Number Plate Detection in Different Environments using Generative Adversarial Networks

Alden Boby
Supervisor: Dr. Dane Brown
Computer Science Department, Rhodes University

May 21, 2021

Abstract

Licence plates are a form of vehicle identification with a unique set of characters that can be used to identify details about a car and its owner. This information can be used by law enforcement to promote adherence to traffic rules. Initially, licence plate recognition was accomplished using clear images taken in favourable and controlled conditions. In real-world licence plate recognition scenarios, images are not as clear as the datasets used as input. Real-world data typically come from mediums such as CCTV, which compromises video quality to reduce storage costs as storing 24-hour high-resolution videos is not economical. Newer machine learning approaches such as applying generative adversarial networks (GANs) and convolutional neural networks (CNNs) to upscale and reconstruct images address low-quality input. Training these systems with a suitable dataset will enable them to better detect number plates in unfavourable conditions and environments.

1 Introduction

1.1 Context of Research

Licence plate recognition (LPR) is not a new development and has been explored in the past (Chang et al., 2004). There are working systems that can detect number plates on a vehicle, isolate them and successfully deduce the numbers and characters on the plate with either character segmentation or optical character recognition (OCR) (Anagnostopoulos, 2014). Given input that is flawless, i.e. high-resolution, with favourable lighting and no distortion or occlusion, existing systems will work effortlessly (Du et al., 2012). A problem arises when the input is unclear, this is often the case with real-world data such as footage captured from CCTV cameras which are compressed, leading to a loss of image quality, decreasing the available data points (pixels) which can be used for character recognition, edge detection and so on (Lee et al., 2018). This specific problem can be addressed by upscaling low-resolution images using machine learning techniques to produce a high-resolution image with more information in it (Lee et al., 2019). The high-resolution image containing more information to extrapolate can then be used to recognise the characters on the number plate through more traditional means. LPR is applicable in several areas such as security, traffic regulation and parking (Anagnostopoulos, 2014; Du et al., 2012). Licence plates can be used to identify stolen vehicles and even automatically bill car owners for parking and using toll gates, reducing the amount of manual work that needs to be done. There is a need to create a system that can sustain good accuracy when fed varying quality data. The use of generative adversarial networks and convolutional neural networks will be investigated to achieve this goal.

2 Image Processing

Image processing refers to a number of techniques that can be applied to an input image to extract useful information from it (Anagnostopoulos, 2014). Chang et al. (2004) offers some insight into how LPR was dealt with historically. The basic stages of LPR involve locating a number plate within an image (localisation), isolating it, and then afterwards determining the actual data on the number plate. LPR requires an image as input, which is transformed with the relevant image processing techniques. The resultant image is then used to produce a string of characters as the final output (Anagnostopoulos, 2014). The performance of such systems is measured through two variables; speed and accuracy. This particular project will favour accuracy as speed is of no use if the system cannot produce accurate output.

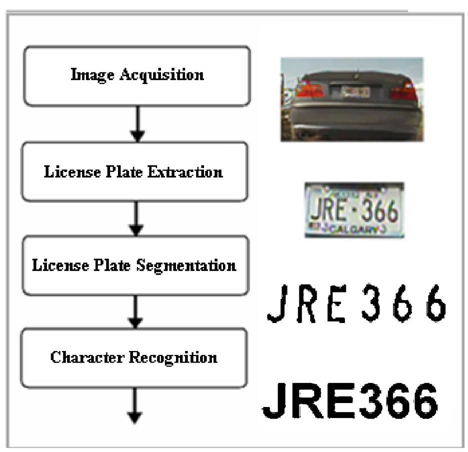

Figure 1: Basic stages of LPR (Du et al., 2012)

Chang et al. (2004) did not utilise much of the machine learning approach as it was in its infancy during the paper's time of publication, coupled with the relatively high cost of capable machines.

2.1 Edge Detection

Edge detection is one of the many techniques applied to digital images to extract useful features (Anagnostopoulos, 2014). Its purpose is to reduce the amount of data to be processed while maintaining structural information such as the edges and vertices in the image that give it meaning (Canny, 1986). Chang et al. (2004) utilised an edge detection technique limited to a small subset of colours related to the plates they were concerned with, followed by fuzzy maps and OCR using a neural network. The system by Chang et al. (2004) was very successful for head-on images of vehicles. However, when using a dataset with more complex backgrounds and angles, detection accuracy decreased. However, the accuracy remained above 90%, which is still relatively good. It is important to note that although successful, the set of data used for this experiment was clear and well illuminated. The overall success rate of their LPR system was 93.7%.

As mentioned previously in the introduction, one of the main factors affecting the performance of such systems is the quality of the data. Images can suffer from occlusion (Anagnostopoulos, 2014) and other issues, resulting in incomplete data, making the detection process more difficult (Du et al., 2012). One of the aims of this research project is to explore how images can be restored before they are used as input to detect characters on a number plate.

2.2 Connected Component Analysis

The connected component analysis (CCA) method groups pixels of similar values together and gathers information about regions in the image; given that the images being dealt with are binary, the CCA only considers two values (Anagnostopoulos, 2014; Babbar et al., 2018). CCA can help locate number plates within images after pre-processing, such as thresholding (Babbar et al., 2018). Features such as aspect ratio and area can be used to single out locations of number plates within an image (Du et al., 2012).

Babbar et al. (2018) uses CCA in a relatively recent paper. Regular image pre-processing is done to make the data more digestible for the models that utilise it afterwards. Babbar et al. (2018) use a unique approach at the character recognition stage. The authors consider the use of multiple classifiers in order to improve number plate recognition in poor light conditions. This benefit is not seen in the paper published by Chang et al. (2004), who similarly used connected component labelling but did not supplement it with machine learning algorithms, which is understandable given the period between publication dates.

Babbar et al. (2018) used CCA in their approach because connected regions are a feature of number plates. Using python, the authors were able to visualise the regions found by CCA. Further processing was done to ensure that the CCA regions accurately represented the location of number plates in images and did not just match segments with characteristics that met the criteria of a number plate.

CCA can also be applied at the character recognition stage as characters can also be defined as a connected region (Anagnostopoulos, 2014). The size of characters on number plates are defined and uniform; ratio analysis can also be used to supplement CCA at this stage.

3 Number Plate Localisation

Number plate localisation is the process of locating the position of a number plate within a given image. Regarding LPR, this step is used to segment the image to isolate the region of interest (ROI). Following localisation, image processing techniques can be applied to the resulting output to extract the relevant features to identify the contents on the number plate.

There are a number of ways in which the coordinates of a licence plate can be deduced from an image. The first and most straightforward method is manually highlighting the ROI within a given dataset; this requires no computational tasks but is impractical on large datasets. Shaikh et al. (2016) used this method on their own dataset, which was small enough to make tagging the images and marking number plates with red bounding boxes feasible.

Other alternative methods can be either through CCA as done by Babbar et al. (2018) or edge detection, as number plates have specific features which can be realised using the aforementioned image processing techniques. With advances in technology, machine learning algorithms can actually bypass certain image processing stages and perform feature extraction independent of image processing. Existing frameworks such as the You Only Look Once (YOLO) framework, which is almost explicitly used for object detection, is an example of a machine learning approach, which will be discussed in subsequent sections of this paper.

4 Machine Learning

Machine learning is classed under two distinct categories, supervised and unsupervised learning. The former makes use of known outputs and trains an algorithm to learn specific features of a labelled dataset by minimising the error between the ground truth and predictions made by the system (Bonaccorso, 2017). One must be careful with this method as it can lead to overfitting; the model needs to have generalisation capabilities, so that it can predict unseen data correctly. Unsupervised learning removes the element of comparison between input and expected output and then interprets as much as it can when fed data.

A support vector machine (SVM) is a binary classifier that utilises a hyperplane to separate data into two separate categories. The shape of the separator is determined by the kernel applied by the SVM (Ayodele, 2010). Babbar et al. (2018) utilise several supervised machine learning techniques at the character recognition stage; all models were trained with a dataset of 20 images. The images were converted to a one-dimensional array of pixels, of which each pixel was a separate feature aiding in training the model. The classifiers considered in the paper were: SVM, K-Nearest Neighbours (K-NN), Logistic Regression, Extra Tree Classifier, Random Forests and lastly, a combination of all the mentioned classifiers.

The model proposed by Babbar et al. (2018) is said to have improved accuracy and detection in low and over-bright lighting conditions. Out of all the classifiers used, the best accuracy was obtained from using an SVM with a linear kernel and achieved an accuracy value of 97.1%. This model was tested on a new dataset and was not compared to any other existing systems. Khan et al. (2017) also made use of an SVM for number plate classification. They first performed some pre-processing in order to refine the input, this involved erosion and dilation to improve the visibility of characters in their images. Following the image processing stage, their model extracted features from their dataset using histogram oriented gradients (HOG) and geometric features. Finally, the SVM was trained on the extracted features.

Shaikh et al. (2016) propose a solution LPR by using Deformable Part Models (DPM) (Felzenszwalb et al., 2008) to extract number plate features from a training dataset and a Structural Support Vector Machine (SSVM) in order to train their number plate detector to recognise said features. DPM utilise features extracted from histogram oriented gradient models (Shaikh et al., 2016). With DPM models, the use of higher resolution features correlates to higher recognition rates (Felzenszwalb et al., 2008).

The technique proposed by Shaikh et al. (2016) can be used to train more than one classifier and is not limited to their SSVM. Linear SVMs are binary classifiers that produce limited output and do not provide much valuable information other than positive and negative identification. The SSVM, however, inherits the features of the linear SVM and brings in its own set. Although traditionally used in predicting labelled trees, its ability to learn complex outputs and its high prediction accuracy made the authors consider its use for number plate detection.

Shaikh et al. (2016) state that a combination of image processing and machine learning results in a more accurate system. Meaning there is potential to improve upon solely image processing based approaches. Rather than looking for features within an image, the number plate detection system was trained using pre-extracted features.

The process of the system includes extracting relevant features from the ROI within the images in the dataset using the DPM. The ROI was marked manually by highlighting it with a red square; this was possible given the small dataset of 25 - 30 images. The extracted DPM features are then used to train the number plate detector and are input for the SSVM.

The number plate detection system by Shaikh et al. (2016) is fed images and looks for numbers plates within them; if found, the coordinates of the number plate are returned. These are then used to extract the number plate from the input image. The extracted image then goes through further image processing to enhance it. These include Bilateral Gaussian Filtering, Adaptive Thresholding and De-skewing Shaikh et al. (2016). Afterwards, the characters are extracted from a bounding box as surrounding details on the number plate will confuse the OCR process.

The number plate system was compared against other pre-existing ones and was tested using the same dataset (Caltech dataset) as all others to ensure fair testing. The new system by Shaikh et al. (2016) proved to be more accurate and more efficient in terms of speed when evaluated against specific existing systems. Khan et al. (2017) also tested their data against the Caltech dataset, reporting an accuracy of 99.8% showing the potential of combining HOG and SVMs.

4.1 Convolutional Neural Network

CNNs are a form of deep learning with increasing popularity as computing power limits continue to increase. The name is derived from the mathematical operation of convolution using matrices. CNNs have been proven to have excellent results when it comes to computer vision tasks, hence the usage of the said algorithm for LPR, which deals with images as input. A CNN can be looked at as a specialised artificial neural network (ANN) that is able to select features and detect patterns that can be used for classification. The sliding windows within a CNN allow the network to learn features of the input image regardless of where the target is located. CNNs have enabled more complex problems as they have reduced parameters when compared to ANNs (Albawi et al., 2017).

The implementation of computer vision via CNNs and YOLO is explained by Du (2018). These models try to replicate how a human perceives their environment. Existing models are approaching human levels of performance, and machine learning provides promising results in giving computers the visual capabilities of the human being.

CNN began with LeNET in 1998 and was further advanced with AlexNET in 2012. By 2015 the error rate had dropped below that of the human eye 5.1% to 3.6%, meaning that deep learning algorithms were better at recognition than humans (Du, 2018). CNNs are relatively strong when it comes to feature extraction from single objects; however, once other objects enter the scene, there is an added layer of complexity affecting performance.

CNNs are comprised of input, convolutional, active, pooling and fully connected layers. This architecture allows CNNs to extract features from images using mathematical logic; from there, they are able to map data with the same features using weights (Albawi et al., 2017). CNNs, however, are limited by the datasets used to train them as well the hardware used for computation.

YOLO is a specialised object detection method that is able to predict objects within a given input image in one go. YOLO uses GoogleLeNet as a base (Du, 2018). The selling point of YOLO is its relative speed and accuracy. YOLOv2 has significant improvements over its first iteration, and these come through using batch normalisation on the input data and cutting some convolutional layers from the network.

Lee et al. (2018) utilised a YOLO network (YOLO9000) to detect number plates specifically. The network was trained on general imaging data and was not trained using vehicle or number plate specific data. Their model had a total recall of 93.2% when detecting licence plate and vehicles from their own dataset.

Similarly, Laroca et al. (2018) made use of the YOLOv2 object detector at two stages of the LPR pipeline, first to detect vehicles in an image or frame, followed by detection of the number plates within the vehicles located in the first instance.

Related studies mention the use of convolutional neural networks (CNNs) to automatically detect licence plates within an image (Masood et al., 2017), training the CNN to recognise characters and do character segmentation. Laroca et al. (2018) did this using a special CNN called CR-NET and trained two networks separately for the detection of alphabet characters and numerical characters. The use of CNNs in this area is not limited to detection; they have also been used as a means of upscaling low-resolution images. This is done by mapping low-quality images to high-quality images allowing the CNN to produce high-resolution images using this information (Dong et al., 2015).

This version of a CNN was called the Super-Resolution Convolutional Neural Network (SRCNN). After being fed a low-resolution image, the SRCNN will then extract a set of feature maps; these features are then mapped non-linearly to high-resolution patches. The processes are done by the first and second convolutional layers, respectively. The last layer then uses information gathered to produce a high-resolution image.

4.2 Generative Adversarial Network

A GAN is made up of two neural networks: A discriminator and a generator. The role of these two networks is to work against each other in order to train the networks to help with output.

The use of GANs to enhance number plate detection has been explored by Lee et al. (2018). They aimed to use machine learning super-resolution techniques to upscale an image. This was achieved using a GAN instead of traditional CNN methods, which would create high-resolution images that would be blurry and lack clarity. With the addition of GANs, resulting images would be sharper as it would be utilising adversarial loss rather than mean squared error (MSE) loss in CNNs. Supported by Lucas et al. (2019) MSE methods used by existing deep neural networks produce reasonable high-resolution output; however, they do not tap into the full potential of such networks resulting in the said blurry images. Furthermore, MSE does not equate to the human perception of image quality and fidelity; an obscure image can have the same MSE as one that is perceptually clearer to the human eye (Wang and Bovik, 2009). Coupling MSE with feature-based losses can yield improved results when using a discriminatory network pre-trained with feature spaces to map the l_2 distances between a ground truth image and synthesised image (Lucas et al., 2019).

Lee et al. (2019) touch upon how a GAN can be used to enhance an image. Single image super-resolution and de-blurring are suitable methods to run through a GAN for image enhancement (Lee et al., 2019). Lee et al. (2019) talk of using GANs in conjunction with CNNs in order to produce better results than a CNN alone, which was similarly discussed by Lee et al. (2018). More applicable to this field of study, a super-resolution generative adversarial network (SRGAN) can be used to achieve the desired effect. The GAN, which is made up of a discriminator and a generator, will be fed low-resolution images while outputting high-resolution images; the discriminator's job will then be to distinguish whether the image that it has been fed is from the generator or an actual high-resolution image (Lee et al., 2019). The discriminator and generator then learn from each other through adversarial loss. Once the training process is complete, and the generator can create images that the discriminator cannot distinguish, only the generator will be required. The generator will then take in low-resolution images and generate high-resolution ones.

Yuan et al. (2018) argue that an end to end mapping of low-resolution and high-resolution images is impractical in the real world as low-resolution images rarely have a corresponding high-resolution counterpart. They aimed to use generative adversarial networks to achieve super-resolution via unsupervised learning. This is in contrast to Lee et al. (2018) who utilised end to end mapping to train a GAN.

Existing literature use low-resolution images that have been downsampled using the same kernel, meaning that there is an artificial consistency within the dataset contrary to real-world data, which is more erratic (Yuan et al., 2018). GANs tackle the problem of unsupervised learning as they are able to hallucinate data that is not directly present within the input.

Yuan et al. (2018)'s solution included first mapping low-resolution images to de-noised images of the same resolution. Following that step, the clean image is then upsampled to a higher resolution using an already existing GAN. The performance of the proposed system is compared to existing super-resolution capable systems such as CNNs and SRGANs.

The results of the model in this paper were evaluated using peak signal to noise ratio (PSNR) and structural similarity index measures (SSIM) which are more or less used to measure the difference between two images. The cycle in cycle GAN (CinCGAN) achieved an SSIM of 0.69 and a PSNR of 24.33, which are slightly above alternative methods such as SRGAN with an SSIM of 0.67 but the same PSNR of 24.33

5 Concluding Remarks

LPR is a widely studied field in computer vision. The accuracy of image processing techniques has reached its limits. Current image processing techniques alone cannot further improve results in this field; it is best to look toward more advanced and up to date techniques and algorithms to apply to LPR. Machine learning techniques such as CNNs and GANs combined with image processing are the best direction to move forward (Shaikh et al., 2016). The use of GANs to hallucinate missing input data is an area worth exploring as GANs have majorly been used for face generation, environments and more recently for single image super-resolution. It would be reasonable to explore the accuracy a GAN can achieve with LPR as few studies have explored the use of GANs to detect number plates. More specifically, an SRGAN can be used to increase the resolution of input images to aid with character recognition. A case for this is that character height should be at least 20 pixels tall for OCR to work effectively Anagnostopoulos (2014) therefore, upscaling an image of a number plate before getting information from it should improve the accuracy of LPR.

6 Plan of Action

6.1 Research Statement

GANs can improve licence plate recognition accuracy by synthesising high-resolution images from low-quality input – sustaining high accuracy across varying quality data.

6.2 Approach

Time will be allocated to the familiarisation of image processing techniques relevant to this area of study, such as edge detection, CCA and OCR. The majority of this will be done using OpenCV, which is a library tailored explicitly for computer vision. Lastly, research will be conducted to better understand GANs and CNNs together with their specific uses and how best to apply them. Through this research, it will be possible to tell which image processing techniques are considered redundant with advances in machine learning.

YOLOv2 or subsequent versions will be used for number plate localisation.

The potential system will be fed an input image, and YOLO will detect the location of the number plate. The ROI will be extracted and sent to a trained GAN that will upscale the image of the number plate. From there, image segmentation and OCR will be performed on the high-resolution image; the increased resolution of the image should provide complete information making the OCR results more accurate when high-resolution input is used.

The system's performance will be evaluated by feeding the number plate recognition system the original low-resolution dataset and then an augmented high-resolution dataset and then compare the accuracy obtained from each dataset. This approach will hopefully result in higher accuracy scores and give some insight as to how practical using super-resolution for LPR can be as it is not significantly addressed in current literature (Anagnostopoulos, 2014).

6.3 Deliverables and Dates

Deliverables and Dates	
6 April 2021	Seminar Series 1: Oral presentations
8 April 2021	Review related literature
	Test an existing LPR program with a few images from a dataset (both available online)
15 April 2021	Review more literature
	Test an existing GAN from demo code
22 April 2021	Review related literature
23 April 2021	Final proposal submission
26 April 2021	Ongoing process of learning practical skills needed to create system
21 May 2021	Hand in literature review
18 June 2021	Thesis Design chapter to be handed in to supervisor
23 June 2021	First Semester Exam Period Begins
5 August 2021	Have prototype version of a number plate detection system
9 August 2021	Seminar Series 2: Oral presentations
20 September 2021	Progress Report handed to supervisor
4 October 2021	Draft thesis sent to supervisor
11 October 2021	Short ACM paper hand-in
18 October 2021	Seminar Series 3:Final Presentations
29 October 2021	Project submission Deadline
	Website finalised
4 November 2021	Second semester exams begin
17 November 2021	Oral exams
3 December	Corrected ACM short style paper hand in
	Corrected project thesis hand in

References

Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. Understanding of a convolutional neural network. In *2017 International Conference on Engineering and Technology (ICET)*, pages 1–6. Ieee, 2017.

Christos-Nikolaos E Anagnostopoulos. License plate recognition: A brief tutorial. *IEEE Intelligent transportation systems magazine*, 6(1):59–67, 2014.

Taiwo Oladipupo Ayodele. Types of machine learning algorithms. *New advances in machine learning*, 3:19–48, 2010.

Sarthak Babbar, Saommya Kesarwani, Navroz Dewan, Kartik Shangle, and Sanjeev Patel. A new approach for vehicle number plate detection. In *2018 Eleventh International Conference on Contemporary Computing (IC3)*, pages 1–6. IEEE, 2018.

Giuseppe Bonaccorso. *Machine learning algorithms.* Packt Publishing Ltd, 2017.

John Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, 6:679–698, 1986.

Shyang-Lih Chang, Li-Shien Chen, Yun-Chung Chung, and Sei-Wan Chen. Automatic license plate recognition. *IEEE transactions on intelligent transportation systems*, 5(1):42–53, 2004.

Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015.

Juan Du. Understanding of object detection based on cnn family and yolo. In *Journal of Physics: Conference Series*, volume 1004, page 012029. IOP Publishing, 2018.

Shan Du, Mahmoud Ibrahim, Mohamed Shehata, and Wael Badawy. Automatic license plate recognition (alpr): A state-of-the-art review. *IEEE Transactions on circuits and systems for video technology*, 23(2):311–325, 2012.

Pedro Felzenszwalb, David McAllester, and Deva Ramanan. A discriminatively trained, multiscale, deformable part model. In *2008 IEEE conference on computer vision and pattern recognition*, pages 1–8. IEEE, 2008.

Muhammad Attique Khan, Muhammad Sharif, Muhammad Younus Javed, Tallha Akram, Mussarat Yasmin, and Tanzila Saba. License number plate recognition system using entropy-based features selection approach with svm. *IET Image Processing*, 12(2):200–209, 2017.

Rayson Laroca, Evair Severo, Luiz A Zanlorensi, Luiz S Oliveira, Gabriel Resende Gonçalves, William Robson Schwartz, and David Menotti. A robust real-time automatic license plate recognition based on the yolo detector. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–10. IEEE, 2018.

Donghyeon Lee, Sangheon Lee, Hoseong Lee, Kyujoong Lee, and Hyuk-Jae Lee. Resolution-preserving generative adversarial networks for image enhancement. *IEEE Access*, 7:110344–110357, 2019.

Younkwan Lee, JaeWoong Yun, Yoojin Hong, Juhyun Lee, and Moongu Jeon. Accurate license plate recognition and super-resolution using a generative adversarial networks on traffic surveillance video. In *2018 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*, pages 1–4. IEEE, 2018.

Alice Lucas, Santiago Lopez-Tapia, Rafael Molina, and Aggelos K Katsaggelos. Generative adversarial networks and perceptual losses for video super-resolution. *IEEE Transactions on Image Processing*, 28(7):3312–3327, 2019.

Syed Zain Masood, Guang Shu, Afshin Dehghan, and Enrique G Ortiz. License plate detection and recognition using deeply learned convolutional neural networks. *arXiv preprint arXiv:1703.07330*, 2017.

Zuhaib Ahmed Shaikh, Umair Ali Khan, Muhammad Awais Rajput, and Abdul Wahid Memon. Machine learning based number plate detection and recognition. In *ICPRAM*, pages 327–333, 2016.

Zhou Wang and Alan C Bovik. Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE signal processing magazine*, 26(1): 98–117, 2009.

Yuan Yuan, Siyuan Liu, Jiawei Zhang, Yongbing Zhang, Chao Dong, and Liang Lin. Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 701–710, 2018.