Parking Occupancy Detection from CCTV Images

Jule Valendo Halim -1425567

GEOM90038 - Advanced Imaging

Parking Occupancy Detection from CCTV Images

Contents

Introduction

Recent developments in data has undergone significant changes from data processing into machine learning due to increased volumes of readily available data (Khan and Al-Habsi (2020)). The introduction of image-based neural network architecture has allowed the field of computer vision to flourish. Computer vision leverages the power of machine learning to derive meaningful information from visual data, which allows them to take actions and recommendations when they detect issues. Powerful advances in machine learning such as the widely popular transformer model that Chat-GPT is based on has also been adapted to be trained on visual information (El-Nouby et al. (2021)). Figure 1 provides a representation of a neural network, which consists of multiple layers. The input layer is the initial data, and hidden layers calculate the weights of each of these data and propagates them forward to other hidden layers. Finally, the output layer provides probabilities of the desired output (e.g., some classification label).

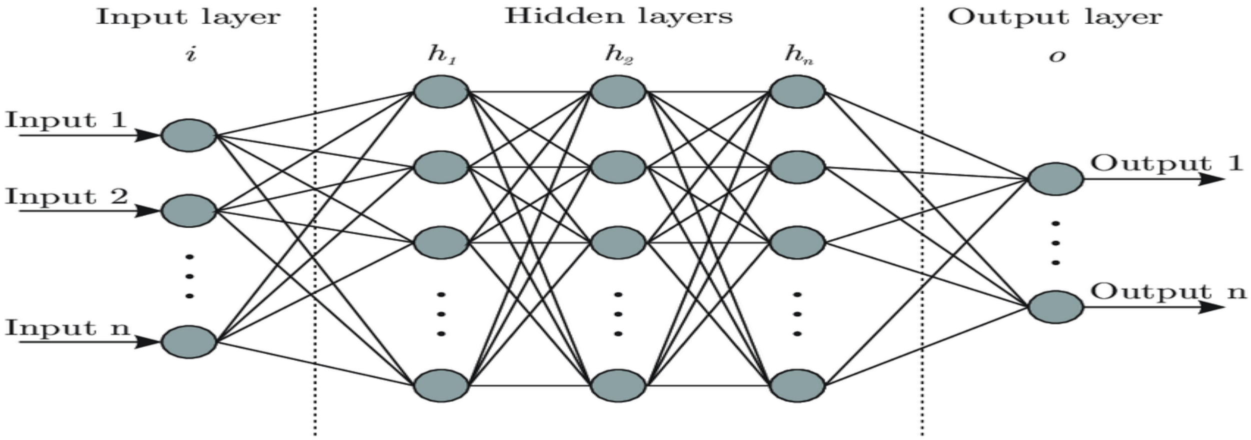

Figure 1

Visual Representation of Layers That Makes Up a Neural Network (Shukla (2019))

Figure 2 provides a simplified neural network architecture for visual data. The visual data that is placed into the input layer depends on the specific architecture and design. In the case of figure 2, a section of the image is placed into the input layer. It is then propagated forward towards additional layers before finally obtaining the results of the output layer, which in this case, attempts to classify the image into four possible species.

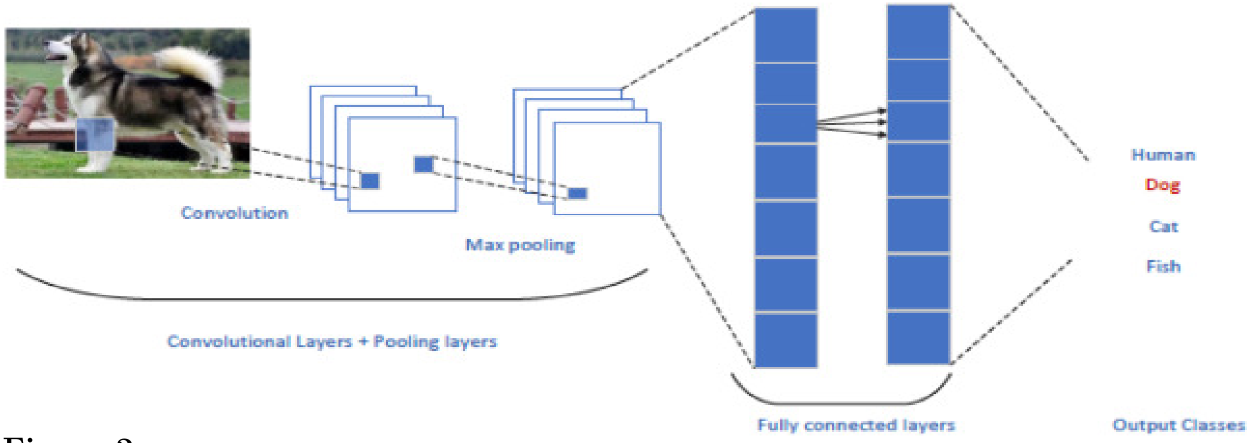

Figure 2

Sample Architecture of Machine Learning Being Used in Computer Vision (Chai et al. (2021))

The use of computer vision has been applied to wide range of different fields. For example, Esteva et al. (2021) discussed the way medical imaging applications in multiple medical areas such as pathology and dermatology has enhanced the level of care provided. Another field in which machine learning has enhanced computer vision is in the automation and digitization of fruit quality measuring and maintainence (Rathnayake P et al. (2022)). Figure 3 shows how computer vision can be used to detect cars and humans in images. These detections are generated post-training, where multiple images are provided for training the neural network. The trained neural network can then be given new images or even real time video to detect what they were trained to. However, as seen in figure 3, such neural networks are not perfect, and can misclassify or not detect parts of the image that they are supposed to (e.g., the model did not successfully classify one of the humans walking).

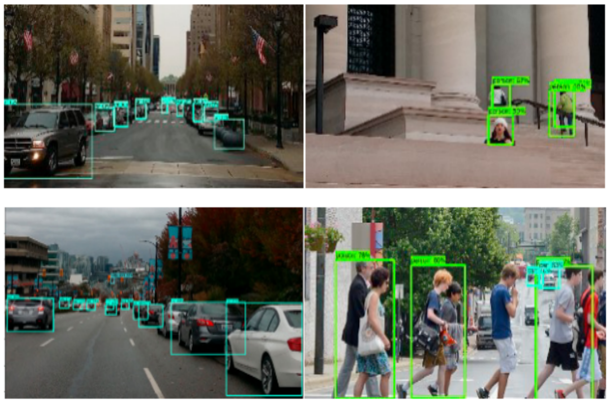

Figure 3

Sample Image Detection Results of Cars and People Using Deep Learning and RCNN Models(Khan and Al-Habsi (2020))

Faster Recurrent Neural Network models (FasterRCNN model) are a form of neural network that can detect objects in images. The FasterRCNN model is based off Ren et al. (2016). It consists of two main components, the Region Proposal Network (RPN) and the FastRCNN. RPNs are a convolutional neural network, which are a form of neural network such as those described above, but are specialized in taking three dimensional data (such as images) as inputs. RPNs are further specialized to handle different schemas of different images. Figure 4 shows the different schemas that the RPN can handle, namely differing scales, differing filter sizes, and multiple references of the same image.

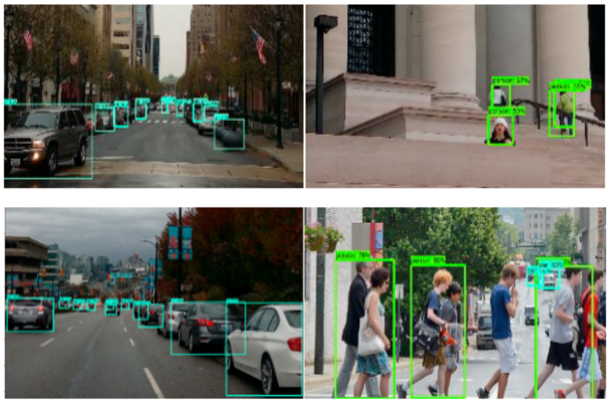

Figure 4

Different Schemas that RPNs Are Adapted To(Ren et al. (2016))

The other component, the FastRCNN (Girshick (2015)) is a neural network that has been shown to be much faster than other image detection neural networks. The main innovation that causes this increased detection speed is the use of of Regions of Interest (RoI). The RoI is defined with a four-tuple (r,c,h,w). The r and c specifies the top-left corner, while the h and w defines its height and width respectively. These tuples are then max-pooled to convert features inside regions of interest into a small feature map with fixed spatial size (e.g., height and width of 6 x 7). Figure 5 shows the overall architecture of the fastRCNN. As seen in this figure, the image is projected into an RoI format before being pooled and trained on a neural network.

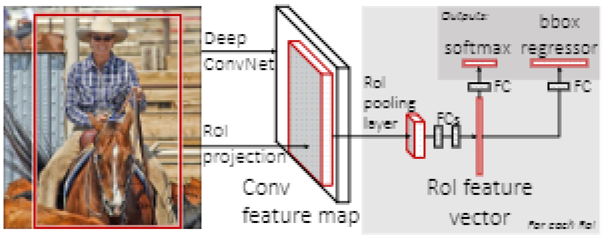

Figure 5

The FastRCNN Architecture, Including the Conversion of Images into ROI(Girshick (2015))

The FasterRCNN architecture has layers that provide output towards the RPN. This is then used to create a proposal, which basically proposes what sort of schema the image is using (see figure 4). The image that is passed through the RPN is also used to create feature maps, which create is then passed to the the RoI pooling before being passed to a classifier. In this architecture, the RPN is performing the 'attention' task of FasterRCNN. Attention allows the model to incorporate sequential information into the model (e.g., in a picture of a man on a horse, the model is able to remember that the section of the image with a hat is in the upper-middle area of the whole image). Figure 6 shows the entirety of the FasterRCNN architecture.

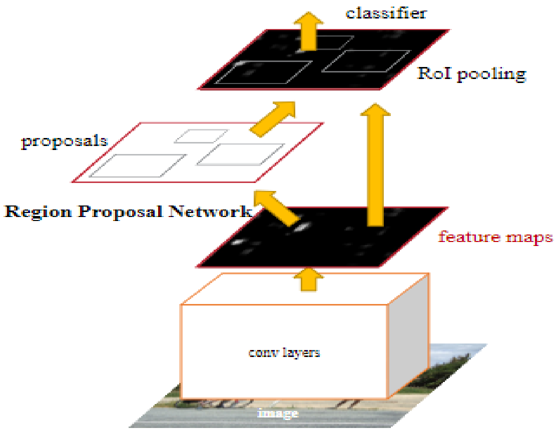

Figure 6

The FasterRCNN Architecture, Showing the Components of RPN and FastRCNN RoI (Ren et al. (2016))

The other model used in this report is the Resnet50 CNN model (He et al. (2015)). The Resnet50 CNN model's building block consists of a series of weight layers that have an relu activation function (an activation function converts the weights provided by the layer mapped to an output). Relu takes the maximum of two values and returns an output depending on the input (e.g., a relu that is max(0,x) and has a binary classification task to classify 0 and 1 would classify an output as 1 if x is greater than 0, and classify it as 0 if it is less than 0). Figure 7 shows the building block layer of the Resnet50. The Resnet50 CNN consists of multiple of these layers along with other kinds of layers, one of which is the max-pool layer. This increased depth of representation is vital for visual recognition tasks, and this model is popular in many visual recognition tasks.

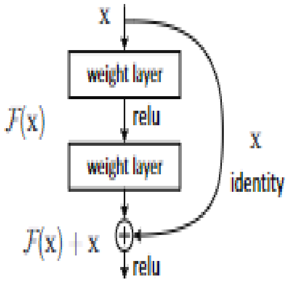

Figure 7

The Resnet50 CNN Building Block (Layer)(He et al. (2015))

In this report, I aim to use and investigate a pre-trained FasterRCNN model and

Resnet50 CNN model(trained on the PKPlot dataset) to detect cars and delineate parking spaces in a picture of the Barry Street parking lot. The software used will be MATLAB 2020a, along with additional toolboxes such as the deep learning toolbox. The methodology of this begins with a visualization of the dataset, before using a pre-trained Resnet50 CNN classifier to identify whether a parking slot is occupied or empty. Afterwards, parking slot delineation is done using FasterRCNN, which detects cars. The dataset for this step involves images of the parking slot and bounding boxes, which identifies the space in which an object is identified (see figure 3, where cars and human areas and identified using bounding boxes).

These results will then be evaluated by calculating recall, precision and accuracy. Afterwards, the object performance will be post-processed to improve the resulting classification. Visual inspection of the resulting bounding boxes will also be compared to the ground truth of where bounding boxes should be. Additional information is found in the methods and results section.

Methods and Results

Visualizing Dataset

This step involves loading in the PKPlot and Barry Street annotated images. These annotations are bounding boxes that are manually set. Figure 8 shows the annotated images of the PKPlot (left) and Barry Street (right) datasets. The PKPlot dataset consists of over 695,000+ parking space images. For this report, only a segment of the total dataset was used.

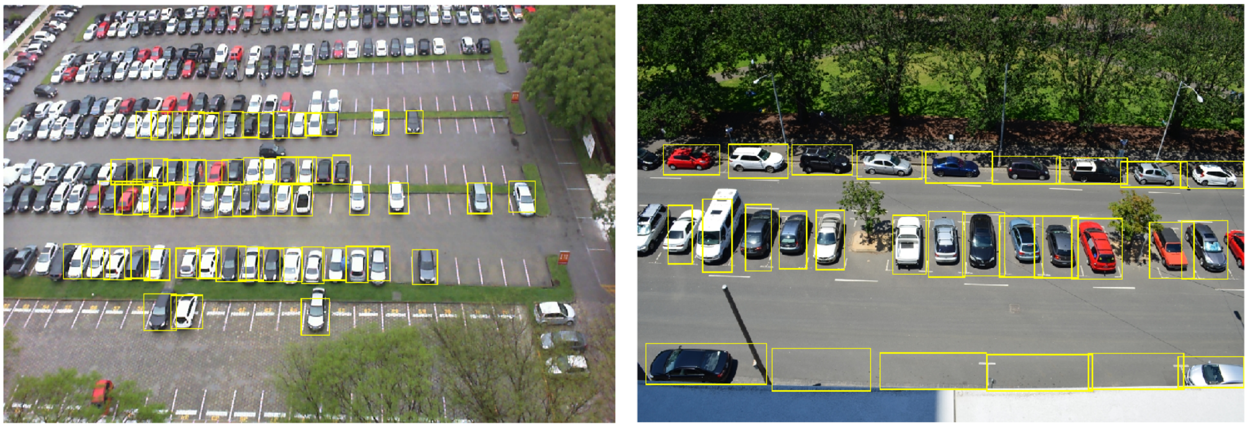

Figure 8

Images of Annotated from the PKPlot Dataset (left) and Barry Street Dataset (right)

Figure 9 shows a sample empty slot and occupied slot found in the annotated PKPlot datasets.

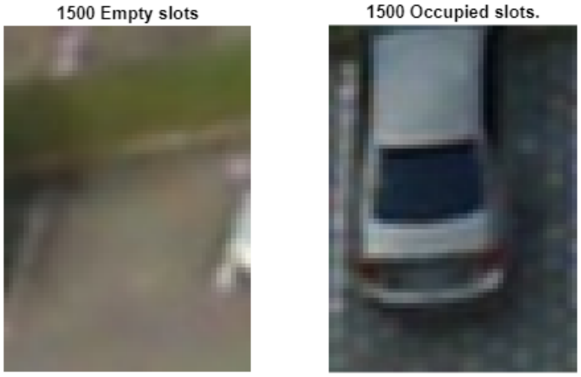

Figure 9

Sample Image of Empty (left) and Occupied (right) Slots of PKPlot, Sampled From 1500 Empty Spaces and 1500 Occupied Spaces

Creating Car Detector

To create the car detector, the pre-trained Resnet50 CNN

Automatic Delineation of Parking Spaces

Evaluation

Post-Processing Improvements

Discussion

Accuracy, Precision, and Recall Evaluation

Improvement Based on Assumptions

Challenges and Shortcomings

Scopes of Improvement

Conclusions and Future Directions

Appendix

References

Chai, J., Zeng, H., Li, A., & Ngai, E. W. (2021). Deep learning in computer vision: A critical review of emerging techniques and application scenarios. *Machine Learning with Applications*, *6*, 100134. https://doi.org/https://doi.org/10.1016/j.mlwa.2021.100134

El-Nouby, A., Neverova, N., Laptev, I., & Jégou, H. (2021). Training vision transformers for image retrieval.

Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., Liu, Y., Topol, E., Dean, J., & Socher, R. (2021). Deep learning-enabled medical computer vision. *npj Digital Medicine*, *4*, 1–10. https://doi.org/https://doi.org/10.1038/s41746-020-00376-2

Girshick, R. (2015). Fast r-cnn.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition.

Khan, A. I., & Al-Habsi, S. (2020). Machine learning in computer vision [International Conference on Computational Intelligence and Data Science]. *Procedia Computer Science*, *167*, 1444–1451. https://doi.org/https://doi.org/10.1016/j.procs.2020.03.355

Rathnayake P, W., D, G. D., Punchihewa G, A., Anjana G, N., Suriya Kumari, P. K., & Samarakoon, U. (2022). Certimart: Use computer vision to digitize and automate supermarket with fruit quality measuring and maintaining. *2022 4th International Conference on Advancements in Computing (ICAC)*, 36–41. https://doi.org/10.1109/ICAC57685.2022.10025119

Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster r-cnn: Towards real-time object detection with region proposal networks.

Shukla, L. (2019). Designing your neural networks. *Towards Data Science.* https://towardsdatascience.com/designing-your-neural-networks-a5e4617027ed