2022 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD) | 978-1-6654-8422-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/icABCD54961.2022.9856380

Investigating the Effects of Image Correction Through Affine Transformations on Licence Plate Recognition

Alden Boby[1] and Dane Brown[2] and James Connan[3] and Marc Marais[4]
Computer Science, *Rhodes University*
Grahamstown, South Africa
[1]boby.alden128@gmail.com, [2]d.brown@ru.az.za, [3]j.connan@ru.ac.za@, [4]marcmarais07@outlook.com

***Abstract*—Licence plate recognition has many real-world applications, which fall under security and surveillance. Deep learning for licence plate recognition has been adopted to improve existing image-based processing techniques in recent years. Object detectors are a popular choice for approaching this task. All object detectors are some form of a convolutional neural network. The You Only Look Once framework and Region-Based Convolutional Neural Networks are popular models within this field. A n ovel a rchitecture c alled t he W arped P lanar Object Detector is a recent development by Zou et al. that takes inspiration from YOLO and Spatial Network Transformers. This paper aims to compare the performance of the Warped Planar Object Detector and YOLO on licence plate recognition by training both models with the same data and then directing their output to an Enhanced Super-Resolution Generative Adversarial Network to upscale the output image, then lastly using an Optical Character Recognition engine to classify characters detected from the images.**

***Index Terms*—object detection, optical character recognition, generative adversarial network, spatial network transformer, super-resolution**

I. Introduction

Licence plates are a unique feature on vehicles used for identification [2]. T hrough t his u nique i dentifier, da ta ab out a registered vehicle can be obtained. This information is useful for surveillance, access control and traffic r egulations [3]. Licence plate recognition (LPR) systems are robust when working on data in constrained environments, i.e. for most access control systems, cameras are directly facing the car's licence plate entering the access control point. This produces a frontal view of the licence plate, presenting a best-case scenario to extract information.

Current literature tends towards the application of deep learning to improve the performance of LPR systems [4]. These methods are favoured over traditional image processing based techniques due to higher recognition rates. A specific problem arises when capturing licence plates in unconstrained conditions [5]. The angle at which an image is taken can greatly affect the results of LPR [6], specifically t he optical character recognition (OCR). This is because character recognition is sensitive to distortion [7]. A twisted or skewed character may not correlate to the features learnt by an OCR engine.

Existing LPR systems make use of deep learning models such as Region-Based Convolutional Neural Networks (R-CNNs) and the You Only Look Once framework (YOLO) [7]. These object detectors can automatically detect objects they have been trained to look for. Once an object is located, they place a bounding box around the Region of Interest (ROI). This favours licence plate detection, as all licence plates are quadrilaterals with varying dimensions depending on the vehicle type and region. Detection is near perfect for frontal images; the problem arises when the input images have licence plates at oblique angles [8]. This leads to a skewed image which is difficult to capture with a standard square or rectangle without introducing noise in the ROI. It also contributes to the distortion of characters within the image [7]. Correcting this requires a bounding box that fits the licence plate tightly and then applying a suitable affine transformation to correct the extracted ROI so that it appears as if it was captured from a frontal position. This cannot be done with standard YOLO as it only outputs a regular quadrilateral. One approach is to make use of the Warped Planar Object Detection Network (WPOD-net). This paper analyses and compares the performance of YOLOv3 and WPOD-net on licence plate detection as well as the performance of OCR through the use of output from both systems. Moreover, at an intermediate stage, an Enhanced Generative Adversarial Network will be used with each network to upscale output images to improve OCR.

II. YOLO

You only look once is a framework for identifying objects within an image in a single pass using a specialised convolutional neural network (CNN) architecture. The framework is described as 'state-of-the-art'. YOLO is based on the Inception Net backbone [9]. YOLO is popular for its high inference speed and accuracy. Its real-time architectural design allows it to be used on both images and video. Objects are detected and tracked through each frame in the case of videos. YOLO currently exists in five versions; the first three by the original authors, Redmon and Farhadi, and the last two revisions were contributions from the machine learning community.

This work was undertaken in the Distributed Multimedia CoE at Rhodes University.

The first version of YOLO is called YOLO9000; this name was derived from its ability to identify 9000 different classes or objects. YOLOv2 has significant improvements over its first iteration [11]. These come through using batch normalisation on the input data and pruning some convolutional layers from the network. The YOLO detector is a general-purpose object detector, making it attractive for licence plate detection.

YOLOv3 features additional improvements upon its first two iterations. Redmon and Farhadi [10] state that YOLOv3 is substantially better at utilising GPU resources. YOLOv3 uses a backbone called Darknet-53, which is comprised of fifty-three convolutional layers. Darknet-53 is said to outperform popular but more complex backbones such as ResNet-101 and ResNet-152 in accuracy whilst providing a $1.5\times$ speedup. Object detectors are commonly evaluated using mean average precision as well as intersection over union; these metrics are used to compare the performance of existing CNN based solutions [9]. YOLOv3 has improved performance on predicting bounding boxes on small objects compared with the previous iterations of the model [10]. One of the issues addressed is that input images for LPR systems may have cars far away from the camera, meaning the ROI will be very small and misaligned, which presents a use case for YOLOv3.

III. Convolutional Neural Networks

CNNs are a specialised artificial neural network (ANN) that extract features from a given set of training data. The sliding windows within a CNN allow the network to learn the features of an image regardless of where the region of interest lies. CNNs have simplified complex problems by doing most of the hard work, which would be complex if coded manually, as they have reduced parameters when compared to ANNs [12]. Convolutional Neural Networks are well suited to computer vision tasks, hence why many computer vision programs are based on CNNs.

CNNs are comprised of input, convolutional, active, pooling and fully connected layers. This architecture allows CNNs to extract features from images using mathematical logic and map data with the same features using weights [12]. CNNs are limited by the datasets used to train them and the hardware used for computation.

IV. Generative Adversarial Networks

A GAN comprises two neural networks: A discriminator and a generator. These networks work against each other, learning through adversarial loss [13]. The resulting images formed by GANs are sharper as they use adversarial loss rather than mean squared error (MSE) loss. MSE methods used by existing deep neural networks produce reasonable high-resolution output [14]. However, they do not utilise such networks' full potential, resulting in 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 [15]. More applicable to this field of study, a super-resolution generative adversarial network (SRGAN) can be used for image upscaling [16]. The SRGAN is fed low-resolution images and produces high-resolution images; the discriminator's job is to distinguish whether an input image is generated or a true high-resolution image [17]. The enhanced super-resolution generative adversarial network (ESRGAN) further improves the appearance of upscaled images by providing sharper details typically lost with standard super-resolution techniques [18].

A. Spatial Transformer Network

The Spatial Transformer Network (STN), developed by members of the DeepMind team at Google [19], is a module that can fit directly into an existing CNN and can transform feature maps to make the existing CNN more robust and improve its performance. The STN makes the CNN invariant to certain transformations such as rotation and skewing, enabling 'state-of-the-art' performance. Weihong and Jiaoyang [7] explored the use of STNs specifically for affine transformations for licence plate detection. They stated that the transformer itself can transform individual characters and not just a whole licence plate. This paper will not be looking directly at STNs but looks at a novel CNN method that draws inspiration from it [8].

B. Warped Planar Object Detection Network

WPOD-net is a novel CNN crafted specifically to detect and deskew licence plates using coefficient regression to perform affine transformations on an extracted bounding box [8]. What differentiates WPOD-net from other object detectors such as YOLO and R-CNN, which have been used for licence plate detection, is its ability to create irregular quadrilaterals as bounding boxes. Since WPOD-net retrieves accurate bounding boxes, it can utilise the coordinates of the bounding box to transform oblique bounding boxes into rectangular images that are more akin to an image taken from the frontal view.

V. Experimental Setup

The performance of an OCR system can be measured using character recognition rate (accuracy), precision and recall; All of which are defined below.

A. Detection Rate

The detection rate is a simple metric. It is the number of images that were returned with a bounding box after being fed into one of the object detectors. Some human input is required when inspecting these results as not all bounding boxes will be placed on a licence plate. Results with erroneous bounding boxes are not considered a positive detection.

B. Character Recognition Performance

There are three existing metrics used to measure the performance of text detection, character recognition rate, recall and precision.

The character recognition rate was used to measure the accuracy of predictions made at the OCR stage.

$$\text{character recognition rate } = n/(all + m). \tag{1}$$

where *n* represents the amount of correctly guessed characters, *all* represents the total number of characters present in the ground truth, and *m* represents the number of incorrect predictions or changes required to rectify errors in the prediction [20].

The formula for recall and precision are shown in equation 2 and 3

$$\text{recall} = n/all. \tag{2}$$

$$\text{precision} = n/(n+m). \tag{3}$$

C. Datasets

Four datasets were divided into groups. The MediaLAB LPR and Croatian Licence Plate datasets were combined as a training dataset for the YOLO model, WPOD-net and ESR-GAN. The validation of the object detection models followed an 80:20 split. One hundred images were used as a training set, and twenty-five pre-annotated images were used as a validation set. Transfer learning was used to supplement the small pre-annotated data set for both frameworks. All images fed into the YOLO model were downsampled to a resolution of 416×416. MakeML Car Licence Plates and Caltech dataset were used for testing. The following subsections break down the structure of the four datasets.

1) MediaLAB LPR Dataset: Sixty images were selected from the MediaLAB LPR Dataset to train the object detectors. The MediaLAB database consists of 12 categorised sets of images. The sets chosen were plates with dirt and shadows to familiarise the object detection models with challenging conditions.

2) Croatian Licence Plate Dataset: The dataset comprises 500 images containing several vehicles such as trucks, vans, SUVs and sedans. Images in the dataset are taken from the rear, frontal view and a few at oblique angles. All images in this dataset have a fixed resolution of 640×480.

3) MakeML Car Licence Plates Dataset : A dataset containing 433 images of cars taken at varying angles. There is inconsistency in the dataset in terms of angle and orientation of the ROI, making it prone to misalignment and consequently challenging. The resolution of the images varies between 400×400 and 600×450.

4) Caltech Dataset: The Caltech dataset consists of 126 images of vehicles from the Caltech Institute of Technology carpark. All images feature a resolution of 896×592. All licence plates in the dataset are of the rear end of a car. The majority of licence plates are Californian, with few from out of state.

VI. Results and Discussion

A. Experiment 1

Fifty images were passed through the YOLO-based LPR and WPOD-net LPR systems. The MakeML dataset was used to measure the performance of the object detectors; these fifty images were selected specifically because of their relative difficulty. The final output from the system was a string of characters representing what was detected on a licence plate. In Table I, a comparison between the performance of WPOD-net and YOLO can be seen.

TABLE I
Detection Rate.

Dataset	Architecture	Detection Rate (%)
MakeML	YOLOv3	92
	WPOD-net	98
Caltech	YOLOv3	75
	WPOD-net	100

As evident from the results in Table I, the WPOD-net architecture performed better when detecting images from the selected datasets, with an impressive detection rate of 100% on the Caltech dataset, 25% higher than the detection rate of the YOLO model on the same dataset. The WPOD-net architecture has a slight advantage as it is developed with the intention of detecting licence plates and produces more precise bound box coordinates as opposed to YOLOv3, which can only produce equiangular bounding boxes. Additionally, the YOLOv3 architecture uses the COCO image set as it is a general-purpose object detector. Therefore, the WPOD-net already has licence plate features to work with at the transfer learning stage, while the YOLO model does not.

The WPOD-net effectively detected licence plates that the YOLOv3 model was unable to detect, leading to higher detection rates. The images the YOLO model failed to predict were either at an oblique angle or had small ROIs. Shown in Fig 1 is a licence plate that was correctly predicted by the WPOD-net and entirely missed by the YOLO architecture.

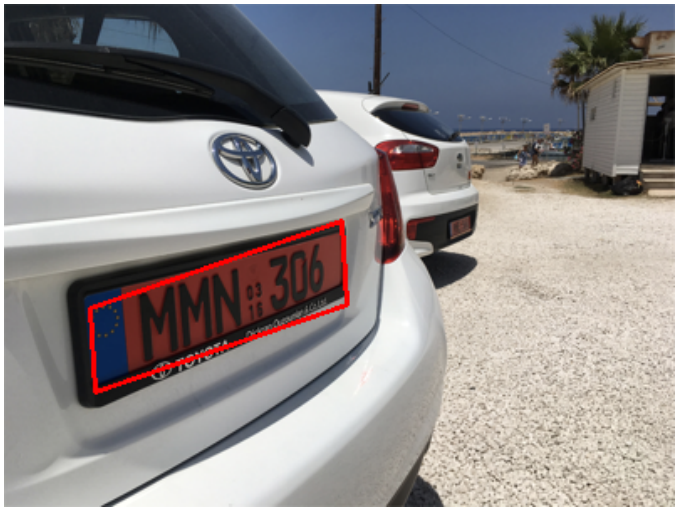

Fig. 1. The skewed licence plate was not captured by the YOLO model but was detected by the WPOD-net.

Fig. 2 and Fig. 3 show a comparison between the bounding boxes predictions by WPOD-net vs the ones produced by YOLOv3. The WPOD-net produces a much more accurate bounding box as observed through visual inspection. The Bounding box from the YOLO model is large and includes

some background noise which can interfere with OCR at a later stage.

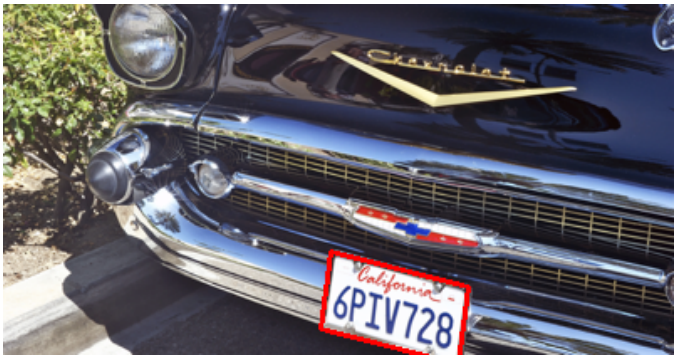

Fig. 2. An angled licence plate predicted by the WPOD-net.

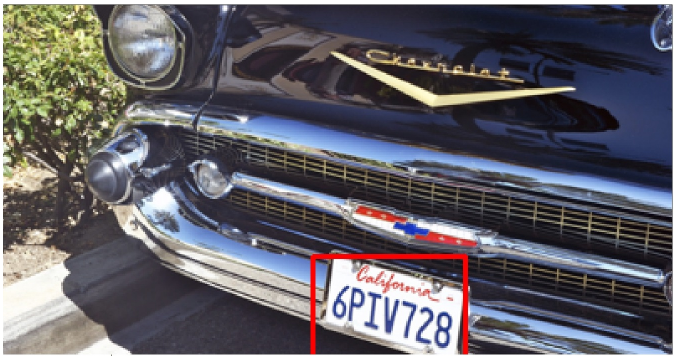

Fig. 3. The same angled licence plate from Fig. 2 predicted by YOLOv3.

The MakeML dataset includes more complex images when compared to the Caltech dataset. Both models failed to predict the licence plate shown in Fig. 4. The vehicle's licence plate closely matches the colour of the car, masking the edges of the licence plate making it difficult to detect. While the WPOD-net detected most images with ease, the YOLO model struggled with a few more licence plates with similar characteristics to the example from the MakeMl dataset shown in Fig. 4.

The WPOD-net has the ability to detect more than one licence plate in a given image. This, on rare occasions, caused the detector to detect a licence plate and any additional text on the vehicle that resembled features from a licence plate. Although a minor issue, it can be rectified by training the model on more data. Fig. 5 shows output from the WPOD-net, including one positive and one negative detection in the same image.

B. Experiment 2

The output from both object detection architectures was upscaled using an ESRGAN to measure its effectiveness on improving OCR. A total of fifty images were used for the OCR data. High-resolution (HR) licence plate images and their low-resolution (LR) counterparts from the ESRGAN were fed into the OCR engine, and the results were recorded. Character

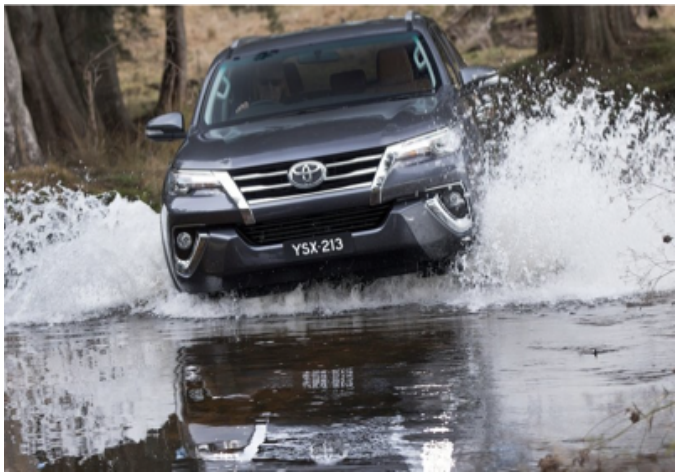

Fig. 4. The colour of the licence plate made it difficult for both models to locate the ROI.

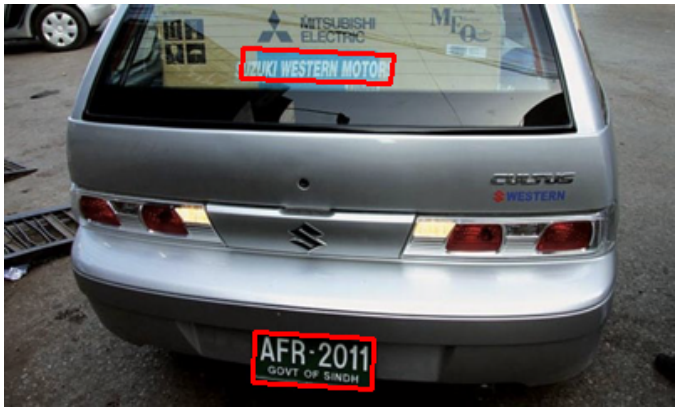

Fig. 5. A false positive is highlighted on the windshield of the car in the image.

recognition rate, precision and recall were calculated for the HR and LR images and are displayed in II.

The overall OCR metrics show that the WPOD-net comes out on top but only by a slight margin. However, the performance increase observed from using the ESRGAN was more effective for the YOLOv3 model as increasing the image resolution saw a 7.46% jump in accuracy. In contrast, the WPOD-net model only saw a 1.32% increase in OCR accuracy. The WPOD-net, however, did have a higher overall score for all the metrics shown in Table. II.

The WPOD-net output increased the OCR accuracy by 6.86% without upscaling the images at all, proving the effectiveness of correcting the image through an affine transformation before sending it to the OCR engine for classification. Conversely, the YOLO model benefits largely from upscaling from the ESRGAN, as after upscaling the images, the gap between the performance of the two models is reduced from 6.86% to just 0.72%.

Shown in Fig. 6 is an example of an image with a skew bounding box, and the resulting image transformed by the WPOD-net in Fig. 4.

TABLE II
OCR RESULTS.

Dataset	Architecture	Resolution	
		LR(%)	*HR(%)*
Accuracy	YOLOv3	58.44	65.90
	WPOD	65.30	66.62
Recall	YOLOv3	67.59	73.88
	WPOD	75.44	78.04
Precision	YOLOv3	64.22	70.75
	WPOD	72.84	72.89

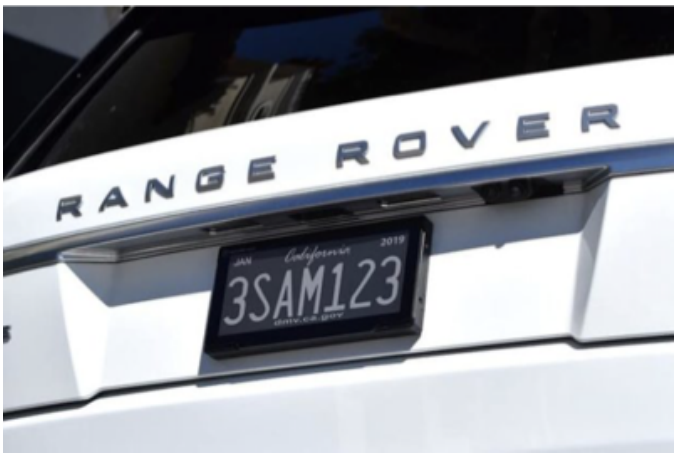

Fig. 6. The angled licence plate was only predicted by the WPOD architecture.

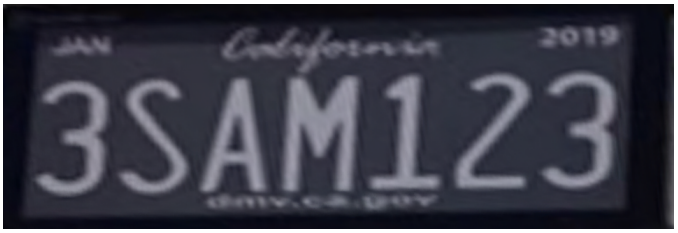

Fig. 7. The ouput image from the WPOD model for the image in Fig. 6.

VII. CONCLUSION

In conclusion, the results signify that the affine transformations and tight bounding boxes from the WPOD-net improved accuracy when compared with an equivalent YOLO model trained on the same dataset. Image correction improves licence plate recognition indeed, and this is best signified by the increase in detection rate as well as character recognition rate when comparing the two models.

It was shown that the WPOD-net was consistent in applying affine transformations to correct an image. The additional transformation step enabled the OCR algorithm to produce better results than when presented with a regular oblique image, as perspective can skew characters and alter predictions from the OCR engine. WPOD-net, however, saw few benefits from upscaling through the ESRGAN when compared to YOLO. The results also show that upscaling an image significantly improved recall at the OCR stage for both object detection models.

In future, it would be beneficial to alter the parameters for the WPOD-net so that it does not reduce the resolution of its output images at the expense of more computational power. Moreover, for the fairness of comparison, Tesseract OCR was used instead of the usual CNN attached to WPOD-net for OCR. It would be worth exploring the effect upscaling through an ESRGAN would have on the results from the said OCR method.

REFERENCES

[1] Y. Zou, Y. Zhang, J. Yan, X. Jiang, T. Huang, H. Fan, and Z. Cui, "License plate detection and recognition based on YOLOv3 and ILPRNET," *Signal, Image and Video Processing*, pp. 1–8, 2021.

[2] X. Liu, W. Liu, T. Mei, and H. Ma, "Provid: Progressive and multimodal vehicle reidentification for large-scale urban surveillance," *IEEE Transactions on Multimedia*, vol. 20, no. 3, pp. 645–658, 2017.

[3] J. Han, J. Yao, J. Zhao, J. Tu, and Y. Liu, "Multi-oriented and scale-invariant license plate detection based on convolutional neural networks," *Sensors*, vol. 19, no. 5, p. 1175, 2019.

[4] R. Laroca, E. V. Cardoso, D. R. Lucio, V. Estevam, and D. Menotti, "On the cross-dataset generalization for license plate recognition," *arXiv preprint arXiv:2201.00267*, 2022.

[5] M. Lin, L. Liu, F. Wang, J. Li, and J. Pan, "License plate image reconstruction based on generative adversarial networks," *Remote Sensing*, vol. 13, no. 15, p. 3018, 2021.

[6] L. Xie, T. Ahmad, L. Jin, Y. Liu, and S. Zhang, "A new cnn-based method for multi-directional car license plate detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 2, pp. 507–517, 2018.

[7] W. Weihong and T. Jiaoyang, "Research on license plate recognition algorithms based on deep learning in complex environment," *IEEE Access*, vol. 8, pp. 91 661–91 675, 2020.

[8] S. M. Silva and C. R. Jung, "A flexible approach for automatic license plate recognition in unconstrained scenarios," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[9] J. Du, "Understanding of object detection based on CNN family and YOLO," in *Journal of Physics: Conference Series*, vol. 1004. IOP Publishing, 2018, p. 012029.

[10] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018.

[11] S.-H. Park, S.-B. Yu, J.-A. Kim, and H. Yoon, "An all-in-one vehicle type and license plate recognition system using yolov4," *Sensors*, vol. 22, no. 3, p. 921, 2022.

[12] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)*. Ieee, 2017, pp. 1–6.

[13] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 53–65, 2018.

[14] A. Lucas, S. Lopez-Tapia, R. Molina, and A. K. Katsaggelos, "Generative adversarial networks and perceptual losses for video super-resolution," *IEEE Transactions on Image Processing*, vol. 28, no. 7, pp. 3312–3327, 2019.

[15] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? a new look at signal fidelity measures," *IEEE signal processing magazine*, vol. 26, no. 1, pp. 98–117, 2009.

[16] T.-G. Kim, B.-J. Yun, T.-H. Kim, J.-Y. Lee, K.-H. Park, Y. Jeong, and H. D. Kim, "Recognition of vehicle license plates based on image processing," *Applied Sciences*, vol. 11, no. 14, p. 6292, 2021.

[17] D. Lee, S. Lee, H. Lee, K. Lee, and H.-J. Lee, "Resolution-preserving generative adversarial networks for image enhancement," *IEEE Access*, vol. 7, pp. 110 344–110 357, 2019.

[18] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, "Esrgan: Enhanced super-resolution generative adversarial networks," in *The European Conference on Computer Vision Workshops (ECCVW)*, September 2018.

[19] M. Jaderberg, K. Simonyan, A. Zisserman *et al.*, "Spatial transformer networks," *Advances in neural information processing systems*, vol. 28, 2015.

[20] M. Shen and H. Lei, "Improving OCR performance with background image elimination," in *2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*. IEEE, 2015, pp. 1566–1570.