and experimental results in Sections 3 and 4, respectively. We discuss the limitations of our system and potential future work in Section 5. Finally we conclude in Section 6.

2. Related Work

Parking occupancy detection is a highly active research area. We focus our survey on the latest deep learning-based approaches because they are most relevant to our method.

Amato et al. [11] proposed a decentralized and efficient solution for visual parking occupancy detection based on a deep convolutional neural network specifically designed for smart cameras. Their algorithm performs the detection task directly on-board of a smart camera without a central server.

Jain [12] patented a method for detecting available parking spaces using data from a radar system and a neural network model. Specifically, the neural network model processes the radar data and approximates the parking space boundaries as splines. The spline estimates are updated as the vehicle navigates the parking environment, and a computer system utilizes the spline estimates to detect available parking spaces.

Sarkar et al. [13] proposed a UAV-assisted quick and efficient monitoring solution for real-time parking

occupancy and license plate detection. In their approach, a drone-mounted camera is used to capture images of the parking lot under consideration.

Additionally, Regester et al. [14] proposed a pattern recognition algorithm to map the location of parking spaces from aerial images of parking lots. Paidi et al. [15] aimed to identify parking occupancy in an open-air parking lot, which consists of free parking spaces using a thermal camera.

In our study, we aim to apply a deep neural network model to video signals acquired from various types of surveillance cameras, and identify vacancies by detecting parked vehicles leaving their locations.

3. Methods

In this section, we first give a brief introduction to the Mask R-CNN model and the Intersection over Union (IoU) ratio. They form the foundation of our PSN system. We then present the three core components of PSN, i.e., detecting vehicles in a given image, identifying cars leaving parked locations or reoccupying a freed spaces, and broadcasting new information to the users.

3.1. Mask Region-based CNN (Mask R-CNN)

Mask R-CNN [9] is the latest object localization model in the R-CNN family [16]. At its inception, Mask R-CNN outperformed all existing single-model entries for instance-level recognition in all three tracks of the COCO challenges [10]. Since then, Mask R-CNN has been successfully applied to various computer vision applications [17–19].

Mask R-CNN provides a general framework to efficiently detect objects in an image while simultaneously generating a high-quality segmentation mask and a bounding box for each instance. Its architecture is designed to detect objects across the entire image in a computationally efficient manner without using the traditional sliding window approach [20].

Conceptually, a Mask R-CNN model employs two deep neural networks. The first one is a region proposal network (RPN) that produces a collection of boxes/proposals for potential locations of objects in an image. The second one is a detection neural network that predicts the classes of the objects, refines the bounding boxes, and generates pixel-level masks based on proposals from the RPN. In the implementation of the Mask R-CNN model, the two networks are merged into a single network by sharing their convolutional features and the RPN component serves as the “attention”



Figure 1: Sample Output of a Mask R-CNN Model. A Mask R-CNN model trained to recognize various types of objects in an image, including cars, traffic lights, parking meters, etc. The number next to each object label is the confidence of the class prediction. *Image reproduced from [9].*

mechanism in the unified framework. Figure 1 presents sample outputs from a Mask R-CNN model trained to recognize various types of objects in an image, including cars, traffic lights, parking meters, etc. The number next to each object label is the confidence of the class prediction.

3.2. Intersection over Union (IoU)

As illustrated in Figure 2(a), IoU is a ratio between the intersection and the union of two areas. A higher IoU implies a larger overlapping region, and $\text{IoU}=1.0$ represents a complete alignment of the two areas.

IoU is typically used to evaluate the quality of predicted bounding box locations in an object detection or image segmentation task. Figure 2(b) presents an example of the ground-truth (green) vs. predicted (red) bounding boxes by an object detector. We observe that a higher IoU ratio indicates a more accurate prediction.

In our PSN model, we leverage the IoU concept to measure the overlapping area between a vehicle and its parked space to detect leaving cars. We outline the details of our algorithm in Section 3.4.

3.3. Detecting Vehicles (V-MRCNN Model)

An essential task for our PSN system is detecting vehicles in each video frame. To this end, we favor models in the R-CNN family [16] due to their solid performance and computational efficiency. Because we built our model by fine-tuning pre-trained weights, training time was not a concern in model selection. Mask R-CNN was adopted over the other models because it provides extra object segmentation which could be beneficial in facilitating future enhancements.

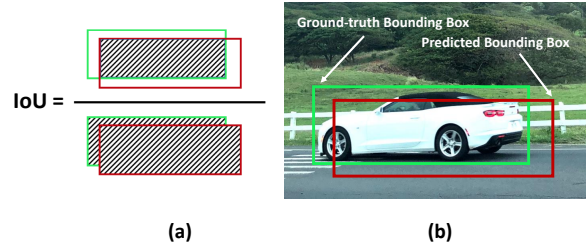


Figure 2: Intersection Over Union (IoU) of Two Bounding Boxes

Despite the overhead in generating the object masks, our calibrated Mask R-CNN model performs at a satisfactory speed of several frames per second in high-resolution videos.

Using the COCO dataset [10], we trained a customized Mask R-CNN model (V-MRCNN onwards) to recognize three types of vehicles: cars, buses, and trucks. Each training sample in the dataset was annotated with object labels, instance locations (bounding boxes), and pixel-wise object masks. We anticipated that V-MRCNN would comfortably handle this localization task because of Mask R-CNN's success in recognizing 91 categories of objects in the COCO dataset. Furthermore, the objects in our study were relatively large and similarly shaped. Our experimental results confirmed this expectation.

3.4. Detecting Available Parking Spaces

To detect vacant parking spaces, we first need to recognize legal parking locations. Due to inter-object occlusion in a crowded parking area, this is not a straightforward task.

Figure 3 illustrates some existing techniques used to identify legal parking spaces. The first one (Figure 3(a)) is to manually outline each parking spot boundary [21]. However, this approach involves tedious manual work and cannot be generalized to new parking areas with different layouts. Figure 3(b) utilizes special indicators, such as meters, to distinguish parking spots from other areas [1]. This method is not reliable because every parking meter corresponds to a legal parking spot, but not vice versa. Figure 3(c) presents the idea of training a model to recognize parking boundaries painted on the road [22]. This method requires the boundaries to be highly visible and distinguishable from lane dividers and crosswalks, which is not the case in many places.

In our study, we make the PSN system start its observations from a moment when nearly all parking spaces are occupied. This is an easily satisfiable requirement in places where PSN would be most useful,

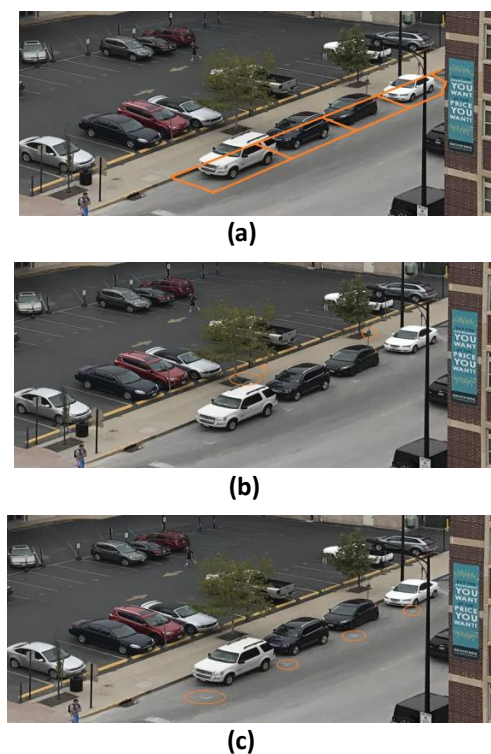


Figure 3: Existing Methods to Identify Legal Parking Areas. (a) manually outline each parking spot boundary. (b) utilizes special indicators such as meters. (c) training a model to recognize parking boundaries painted on the road.

such as large metropolitan areas where complete parking saturation is practically guaranteed during business hours. We assume that cars that remain stationary for a certain period of time at a fixed location have found legal parking spaces. This assumption allows us to indirectly recognize legal parking locations by detecting stationary cars, or more precisely, by calculating the bounding boxes of vehicles identified by our V-MRCNN model at the beginning of a video stream. The PSN system will subsequently track parking spaces freed by leaving vehicles and reoccupation of these spaces.

To quickly and accurately detect vehicles leaving their parked locations, we utilize the IoU (Section 3.2) ratio between the two bounding boxes encompassing the vehicle and its corresponding parking space, respectively. In particular, we observe that a parked car has nearly 100% overlap with a bounded parking space. When a car is leaving, this overlapping region gradually decreases to 0. On the other hand, when a new car reoccupies a freed space, the IoU will increase back to nearly 100%. Thus, we can monitor the change of

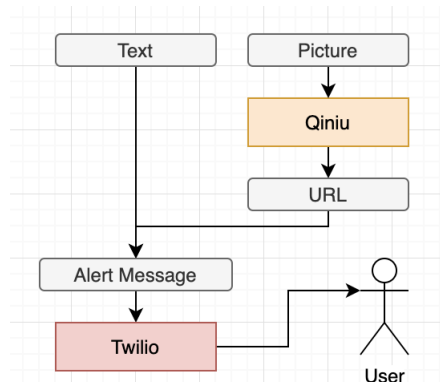


Figure 4: PSN Message Broadcasting Architecture

IoUs of cars and their corresponding parking boundaries in temporal video frames to detect leaving vehicles and reoccupying of the freed space. In calculating the IoU ratio, we obtain the bounding boxes of the vehicles from our V-MRCNN model. For the boundaries of the parking spaces, we use the bounding boxes of the vehicles when the park spaces were first registered.

3.5. Broadcasting Notification Messages

When a parking space becomes unoccupied, our PSN system will send an SMS message to the users. We investigated a few cloud communication platforms and chose the service provided by Twilio [23]. Twilio allows users to send and receive text messages programmatically using its web service API. However, Twilio does not directly accept images in its messaging API. To include pictures of the new parking locations in our SMS messages, we resorted to Qiniu [24], which provided us with a server to store images and their corresponding URLs. Thus, the Twilio API could retrieve an image using its URL and send it along with our alert text. Figure 4 illustrates our message broadcasting architecture.

4. Experimental Results

4.1. Training the V-MRCNN Model

We trained our V-MRCNN model to recognize three types of vehicles: cars, buses, and trucks. Our training data consists of 35,185 images from the COCO dataset [10], among which over 12,000 images contain vehicle objects. We initialized our model using the pre-trained weights of the Mask R-CNN model trained on the COCO dataset. We experimentally adjusted our learning rate to 0.002 with a momentum factor of 0.9. The validation loss of class labels,

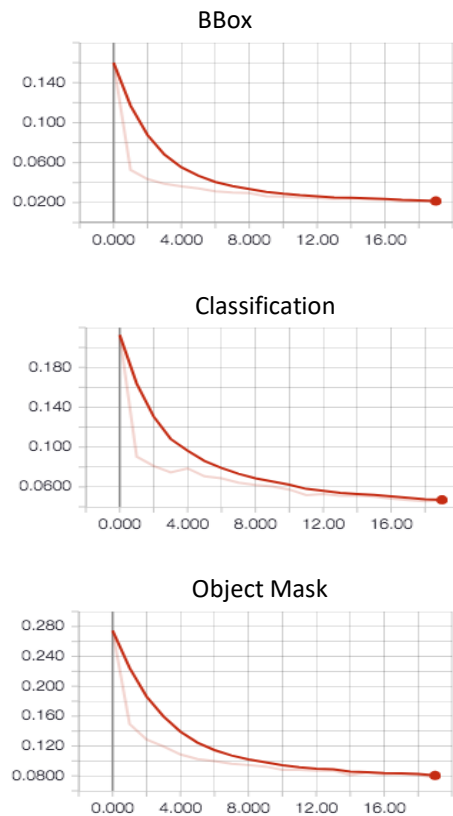


Figure 5: Learning Curves of BBox, Classification, and Object Mask in V-MRCNN Training. The x-axis is the number of epochs. The y-axis is the value of the loss function. Each diagram contains a validation (bold) curve and a training (thin) curve.

bounding boxes, and object masks for our V-MRCNN model are approximately 0.015, 0.02, 0.08, respectively. Compared to the original Mask R-CNN model trained on the entire COCO dataset, our V-MRCNN model improved the detection precision from 0.347 to 0.48 for the car objects in our study. Figure 5 presents the learning curves of BBox, classification, and object mask in training the V-MRCNN model.

4.2. Assessing the PSN System

To evaluate the PSN system, we collected 20 short videos of urban parking activity comprising over one hour of footage. Approximately half of the total footage came from the public domain (YouTube) while the rest came from a CCTV camera we installed in a private parking lot. The videos we obtained are about 40 frames/second. Our PSN system applied the V-MRCNN model to one image every 40 frames (i.e., every second). We experimentally set the IoU threshold

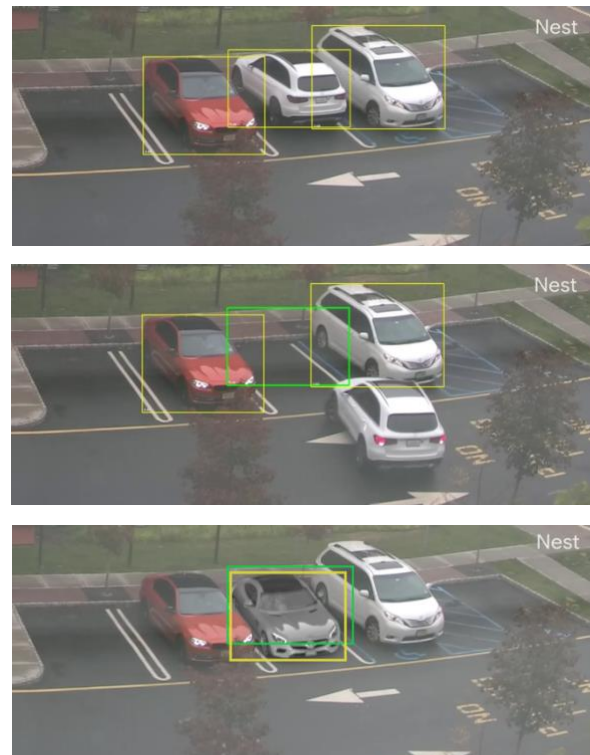


Figure 6: PSN Parking Vacancy Monitoring. Top frame: recognizing and registering three legal parking spaces by detecting bounding boxes of stationary cars (yellow rectangles). Middle frame: detecting a new vacancy (green box) by observing a drop below 0.3 in the IoU of the leaving car and its parking spot. Bottom frame: detecting re-occupancy of the freed space after a new car parked in. The IoU of the new car (yellow box) and the parking space (green box) increased to above 0.7.

to 0.3, which means a car is considered leaving if the IoU ratio between its bounding box and the parking space decreases by at least 70% of the initial IoU ratio.

It is worth noting that, even with a state-of-the-art object detection algorithm, it is still possible for a model to fail to detect the vehicles accurately. To minimize drivers' frustration and disappointment caused by false alerts, we only send a notification if at least two successive video frames confirm an identical vacancy. With this protocol, our PSN system accurately captured all (10) new parking vacancies arising from leaving vehicles in our test videos, and there was no false detections throughout the validation process.

Figure 6 illustrates an example of the PSN system monitoring the parking availabilities. In the top frame, the system recognized three legal parking spaces marked by the yellow boxes. Note that these boxes were the

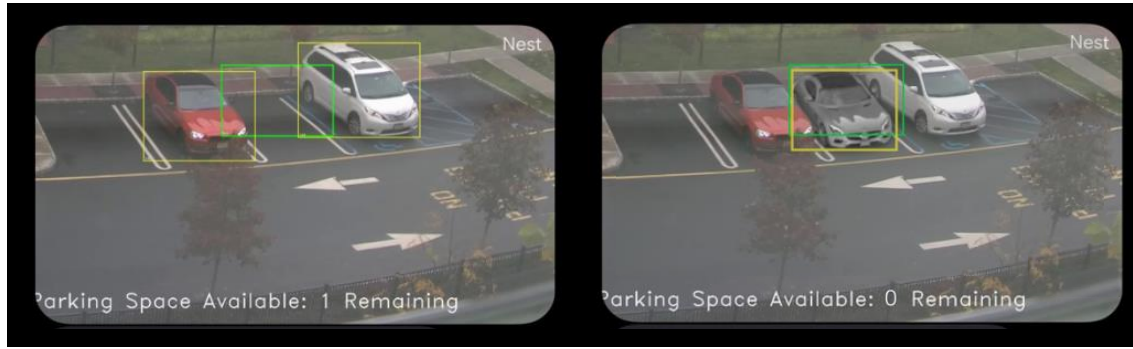


Figure 7: Sample SMS Notification Message. Left: message when a new parking space was detected. Right: message when the available parking space was taken.

bounding boxes of the stationary cars as well. As expected, PSN did not detect the pre-existing empty spaces in the top frame because they had not been previously registered. In public metropolitan areas, such spots will be taken quickly, engendering the corresponding parking locations being registered in the PSN system.

In the middle frame, our system correctly identified the new vacancy (green box) based on the drop in the IoU of the middle car and its parking spot. The bottom frame demonstrates that when the freed space was reoccupied by a new car, the IoU ratio between the car (yellow box) and the parking boundary (green box) increased from 0 to above 70%. Consequently, PSN notified the users that the parking space was claimed.

We used the messaging service described in Section 3.5 to send alerts to subscribers when a new parking space became available or reoccupied. Figure 7(a) presents the SMS message delivered after the middle car in Figure 6 left. Figure 7(b) shows the message notifying the users that the parking space was taken (i.e., bottom frame of Figure 6).

5. Discussions and Future Work

PSN monitors a stationary car for a time interval t before declaring the spot it occupies a legal parking space. In practice, t is an adjustable, neighborhood-dependent parameter only used for the initial parking space establishment. As illustrated in Section 4.2, once a parking spot is identified, PSN will use the registered boundary to monitor the location's subsequent availability. Thus, once the system has been running for an extended period of time, all legal spaces will likely be detected and registered. We believe that the value of t can be successfully adjusted in areas where the system would be most applicable and optimizing value t is not the focus of this paper. Indeed, we focused

on testing PSN's vacancy tracking quality by assuming all cars have passed the threshold time t at the start of a video stream.

Although the PSN system demonstrated promising results when tested on our private CCTV stream and several publicly available videos, it has not been tested on a large scale. Scaling it up requires securing of legal contracts with public parking operators regarding resource sharing, confidentiality, etc., which is a time consuming process that is independent of the involved research. We foresee a few enhancements that could improve the efficacy and robustness of our prototype system.

An assumption of PSN is that every stationary vehicle represents a legal parking spot. While this is mostly the case for parking lots, the situation can be different for street parking. For example, our system can be misled by a car standing by a hydrant with emergency lights on and, consequently, sends a false alert when the car leaves the location. To improve the accuracy in identifying legal parking spaces, one potential solution is to enhance our V-MRCNN model to recognize typical illegal (e.g. hydrant) and legal (parking meters) landmarks and incorporate the information in the detection algorithm.

Another area for potential improvement is PSN's message delivery system. We presented a proof of concept solution using a cloud messaging service. In practice, a more serviceable implementation would be a city- or contractor- operated CCTV network, for which PSN would provide a real-time inventory of vacant parking spots. Additional services such as reservations and variable pricing could also be offered. Agencies/contractors controlling parking space could further develop a subscription service where subscribers pay less than cash/credit card users. A service akin to a tollway system where the drivers with electronic passes

pay lower fees than cash payers.

Lastly, our system can be naturally enhanced by a statistics component to provide real-time expected vacancy/occupancy times for every parking spot in a given vicinity based on past statistics. In addition, by integrating with a GPS system, the broadcasting process can be further optimized to target nearby drivers to increase efficiency in large urban areas.

6. Conclusions

In this paper, we presented a simple, effective, and promising PSN system capable of monitoring individual parking locations and, subsequently, updating subscribing users of their availabilities. Leveraging a unique adaptation of the IoU concept, our PSN system employs a state-of-the-art object localization model to track the changes of car positions in a surveillance video stream. Prompt user notifications are automatically delivered to the users using cloud messaging services. Our experimental results demonstrate the high accuracy and reliability of the system. We further addressed PSN's limitations and future work to improve the efficacy of the system.

References

- [1] N. Ratnakar, "Smart parking meter," July 13 2006. US Patent App. 10/905,467.
- [2] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.
- [3] D. Cireşan and U. Meier, "Multi-column deep neural networks for offline handwritten chinese character classification," in *2015 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–6, IEEE, 2015.
- [4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.
- [5] J. Russell, "Google's alphago ai wins three-match series against the world's best go player," *TechCrunch*, 2017.
- [6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- [7] A. Karpathy and L. Fei-Fei, "Deep visual-semantic alignments for generating image descriptions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3128–3137, 2015.
- [8] X. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in *Advances in neural information processing systems*, pp. 2802–2810, 2016.
- [9] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.
- [10] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*, pp. 740–755, Springer, 2014.
- [11] G. Amato, F. Carrara, F. Falchi, C. Gennaro, and C. Vairo, "Car parking occupancy detection using smart camera networks and deep learning," in *2016 IEEE Symposium on Computers and Communication (ISCC)*, pp. 1212–1217, IEEE, 2016.
- [12] J. J. Jain, S. Kadetotad, H. Banvait, P. G. Joh, *et al.*, "Detecting available parking spaces," May 28 2019. US Patent 10,304,335.
- [13] S. Sarkar, M. W. Totaro, and K. Elgazzar, "Intelligent drone-based surveillance: application to parking lot monitoring and detection," in *Unmanned Systems Technology XXI*, vol. 11021, p. 1102104, International Society for Optics and Photonics, 2019.
- [14] A. Regester and V. Paruchuri, "Using computer vision techniques for parking space detection in aerial imagery," in *Science and Information Conference*, pp. 190–204, Springer, 2019.
- [15] V. Paidi and H. Fleyeh, "Parking occupancy detection using thermal camera," in *Vehicle Technology and Intelligent Transport Systems*, 2019.
- [16] L. Weng, "Object detection for dummies part 3: R-cnn family," *lilianweng.github.io/lil-log*, 2017.
- [17] R. Zhang, C. Cheng, X. Zhao, and X. Li, "Multiscale mask r-cnn-based lung tumor detection using pet imaging," *Molecular imaging*, vol. 18, p. 1536012119863531, 2019.
- [18] W. Zhang, C. Witharana, A. Liljedahl, and M. Kanevskiy, "Deep convolutional neural networks for automated characterization of arctic ice-wedge polygons in very high spatial resolution aerial imagery," *Remote Sensing*, vol. 10, no. 9, p. 1487, 2018.
- [19] S. Elnagar and M. A. Thomas, "Real estate image-based appraisal using mask region based convolutional networks," 2019.
- [20] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," *arXiv preprint arXiv:1511.08458*, 2015.
- [21] S. Valipour, M. Siam, E. Stroulia, and M. Jagersand, "Parking-stall vacancy indicator system, based on deep convolutional neural networks," in *2016 IEEE 3rd World Forum on Internet of Things (WF-IoT)*, pp. 655–660, IEEE, 2016.
- [22] A. Goyal and A. Rohatgi, "Detecting whether a parking zone is vacant or occupied," *nayan.co/blog/AI/Vehicle-Parking-Occupancy-Detection*, 2020.
- [23] M. Beccaria, "How to provide live library information via sms using twilio," *Code4Lib Journal*, no. 14, 2011.
- [24] "Qiniu: A leading cloud service provider in china," <https://github.com/qiniu>.