

Received February 21, 2022, accepted March 1, 2022, date of publication March 8, 2022, date of current version March 24, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3157714

License Plate Image Analysis Empowered by Generative Adversarial Neural Networks (GANs)

IBRAHIM H. EL-SHAL¹, **OMAR M. FAHMY**^{2,3}, (Senior Member, IEEE),
AND MUSTAFA A. ELATTAR¹

¹Informatics Sciences Center, School of Information Technology and Computer Science, Nile University, Giza 41516, Egypt

²Electrical Engineering Department, Badr University, Cairo, Egypt

³Electrical Engineering Department, South Ural State University, 454080 Chelyabinsk, Russia

Corresponding authors: Ibrahim H. El-Shal (i.hamed@nu.edu.eg), Omar M. Fahmy (omar.fahmy@buc.edu.eg), and Mustafa A. Elattar (melattar@nu.edu.eg)

ABSTRACT Although the majority of existing License Plate (LP) recognition techniques have significant improvements in accuracy, they are still limited to ideal situations in which training data is correctly annotated with restricted scenarios. Moreover, images or videos are frequently used in monitoring systems that have Low Resolution (LR) quality. In this work, the problem of LP detection in digital images is addressed in the images of a naturalistic environment. Single-stage character segmentation and recognition are combined with adversarial Super-Resolution (SR) approaches to improve the quality of the LP by processing the LR images into High-Resolution (HR) images. This work proposes effective changes to the SRGAN network regarding the number of layers, an activation function, and the appropriate loss regularization using Total Variation (TV) loss. The main paper contribution can be summarized into presenting an end-to-end deep learning framework based on generative adversarial networks (GAN), which is able to generate realistic super-resolution images. Also, proposed adding a TV regularization to the loss function to help the model enhance the resolution of images. The proposed SRGAN can handle tiny 72×72 images of LPs. The paper explores how SRGAN performed over different datasets from many aspects, such as visual analysis, PSNR, SSIM, and Optical Character Recognition (OCR). The experiments demonstrate that the suggested SRGAN can generate high-resolution images that improve the accuracy of the license plate recognition stage compared to other systems.

INDEX TERMS Computer vision, deep learning, generative adversarial networks, image reconstruction, license plate recognition, single image super-resolution, total variation loss.

I. INTRODUCTION

Automatic License Plate Recognition (ALPR) is a computer vision technology that efficiently identifies a vehicle's registration plate from images without the necessity for human involvement. Traffic management, law enforcement, toll collection, and vehicle owner identification have become major issues globally. Therefore, the ALPR framework should be developed as one of the potential solutions. Recently, several ALPR systems have been proposed.

In recent years, the majority of ALPR applications have been based on real-time detection or recognition of licence plates. As a result, there are certain drawbacks since they depend on the vehicle's availability within a short-range.

The associate editor coordinating the review of this manuscript and approving it for publication was Hengyong Yu.

Otherwise, non-real-time applications rely on improving the quality of images, including license plates, to improve the accuracy of object detection at large distances. Although ALPR systems are based on specific methodologies, it is still a particularly challenging task because some of the variables, such as high vehicle speed, and non-uniform vehicle registration plates, will significantly affect the overall rate of recognition and the expansion of the video camera deployment in every intersection under the Intelligent Transportation System will cause the production of an enormous number of video streams. The environmental conditions and the variety of registration plates are the primary concerns of the license plate recognition problem.

Consequently, the environmental side, such as varying illumination, colour, dirt, shadows, or background patterns, significantly influences number plate recognition. Hence, vary-

ing illumination can degrade the quality of the vehicle image, and background patterns add extra difficulty to the number plate location process. Otherwise, registration plate position, quantity, size, font, colour, character sharpness, language, and tendency all pose significant challenges in constructing a reliable ALPR framework. The recognition unit is often installed at the gate of the residential area, toll gates, or other highly secured facilities like defence institutes and nuclear plant facilities [1].

There are three main phases of a traditional LP recognition system. **Firstly**, LP detection is based on handmade features in the whole image by using bounding boxes to locate license plates in an image. These features can be grouped into four classes based on edges, colours, texture, and characters [2], [3].

- 1) **Edge-based Approaches** are frequently used to detect LP because they are usually rectangular with a particular aspect ratio and have a higher edge density than the rest of the image. Zhang *et al.* [4] have formulated an edge-based strategy by introducing a weak chain classifier that uses a Haar-like feature with adaptive boosting for LP extraction. The gradients of the input vehicle images are used as proposed by Wang *et al.* [5] to detect possible LP regions. They are divided into various adjacent areas, and the one with the highest value is picked. The edge cluster used Expectation-Maximization (EM) [6] to extract areas with dense edge sets identical to the candidate license plates. Sappa *et al.* [7] introduced a collection of arithmetic operations between a given frame and two equidistant ones used to compute a coarse moving edge representation. Non-desired edges are filtered with the filtering technique. The Line Density Filter (LDF) technique [8] links regions of high edge density in the binary edge image using the Sobal operator and Adaptive Thresholding (AT). Although edge-based methods are quick to compute, they cannot be used on complex images because they are very sensitive to sharp edges.
- 2) **Color-oriented Approaches** are based on the fact that the color of the LP is typically distinct from that of the vehicle. Deb *et al.* [9] presented that the LP automatically selects the HSI colour space for statistical threshold values based on two major stages. At first, HSI (Hue, Saturation and Intensity) colour space recognises candidate regions. These candidate regions can include LP regions based on different colours of the LP; then, the license plate's geometrical properties are used for classification. The technique proposed by Chen *et al.* [10] suggests a feature-salient approach. The Hough Transform (HT) detects vertical and horizontal lines. The number plate is then segmented by transforming it from red, green, and blue (RGB) to HSI. They are extremely sensitive to different lighting conditions in natural scene images. They are unable to differentiate other objects in the image that are the same colour and size as the LPs.

- 3) **Texture-based Approaches** focus on detecting LP in plate regions based on the unconventional distribution of pixel intensity. A Discrete Wavelet Transform (DWT) [11] was used to extract details from an image. Also, local features reduce the sensitivity of moving object detection methods to changes in illumination [12]. Texture-based techniques incorporate more discriminative characteristics than edge-based or colour-based strategies, requiring more computational resources.

Secondly, segmenting the whole plate into single blocks of character [13]. **Thirdly**, recognize the segmented characters with handcrafted features [14], [15] and pre-designed classifiers like Support Vector Machine (SVM) [16], Naive Bayes algorithm [17]. The above-mentioned ALPR methods cannot be applied to existing surveillance systems because of their limitations. It can be effective only for specific tasks and not be generalized for different applications or different low-resolution images as well as the vehicle. Therefore, current ALPR [18]–[20] systems only apply to regular high resolution and official LP patterns, which demonstrates that in a tough environment, current solutions do not fully utilize a large number of well-valued traffic surveillance cameras.

Our main contribution is to develop a technique to alleviate the problems of ALPR in challenging environments.

- 1) An end-to-end deep learning LP system with SR and recognition for unrestricted urban surveillance sensing.
- 2) A SR approach could be applied to enhance the quality of vehicle plate images by handling a single Low-Resolution image or a sequence of LR images into a single HR image.
- 3) License plate detection and optical character recognition (OCR) uses the state-of-the-art and real-time object detection method, You Only Look Once (YOLO).
- 4) Training techniques are different from previous works by training low-resolution images for a license plate with a 72×72 size, and evaluation performance of the proposed methods are applied over different datasets.

The structure of this paper is organized as follows. Section II gives a brief discussion of related work. Section III explores the generative adversarial network architectures used in this work to enhance the resolution of license plate images. Experimental verification and evaluation against other techniques are followed in section IV, and discussion and conclusions are drawn in Sections V and VI as follows.

II. RELATED WORK

To overcome the previous methods' drawbacks, Deep Convolutional Neural Networks (DCNNs) [21], [22] or Artificial Neural Networks (ANN) [23] present many hidden layers to learn high-level features in order to extend their capacity so that they can generalize not only the target re-identification function but also other computer vision problems, including image classification, object detection,

semantic segmentation, and video tracking. As an example of using DCNNs in generalization for face detection and recognition, the raw input data is represented in the pixel matrix form, in which the pixels are abstracted. The edges are covered in the first-pixel representation layer; the edge layer is formed and encoded. Next, eyes and noses are encoded in the next layer, and then the face is recognized in the final layer. Features are extracted using a histogram of the directed gradients method from the area of interest in the recognition stage [24], [25].

A. ADVERSARIAL LEARNING

The Generative Adversarial Network (GAN) is a way to train deep neural networks for generative models that attempt to learn probability distributions from input data. GAN was initially intended to generate more realistic fake images, but past surveys have shown that this adversarial strategy can also create complex training algorithms. Examples include generative-focused tasks, super-resolution, style transfer, natural-language processing, and discriminative-focused tasks, as well as human pose estimation.

Car plate Super-Resolution (SR) aims to retrieve and recognize plate images to be identified by humans and computers with higher accuracy. Numerous previous techniques attempt to enhance SR images on the license plate, which can be categorized into three groups.

1) Interpolation-based Approaches:

The approach based on interpolation [26], [27] has its benefits and drawbacks, based on the number of pixels to process. Several researchers have been working on developing improved interpolation-based algorithms. Sarmadi *et al.* [28] suggest a new strategy, which describes the Single Image Super-Resolution (SISR) system in three phases. The first phase of the interpolation is the up-sampling phase, and then in the second phase, which is called the de-blurring phase, the blur is extracted using a single probabilistic model. Finally, the de-noising step, in which noise is extracted using a spatially adaptive iterative process. The SR approach proposed by Ling *et al.* [29] extracts subcategories by subpixel decomposition based on spatial dependency. The feature-level data is used in [30] to enhance the image quality generated by the Self-Interpolation Generative Adversarial Network (SIGAN) and the Channel Interpolation Generative Adversarial Network (CIGAN). Dai *et al.* [31] used the prior images to generate a better image in terms of PSNR.

2) Reconstruction-based Approaches:

The technique relies on reconstruction and does not require a training set, but the HR images must be properly restricted so that super-resolved images improve their quality. The explicit distribution or energy function defined in the class contains the prior knowledge required to resolve the SISR unconditional problem. Sun *et al.* [32] use gradient profiles to estimate a HR

image from a LR image by defining the shape and sharpness of image gradients. The proposed technique that Tai *et al.* [33] suggest is that the gradient map conveys potential edge information and texture information, or high-frequency data, which is crucial for detailed texture synthesis. Sen *et al.* [34] proposed an application for enforcing the restriction in the wavelet domain such that the HR image is sparse. Image patch-based methods have been proposed in [35].

3) Learning-based Approaches:

The following Deep Learning (DL) based studies have been carried out to select the optimal mapping function. A new method has been developed by Yang *et al.* [36] for clustering patch spaces and learning a mapping from low-resolution patches to their high-resolution counterparts. Ledig *et al.* [37] introduce the SRGAN based on the Super-Resolution Restoration (SRR) method for super-resolution. Also, the authors proposed a perception loss function consisting of a content loss and a loss of opposition. The SRR method [38] is based on neural network convolution (CNN). The proposed CNN model combines extraction, sampling, and reconstruction of the features. Then the Deep CNN (DCNN) can be trained end-to-end. The progressive GAN in [39] is typically used to convert LR to HR images using a dataset of retinal colour fundus images. The SRR approach is used in [40] with improved SRGAN (ESRGAN). In this analysis, ESRGAN has developed the architecture, adversarial loss, and perceptual loss of SRGAN. The SR image for LP was produced from blurry, tiny images as proposed by Lee *et al.* [41]. Lai *et al.* [42] suggested that the perceptual loss is a solution to the smoothing SR for the LP problem. Architecture based on the Generator (G) network consists of 16 ResNet blocks with skip-connections using the ReLU activation function. The Discriminator (D) network is implemented with one input convolution layer and 10 hidden convolution layers using LeakyReLU. Shamsolmoali *et al.* [43] summarized several generation methods of synthesis images and described their categories. Furthermore, the researchers analyzed the GAN related to architecture, loss functions, and datasets. For instance, convolution GANs convert an actual image into a multi-scale pyramid image by training the network to generate multi-scale and multi-level feature maps and then merging them to produce the final feature map. CycleGAN proposed a domain adaptation adversarial network regarding unsupervised segmentation.

III. PROPOSED TECHNIQUE

The performance of the ALPR system is based on the precise identification algorithm of the car plate and the acquisition quality. In practice, LR images or videos are often utilized in surveillance systems. The text on the vehicle plate might be unreadable because of distance, light, and distortion of

perspective in low-resolution surveillance systems. A combination of higher accuracy and low processing time should be considered to accomplish this.

The following section describes our optimizations and contributions to the GAN architecture from two main perspectives related to the generator and discriminator networks that enhance the resolution of low images of license plates.

A. SYSTEM MODEL

GAN is a framework represented by Ian Goodfellow [44], as shown in Figure 1. It addresses the issue of unsupervised learning by training two deep neural networks, known as Generator (G) and Discriminator (D), that compete and collaborate to estimate generative models using adversarial methods. The game of zero-sum is the fundamental concept of the GAN model. The GAN trains the network to achieve Nash equilibrium [45] by learning adversaries in order to achieve the goal of estimating data's potential distribution and producing new data samples.

Two input sets of variables are used, x and z . Set z , which indicates the random variables used by the Generator (G) in order to generate $G(z)$. Set x , which refers to the real data used by the Discriminator (D). By comparing the outcome of the discriminator with $G(z)$, which denotes the generated result by the generator that corresponds to the distribution of real-world samples p_{data} to the greatest possible extent. The output will be 1, If the discriminator's input is the actual data, otherwise, the discriminator returns 0.

To keep $V(G;D)$ stable, the following conditions must be satisfied with the following minimax equation.

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)}[\log D(x)] + E_{z \sim p_z(z)}[\log(1 - D(G(z)))] \quad (1)$$

The general idea behind this formulation (1) is that it enables the training of a generative model G with the intention of deceiving a differentiable D trained to differentiate super-reconstructed images from the real-world data. In other words, G and D compete with each other and iteratively optimize until D is unable to differentiate whether the input sample is derived from G or real samples. As of this point, goal G has been achieved. Using this procedure, G can be taught to generate samples that are strikingly close to actual images and thus problematic to classify by D .

B. PROBLEM FORMULATION

Following Goodfellow [44], a discriminator network is also described as D_{θ_D} , which we alternately optimize with G_{θ_G} . Therefore, these subnetworks have two main goals. The first goal is to reduce their own costs, and the second one will be to increase the opposite cost (adversarial min-max problem). Adversarial mathematics is defined as below 2. Using this approach, G can be trained to produce samples that closely correspond to real images, making it challenging for D to

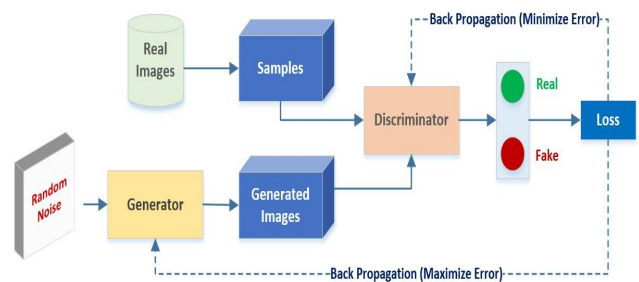


FIGURE 1. The primary architecture of the generative adversarial network (GAN).

classify them correctly.

$$\min_{G_{\theta_G}} \max_{D_{\theta_D}} V(D, G) = E_{I^{HR} \sim p_G(I^{HR})}[\log D_{\theta_D}(I^{HR})] + E_{I^{LR} \sim p_G(I^{LR})}[\log D_{\theta_D}(G_{\theta_G}(I^{LR}))] \quad (2)$$

where $p_G(I^{HR})$ is the distribution of HR image, I^{HR} is the HR image sample for $p_G(I^{HR})$, $p_G(I^{LR})$ is the distribution of generator based on LR image, I^{LR} is the LR image sample for $p_G(I^{LR})$, $G_{\theta_G}(I^{LR})$ is the Generator Network, and $D_{\theta_D}(G_{\theta_G}(I^{LR}))$ is the Discriminator Network.

C. ADVERSARIAL NETWORK ARCHITECTURE

Few modifications have been made to the architecture of SRGAN [37]. Figure 2 shows the overview of our proposed ALPR model based on Super-Resolution GAN. We determined and extracted the Region of Interest (ROI) around license plates of all input images in the dataset and then down-sampled them using the bilinear interpolation approach. The downsampled (LR) images were input to the generator network to begin the training process. The output images from the generator network will be super-resolution images. The generated images are then fed into the discriminator, which distinguishes between generated and authentic images. After the network is trained, the output super-resolution images are passed to the OCR framework, which uses Yolov5 to recognize characters on license plates in the final stage.

1) GENERATOR NETWORK

Our main goal is to develop the G Network to produce realistic super-resolution vehicle plate images from LR images. To accomplish this, we made very few changes to the G network. The figure 2 shows the proposed Generator Network.

- 1) Instead of deep convolution networks, the generator architecture contains residual networks because residual networks are easy to train and much deeper for better results. The entire network uses a kind of link known as skip connections [46].
- 2) The input layer has a convolution layer with a 9×9 kernel size, and a 1×1 stride with a padding of 4.
- 3) Then, with 16 ResNet blocks, there are two convolution layers in each residual block, small 3×3 kernels in

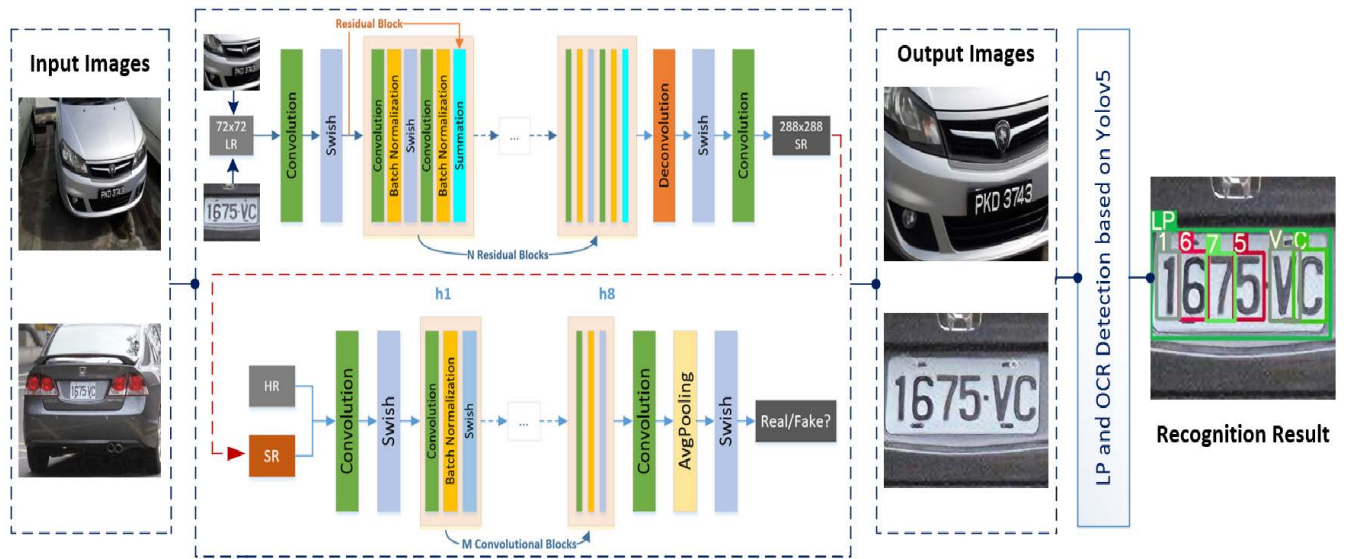


FIGURE 2. An overview of the proposed super-resolution license plate system and YOLO-based recognition system.

each, and a 1×1 stride at the input layer with padding of 1, this is followed by 64 feature maps and then a layer of batch-normalizations.

- 4) The upsampling blocks is performed using a 3×3 filter kernel with 64 feature maps, a stride of 2.
- 5) In each mini-batch, we cropped 288×288 randomly chosen from preprocessed HR images (384×384), then downsampled using scaling factor 4 to obtain LR images of size 72×72 to train the generator network to generate SR sub-images of size 288×288 .
- 6) All the convolution layers use the Swish activation function.

In the training process, the HR image is downsampled to the LR image using the bilinear operator, making the new resolution equal to 72×72 . The Generator architecture then attempts to upsample the images from low resolution to a fake super-resolution image using the same high-resolution image size of 288×288 . Consequently, a fake super-resolution image is fed to the discriminator. The discriminator attempts to differentiate between a fake super-resolution image and a high-resolution image. Then the discriminator computes the adversarial loss and content criterion loss, which is then backpropagated into the generator architecture.

2) DISCRIMINATOR NETWORK

As previously mentioned, the Discriminator Network’s task is to differentiate between the real image and the image generated by the Generator Network.

- 1) The network contains one input convolution layer and seven hidden layers with 3×3 filter kernels.
- 2) For each convolution layer, the feature maps are expanded from 64 to 512.
- 3) When the number of feature maps is raised by a factor of 2, a 2×2 stride is applied to degrade resolution.

- 4) The last convolution layer (h9) applies a 1×1 filter kernel, a 1×1 stride, and a 512 feature map.
- 5) Using the Swish activation function according to equation 3, the sample classification probability is obtained.

A probability range of 0 to 1 reveals how real the supplied image is. In the case of a 0, the D identifies the input image as a generated image, and the generator network updates parameter θ_G according to 2. A higher probability means the D has identified the input image as real, and the Discriminator Network’s parameters have been changed accordingly. The figure 2 shows the proposed Discriminator Network.

D. ACTIVATION FUNCTION

The selection of activation functions in DNNs has a significant impact on the training dynamics and task performance. According to research by the Google brain team [47], the swish activation function 3 is an alternative to ReLU, although the cost is much higher for the computation for both feed-forwarding and back-propagation. Research demonstrates that ReLU is over-performed by the new activation function for the deep neural network for multiple reasons:

$$f(x) = x * \delta(x) = \frac{x}{1 + e^x} \quad (3)$$

- 1) Weights that are highly negative are simply ignored, thus ($x \approx 0$).
- 2) For large values, the output is not saturated to its maximum value.
- 3) A smooth bend implies that its output will be smooth. It has the advantage of optimising the model in terms of convergence to the minimal loss.
- 4) In ReLU, all negative values are zero ($f(x) = 0$ for $x < 0$), where negative values might remain important in the collection of the data underlying patterns. This

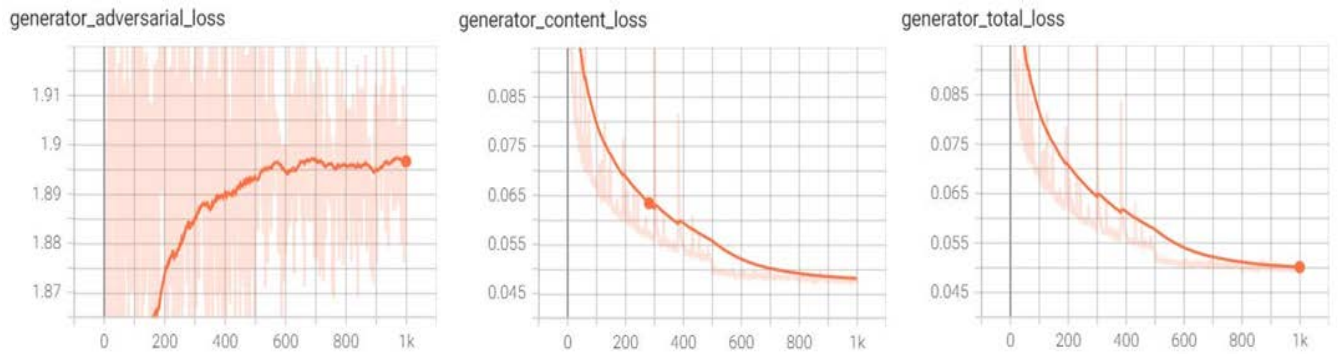


FIGURE 3. Losses of generator during the training process - Model I.

advantage is due to the smoothness of the Swish function and the equivalent values of $(f(x) < 0 \text{ for } x \approx 0)$.

E. LOSS FUNCTION

The loss function is a crucial factor in the efficiency of the generator network. The SRGAN model generates a realistic and sharper image than the previous CNN-based technique [48], with the notion of generator training combining adversarial losses [37] and MSE losses in Resnet [46]. The adversarial loss function of the generator has the following definition:

$$l_G = l_{MSE}^{SR} + \lambda l_D \quad (4)$$

where the content loss l_{MSE}^{SR} represents the MSE loss between the output $G_{\theta_G}(I^{LR})$ of the generator network (G) and the groundtruth I^{HR} . While the adversarial loss l_D is computed from the output of the discriminator (D) network. λ is 10^{-3} multiplied with the adversarial loss to achieve better behaviour of the gradient.

1) MSE LOSS FUNCTION

The pixel-wise MSE loss is the most common type of loss used in super resolution (SR), which is defined as:

$$l_{MSE}^{SR} = \frac{1}{r^2WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} (I_{x,y}^{HR} - G_{\theta_G}(I^{LR})_{x,y})^2 \quad (5)$$

W and H refer to the width and height of the image, respectively. x and y are the width and height of the groundtruth image (I^{HR}), and r is the scale ratio. MSE loss can not handle high-frequency information in the reconstructed image, leading to extremely smooth images.

2) ADVERSARIAL LOSS FUNCTION

The Adversarial loss is a loss function that leads the generator to produce images that are more similar to high-resolution images by employing a discriminator that has been trained to discriminate between HR and SR images. Adversarial losses are more effective at retrieving the high frequency (sharp edge) content of the generated image. Besides the equations

discriminator_loss

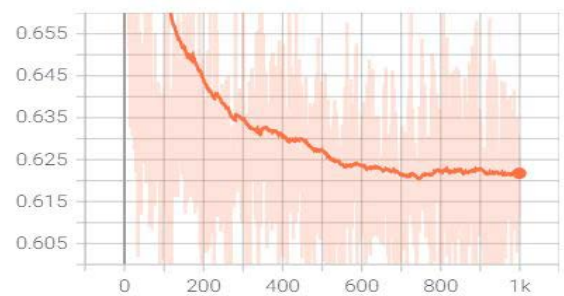


FIGURE 4. Losses of discriminator during the training process - Model I.

mentioned so far, we additionally include the output of the swish function (probability computed) of the discriminating network and compute the discriminating loss as follows:

$$l_D = \sum_{n=1}^N -\log D_{\theta_D}(G_{\theta_G}(I^{LR})) \quad (6)$$

The equation symbols are as follows: N is the number of epochs, $G_{\theta_G}(I^{LR})$ is a parameter of the image that was created from a low-resolution input image, and $D_{\theta_D}(G_{\theta_G}(I^{LR}))$ describe the probability of the generated image from generator network.

3) TOTAL VARIATION LOSS FUNCTION

The total variation (TV) is used as a gradient update, making it significantly less complicated during the backward propagation.

By including 2D-TV regularisation term [49], [50] in the proposed framework, we improve the quality of the generated images. Alternatingly, both the G and D networks are optimized by optimizing this loss function. Consequently, the loss function of the SRGAN model is described as:

$$\begin{aligned} \min_{\theta_G} \max_{\theta_D} V(D, G)_{GAN-TV} \\ = E_{x \sim p_{data(x)}} [\log D(x)] \\ + E_{z \sim p_z(z)} [\log(1 - D(G(z)))] + \lambda TV(G(z)) \quad (7) \end{aligned}$$

F. BATCH NORMALIZATION

Batch normalization (BN) is a covariance shift reduction technique that is applied to the input of each layer (distribution of activation changes in intermediate layers) [51]. Not only does BN removal enhance the image quality [40], but it is also known for permitting a more extensive range of selections for hyperparameters. Batch Normalization allows the network to converge more easily and reduces the computational cost caused by increasing hyperparameters dramatically when we use a different image size.

IV. EXPERIMENTS

The objective of the super-resolution of the LP is to form an accurate recognition of the LP numbers and characters. The conventional assessment criteria for issues with super-resolution often pick a peak signal-to-noise ratio (PSNR) and mean squared error (MSE). The proposed method results introduce evaluation criteria that rely on typical methods with metrics such as the Structural Similarity Index (SSIM), perceptual quality, and automatic license plate recognition. These metrics are used to compare the synthesized output from the proposed models of SRGAN with the existing methods such as Deep Image Prior and SRCNN architectures over the ALOP dataset. For ALPR, Yolo is used to detect the LP and its characters. Consequently, the recognition rate is calculated based on the correct six characters.

This paper discusses two proposed models of SRGAN with different loss approaches that are presented and analyzed. The Model I uses content and adversarial loss, and Model II uses content and adversarial loss while adding the TV-Regularized function. The same parameters and architecture are used to train both models.

A. TRAINING DETAILS AND PARAMETERS

The following detailed implementations are using Pytorch as a learning framework, and all of the super resolution methods have been executed on the NVIDIA GTX 1080 Ti-6 GPU. The images are preprocessed to 384×384 resolution images around LP as the ground truth images used as a region of interest (ROI). Then, before being fed into the network, the images are downsampled by a factor of 4 to create 72×72 low-resolution pixel images. The scaling factor 4 shows the most detail in the image and achieves the best image quality [52], [53].

The MSE loss is determined based on the reconstructed images with intensities ranging from [-1] to [1] and the pre-trained model VGG19 is used for feature extraction. The network is trained for 1000 epochs with a training rate of 0.1 per 500 epochs and a training rate of 10^{-4} . The batch size being used in the network is 4. Adam optimizer, using beta parameter equal to $\beta = 0.9$. The batch normalization layer is disabled during the evaluation. As a result, the reconstructed images are heavily reliant on the input images. The transformed images will normalize a tensor image for each

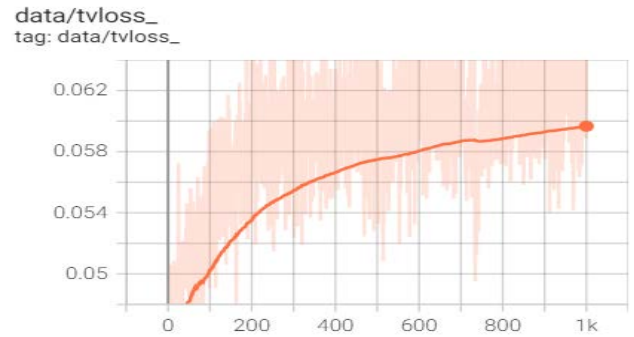


FIGURE 5. Total variation loss during the training process - Model II.

channel of the input with a mean of [0.485, 0.456, 0.406] and standard deviation of [0.229, 0.224, 0.225].

In each iteration of the training process, initially a pair of I^{LR} and I^{HR} are fed into the generator network to determine the I_{MSE}^{SR} through forward propagation. The discriminator network will subsequently take the output from generator network I^{SR} with label = 0 and I^{HR} with label = 1 and use the nominated one as an input to calculate l_D . After this, using backpropagation, the discriminator network is optimized with l_D . Finally, l_G is computed using Equation 4, which is used to optimise the generator network. To accomplish this, the generator network and discriminator network are alternately trained. The generator network can not only carry the super-resolved plate approximate the groundtruth by I_{MSE}^{SR} but also utilize the discriminator network with l_D to prevent the generator network from generating malformed characters in the plate.

This technique is evaluated by using YOLO to recognize letters and numbers. The YOLOv5 that is used on the recognition stage is open source [54] using the Pytorch framework and the models are pre-trained on COCO. Images of license plates are trained via a fine adjustment of their network parameters.

B. DATASET AND DATA PREPROCESSING

The dataset adopted is used for both the training and evaluation datasets. Experiments are done on two datasets. Firstly, we divided the car plate dataset [55] into two parts: training images (1060) and testing images (102). The testing part of the car plate dataset is used to validate the quality of output images in terms of PSNR and SSIM metrics. These results are compared to the mentioned algorithms in section [IV-C] and other published methods [42]. After that, we tested the proposed models on the different datasets, ‘‘Application-Oriented License Plate (AOLP) [6],’’ and evaluated the recognition model (yolov5). The below points illustrate the details of both datasets.

- 1) **Car Plate Dataset:** The training dataset includes 1162 HR images, divided into training and validation sets. The Car Plate Dataset 856 samples for a single object (LP), 172 samples for multiple LP in a single

image, and the remaining samples have a wider range of variation conditions, like partially blocked or spatial cases.

2) Application-Oriented License Plate (AOLP) Dataset:

This dataset contains 2,049 images of LPs which are gathered from unrestricted surveillance scenes. The AOLP dataset consists of three subsets: 681 samples for Access Control (AC), 757 samples for Traffic Low Enforcement (LE) and 611 samples for Road Patrol (RP) depending on a variety of application parameters.

C. COMPARISON OF SUPER RESOLUTION MODELS

This paper compares the suggested model with alternative techniques based on the current work on Deep Image Prior (DIP) and SRCNN.

1) DEEP IMAGE PRIOR (DIP)

This approach encourages the notion of using untrained DNNs as prior models for images, a concept pioneered by the so-called deep image of a prior model. The Deep Image Prior is a non-trained encoder-decoder architecture with constant input whose weights are designed to minimise random initialization. Initially, DIP demonstrated that Convolutional Neural Networks (CNNs) have the innate potential to address inadequately inverse problems without pretraining; they use deep images before utilizing the inner structure of the CNN itself. In addition, the seismic data contains texture patch structures and the adjacent patches are quite similar, which causes the CNN to extract deep seismic prior. Arridge *et al.* [56] proposed that image improvement may be achieved using a DNN architecture without the need for pretraining on a large dataset. Cheng *et al.* [57] suggest the DIP approach implies that the DNN architecture can directly encode the image properties and that a training dataset may not be necessary in order to solve certain issues.

This article develops a Deep Image Prior solution based on [58] to compare with the suggested models. The figure 6 describe the main steps of image improvement using the Deep Image Prior algorithm, from random noise to optimum image representation.

2) SRCNN

Dong *et al.* [48] present a DL technique for SISR that learns an end-to-end mapping between LR and HR images. The mapping is represented by a DCNN that takes a LR picture as input and produces a HR image. They suggested a basic network architecture comprising of three convoluted layers for feature extraction, nonlinear mapping, and image reconstruction called the Super-Resolution Convolutional Neural Network (SRCNN), as shown in figure 7, which outperforms previous machine-learning-based SR methods.

D. VISUAL ANALYSIS OF THE OUTPUTS

In this section, a comparison is made between the following SR methods on the proposed dataset to demonstrate

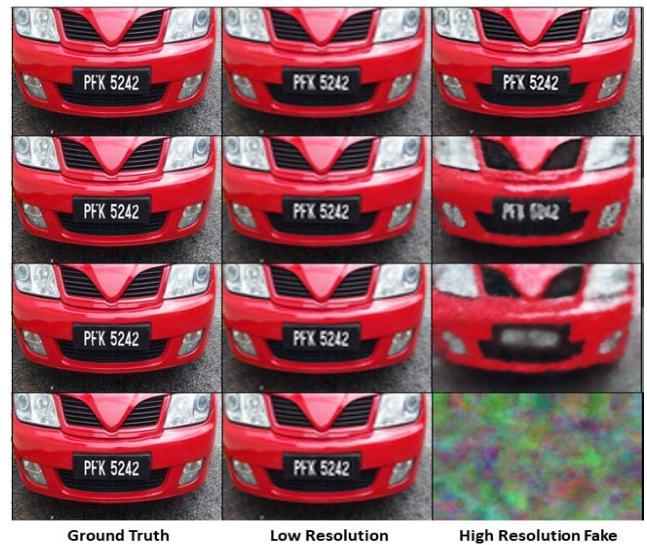


FIGURE 6. Deep image prior example.

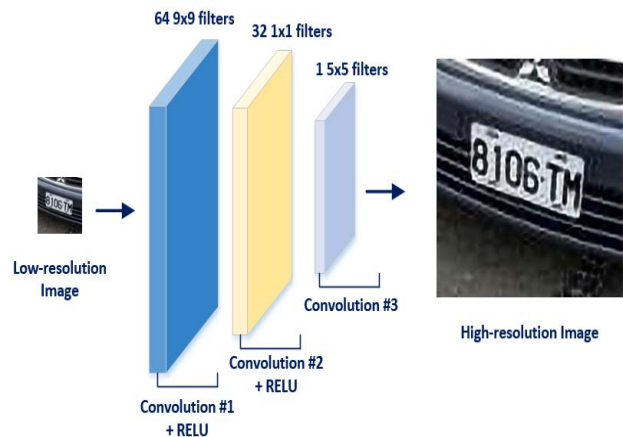


FIGURE 7. SRCNN architecture.

the effectiveness of the proposed methods. Figures 8 show the visualisation performance comparison between SRGAN architectures with other methods. Images with the smallest size on the left represent a low-resolution with a size of 72×72 , whilst after the training process in each pipeline, super-resolution images with a size of 288×288 are generated. However, in comparison to SRGAN models, the SRCNN and deep image prior models have a blurred aspect and less convincing perceptual performance. The suggested models I and II of SRGAN produce better performance in terms of visual quality, but, as shown in many figures, chromatic aberration appears in generated images from model I of SRGAN. Chromatic aberration is a form of aberration that produces undesirable colour fringes along borders within images. Therefore, the paper proposed model II of SRGAN that corrects chromatic aberration by adjusting the colour difference from a degraded image using total variation loss.



FIGURE 8. Sample of comparison with existing methods for visual aspects.

E. EVALUATION OF THE PERFORMANCE METRICS

This article compares and contrasts two well-known objective image quality metrics, the peak-signal-to-noise ratio (PSNR) as well as the structural similarity index measure (SSIM). The suggested SRGAN models are compared to the existing techniques to evaluate the performance according to the proposed method. As shown in 9, the results of the SRGAN are better than the other methods.

Table 1 summarises the quantitative performance comparisons with previous techniques. These results are analyzed over 102 images from the test dataset from the same distribution of the training datasets. As indicated in the table, the two proposed approaches for SRGAN models obtain outstanding performance compared to PSNR and SSIM quantitative findings. Furthermore, the suggested techniques produce better results in terms of perceptual quality. The execution time refers to the average time required to obtain a high-resolution image from a low-resolution one. As observed, the proposed models are executed in the least amount of time.

In addition, our proposed SRGAN techniques, Model I and Model II, are compared with the reported earlier SRGAN architecture proposed by [42]. The proposed SRGAN design

TABLE 1. Quantitative comparison with existing methods.

Model	PSNR	SSIM	Execution Time
Deep Prior	23.756	0.771	157.58 sec
SRCNN	23.494	0.767	0.11 sec
SRGAN	26.449	0.833	0.036 sec
SRGAN-TV	26.621	0.837	0.03 sec

TABLE 2. Quantitative comparison with [42] method.

Model	PSNR	SSIM
SRGAN (MSE)	24.8686	0.7213
SRGAN (MSE and perceptual loss)	23.0307	0.6760
Proposed SRGAN (Model I)	26.449	0.833
Proposed SRGAN-TV (Model II)	26.621	0.837

TABLE 3. Comparison of SRGAN method with other state-of-the-art methods on the ALOP dataset.

Model	LR Images	DIP	SRCNN	SRGAN	SRGAN-TV
Accuracy	18.8%	43%	93.5%	95.4%	93.2%

in this article is subtly different from the design used in the paper [42]. The difference in results is due to that difference in design rather than a difference in method. As reflected in the table 2, our suggested SRGAN achieves better performance in terms of PSNR and SSIM.

F. RECOGNITION EVALUATION ON ALOP DATASET

Besides the super-resolution, the main objective of the suggested strategy is to improve LP recognition performance. We evaluated the suggested model using another dataset (ALOP) for this metric. YOLO recognizes the small license plate from the images, then the SR model improves the resolution of the detected license plate, and finally, the Optical character recognition (OCR) model recognizes the license plate letters.

Therefore, the proposed method is compared with the state-of-the-art recognition techniques. Yolov5 is used as a comparative automatic license plate recognition method. The comparative findings are shown in table 3.

The table 3 displays the precise SRGAN images with super-resolution on AC, LE, and RP subsets. It obtains greater precision than previous methods, which demonstrates that super-resolution can enhance recognition.

The detection and recognition results are shown in the figures below. Firstly, figure 10 shows performance for plate and characters detection over the original images “72 × 72” on all three subsets of the ALOP dataset, with averagely 18.8%. Deep Image Prior model get a slightly higher result which achieve 43%. The poor performance value is due to the majority of the output images of this model being blurred and not clear, as shown in the previous results images of DIP. In addition, DIP is based on a single kernel on every iteration, while the models are dependent on multiple kernels during the training process. In figure 11, the SRCNN output is greater than the previous one, approximately 93% equal with SRGAN-TV, as shown in figure 12. While the SRGAN model in figure 13 achieves high accuracy of 95% compared with the other methods.



FIGURE 9. PSNR comparison of the proposed techniques.



FIGURE 10. The output of the OCR model for LR images.

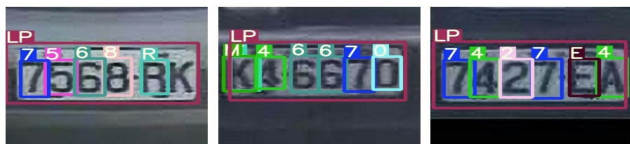


FIGURE 11. The output of the OCR model for SRCNN model.



FIGURE 12. The output of the OCR model for SRGAN model.

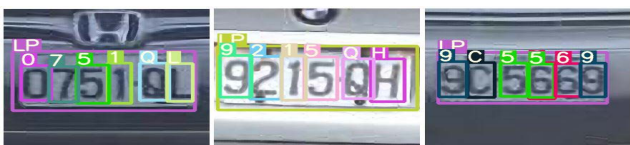


FIGURE 13. The output of the OCR model for SRGAN-TV model.

V. DISCUSSION

The novel proposed a super-resolution model with two separate criteria based on the generative adversarial network architecture. We have designed the method of adding the swish function as an activation function and SRGAN models with content and adversarial loss in the first model (Model I) and adversarial loss with TV-Regularized function in the second model (Model II) are introduced and evaluated.

The Mean Square Error (MSE) loss is applied as the model’s content loss to ensure the consistency of low-frequency information between the reconstructed image and the LR image and reduce the square error between the pixels produced and real high-resolution images. We also

added a term in the content loss equation that computed MSE between the extracted fake and actual features, and the weighting factor was set to 0.006. The purpose of reducing the distance between pixels is to ensure the correctness of the reconstructed image information more rapidly and efficiently. As a result, the PSNR has increased. On the other hand, the adversarial loss is a loss strategy that utilizes a discriminator trained to distinguish between HR and SR images to force the generator to produce an image more similar to an HR image.

Model II combines adversarial loss with Total Variation (TV) regularization loss to improve edge details and make the entire network converge more smoothly, with little need for hyperparameter adjustment to create higher quality samples, as proven by defeating chromatic aberration.

The training performance of the proposed SRGAN (Model I) is shown in figure 3. The horizontal axis indicates epochs number, while the vertical axis represents error as determined by the loss function. As shown in figure 3, the trend of generator content loss is downward and continues to decrease until the changes are tiny in the last epochs. On the other hand, the adversarial generator loss is slightly higher between 1.865 and 1.895. It indicates that if the generator model can generate high-resolution images, then the expectation is that epochs between 700 and 1000 will generate the high-resolution images. Overall the generator total loss curve is downward, which denotes that the model’s performance improves over time.

Figure 4 indicates that the discriminator did learn to distinguish between real and fake images. After 700 epochs, the system reaches the balance, and the loss does not change. Figure 5 demonstrates the relationship between the TV loss of model II and the epoch number, which rises from a small value of 0.048 until it reaches 0.065 at the last epoch. The remaining loss curves in model II follow the same trend as those in model I, but with slightly higher values, resulting in better reconstructed image quality.

The experimental findings demonstrate that our strategy achieves an average PSNR of 26.449 dB for Model I and 26.621 dB for Model II, as well as an average SSIM of 0.833 for Model I and 0.837 for Model II, which

outperforms various comparison methods in the objective evaluation index. Furthermore, character recognition accuracy improved, with 95.4% for Model I and 93.2% for Model II due to enhancement in image resolution. The value of Model II (SRGAN-TV) is lower than Model I (SRGAN) due to the mechanism of the Total Variation (TV) loss. TV loss is based on the sum of the absolute differences for neighbouring pixel-values in the images. So, if TV loss is very high, colours from neighbouring objects in the image get mingled, and the details will deform. The main objective of using the TV in this model is to solve the chromatic aberration that appears in Model I without losing image details.

Moreover, the proposed models are executed in less time, approximately 0.03. The time of OCR is not considered, which is the same for all models. The OCR model tests the recognition accuracy for all high-resolution images from the mentioned models.

VI. CONCLUSION

This article introduced techniques to recognize characters in unconstrained LP according to a deep learning technique for a Single Image Super-Resolution (SISR). Experimental findings on AOLP and Car Plate datasets demonstrate the effectiveness of the proposed method, without any scene-specific modification, surpasses current LP recognition algorithms in accuracy and generates a visual improvement in proposed SR outcomes that are better recognition from the original data. Furthermore, including the YOLO detector with the SR network, which is based on GAN, achieves better performance in terms of perceptual quality than using only the detector model (YOLO). We assess the effectiveness of our method by PSNR, SSIM, and using letter recognition with YOLOv5 for reconstructed images from low-resolution images (72×72 size).

REFERENCES

- [1] S. Du, M. Ibrahim, M. Shehata, and W. Badawy, "Automatic license plate recognition (ALPR): A state-of-the-art review," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 2, pp. 311–325, Feb. 2012.
- [2] C. N. E. Anagnostopoulos, I. E. Anagnostopoulos, V. Loumos, and E. Kayafas, "A license plate-recognition algorithm for intelligent transportation system applications," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 3, pp. 377–392, Sep. 2006.
- [3] W. Zhou, H. Li, Y. Lu, and Q. Tian, "Principal visual word discovery for automatic license plate detection," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 4269–4279, Sep. 2012.
- [4] H. Zhang, W. Jia, X. He, and Q. Wu, "Learning-based license plate detection using global and local features," in *Proc. 18th Int. Conf. Pattern Recognit. (ICPR)*, vol. 2, Aug. 2006, pp. 1102–1105.
- [5] S.-Z. Wang and H.-J. Lee, "Detection and recognition of license plate characters with different appearances," in *Proc. IEEE Intell. Transp. Syst.*, vol. 2, Oct. 2003, pp. 979–984.
- [6] G.-S. Hsu, J.-C. Chen, and Y.-Z. Chung, "Application-oriented license plate recognition," *IEEE Trans. Veh. Technol.*, vol. 62, no. 2, pp. 552–561, Feb. 2013.
- [7] A. D. Sappa and F. Dornaika, "An edge-based approach to motion detection," in *Proc. Int. Conf. Comput. Sci.* Springer, 2006, pp. 563–570.
- [8] Y. L. Yuan, W. B. Zou, Y. Zhao, X. Wang, X. F. Hu, and N. Komodakis, "A robust and efficient approach to license plate detection," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1102–1114, Mar. 2016.
- [9] K. Deb, V. V. Gubarev, and K.-H. Jo, "Vehicle license plate detection algorithm based on color space and geometrical properties," in *Proc. Int. Conf. on Intell. Comput.* Springer, 2009, pp. 555–564.
- [10] Z. X. Chen, C. Y. Liu, F. L. Chang, and G. Y. Wang, "Automatic license-plate location and recognition based on feature saliency," *IEEE Trans. Veh. Technol.*, vol. 58, no. 7, pp. 3781–3785, Sep. 2009.
- [11] S. Arivazhagan and L. Ganesan, "Texture classification using wavelet transform," *Pattern Recognit. Lett.*, vol. 24, nos. 9–10, pp. 1513–1521, 2003.
- [12] M. Heikkilä and M. Pietikäinen, "A texture-based method for modeling the background and detecting moving objects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 657–662, Apr. 2006.
- [13] G. Saadouli, M. I. Elburdani, R. M. Al-Qatoui, S. Kunhoth, and S. Al-Maadeed, "Automatic and secure electronic gate system using fusion of license plate, car make recognition and face detection," in *Proc. IEEE Int. Conf. Inform., IoT, Enabling Technol. (ICIoT)*, Feb. 2020, pp. 79–84.
- [14] K. K. Kim, K. I. Kim, J. Kim, and H. J. Kim, "Learning-based approach for license plate recognition," in *Proc. Neural Netw. Signal Process. IEEE Signal Process. Soc. Workshop*, vol. 2, Dec. 2000, pp. 614–623.
- [15] C. A. Rahman, W. Badawy, and A. Radmanesh, "A real time vehicle's license plate recognition system," in *Proc. IEEE Conf. Adv. Video Signal Based Surveill.*, Jul. 2003, pp. 163–166.
- [16] J. Czajkowska, M. Rudzki, and Z. Czajkowski, "A new fuzzy support vectors machine for biomedical data classification," in *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2008, pp. 4676–4679.
- [17] S. L. Phung, A. Bouzerdoum, D. Chai, and A. Watson, "Naive Bayes face-nonface classifier: A study of preprocessing and feature extraction techniques," in *Proc. Int. Conf. Image Process. (ICIP)*, vol. 2, Oct. 2004, pp. 1385–1388.
- [18] Y.-P. Huang, C.-H. Chen, Y.-T. Chang, and F. E. Sandnes, "An intelligent strategy for checking the annual inspection status of motorcycles based on license plate recognition," *Expert Syst. Appl.*, vol. 36, no. 5, pp. 9260–9267, Jul. 2009.
- [19] M. Yu and Y. Deak Kim, "An approach to Korean license plate recognition based on vertical edge matching," in *Proc. SMC Conf. IEEE Int. Conf. Syst., Man Cybern. Cybern. Evolving Syst., Hum., Organizations, Complex Interact.*, Oct. 2000, pp. 2975–2980.
- [20] H. Ibrahim, M. A. Elattar, and W. Badawy, "On the application of real-time deep neural network for automatic license plate reading from sequence of images targeting edge artificial intelligence architectures," in *Enabling Machine Learning Applications in Data Science*. Springer, 2021, pp. 299–311.
- [21] S. M. Silva and C. R. Jung, "Real-time license plate detection and recognition using deep convolutional neural networks," *J. Vis. Commun. Image Represent.*, vol. 71, Aug. 2020, Art. no. 102773.
- [22] X. Liu, T. Kawanishi, X. Wu, and K. Kashino, "Scene text recognition with high performance CNN classifier and efficient word inference," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 1322–1326.
- [23] K. T. Islam, R. G. Raj, S. M. S. Islam, S. Wijewickrema, M. S. Hossain, T. Razmovski, and S. O'Leary, "A vision-based machine learning method for barrier access control using vehicle license plate authentication," *Sensors*, vol. 20, no. 12, p. 3578, Jun. 2020.
- [24] W. Puarungroj and N. Boonsirisumpun, "Thai license plate recognition based on deep learning," *Proc. Comput. Sci.*, vol. 135, pp. 214–221, Jan. 2018.
- [25] W. Weihong and T. Jiaoyang, "Research on license plate recognition algorithms based on deep learning in complex environment," *IEEE Access*, vol. 8, pp. 91661–91675, 2020.
- [26] A. N. A. Rahim, S. N. Yaakob, R. Ngadiran, and M. W. Nasruddin, "An analysis of interpolation methods for super resolution images," in *Proc. IEEE Student Conf. Res. Develop. (SCORED)*, Dec. 2015, pp. 72–77.
- [27] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-29, no. 6, pp. 1153–1160, Dec. 1981.
- [28] S. Sarmadi and Z. Shamsa, "A new approach in single image super resolution," in *Proc. 6th Int. Conf. Comput. Knowl. Eng. (ICCKE)*, Oct. 2016, pp. 78–81.
- [29] F. Ling, Y. Du, X. Li, W. Li, F. Xiao, and Y. Zhang, "Interpolation-based super-resolution land cover mapping," *Remote Sens. Lett.*, vol. 4, pp. 629–638, Jul. 2013.
- [30] L. Zhang, W. Zhang, G. Lu, P. Yang, and Z. Rao, "Feature-level interpolation-based GAN for image super-resolution," *Pers. Ubiquitous Comput.*, pp. 1–16, Jan. 2021.

- [31] S. Dai, M. Han, W. Xu, Y. Wu, and Y. Gong, "Soft edge smoothness prior for alpha channel super resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
- [32] J. Sun, Z. Xu, and H.-Y. Shum, "Image super-resolution using gradient profile prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.
- [33] Y.-W. Tai, S. Liu, M. S. Brown, and S. Lin, "Super resolution using edge prior and single image detail synthesis," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2400–2407.
- [34] P. Sen and S. Darabi, "Compressive image super-resolution," in *Proc. Conf. Rec. 43rd Asilomar Conf. Signals, Syst. Comput.*, 2009, pp. 1235–1242.
- [35] W. Dong, L. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 1838–1857, Jul. 2011.
- [36] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang, "Coupled dictionary training for image super-resolution," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3467–3478, Aug. 2011.
- [37] C. Ledig, Lucas Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 4681–4690.
- [38] Y. Wang, L. Wang, H. Wang, and P. Li, "End-to-end image super-resolution via deep and shallow convolutional networks," *IEEE Access*, vol. 7, pp. 31959–31970, 2019.
- [39] D. Mahapatra, B. Bozorgtabar, and R. Garnavi, "Image super-resolution using progressive generative adversarial networks for medical image analysis," *Comput. Med. Imag. Graph.*, vol. 71, pp. 30–39, Jan. 2019.
- [40] 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 *Proc. Eur. Conf. Comput. Vis. (ECCV Workshops)*, Sep. 2018, pp. 1–16.
- [41] Y. Lee, J. Jun, Y. Hong, and M. Jeon, "Practical license plate recognition in unconstrained surveillance systems with adversarial super-resolution," 2019, *arXiv:1910.04324*.
- [42] T. K. Lai, A. F. Abbas, A. M. Abdu, U. U. Sheikh, M. Mokji, and K. Khalil, "Super resolution of car plate images using generative adversarial networks," in *Proc. IEEE 15th Int. Colloq. Signal Process. Appl. (CSPA)*, Mar. 2019, pp. 80–85.
- [43] P. Shamsolmoali, M. Zareapoor, E. Granger, H. Zhou, R. Wang, M. E. Celebi, and J. Yang, "Image synthesis with adversarial networks: A comprehensive survey and case studies," *Inf. Fusion*, vol. 72, pp. 126–146, Aug. 2021.
- [44] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014, *arXiv:1406.2661*.
- [45] L. J. Ratliff, S. A. Burden, and S. S. Sastry, "Characterization and computation of local nash equilibria in continuous games," in *Proc. 51st Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2013, pp. 917–924.
- [46] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
- [47] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for activation functions," 2017, *arXiv:1710.05941*.
- [48] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2015.
- [49] S. Minaee, M. Minaei, and A. Abdolrashidi, "Palm-GAN: Generating realistic palmprint images using total-variation regularized GAN," 2020, *arXiv:2003.10834*.
- [50] M. F. Fahmy, G. M. A. Raheem, U. S. Mohammed, and O. F. Fahmy, "A new total variation based image denoising and deblurring technique," in *Proc. Eurocon*, Jul. 2013, pp. 1669–1675.
- [51] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
- [52] M. Zareapoor, H. Zhou, and J. Yang, "Perceptual image quality using dual generative adversarial network," *Neural Comput. Appl.*, vol. 32, no. 18, pp. 14521–14531, Sep. 2020.
- [53] D. Mahapatra and B. Bozorgtabar, "Progressive generative adversarial networks for medical image super resolution," 2019, *arXiv:1902.02144*.
- [54] *YOLOv5-Github*. [Online]. Available: <https://github.com/ultralytics/yolov5>
- [55] E. Liang. (2017). *Car Plate Dataset*. Accessed: Apr. 6, 2021. [Online]. Available: <https://bit.ly/3rzD110>
- [56] S. Arridge, P. Maass, O. Öktem, and C.-B. Schönlieb, "Solving inverse problems using data-driven models," *Acta Numerica*, vol. 28, pp. 1–174, May 2019.
- [57] Z. Cheng, M. Gadelha, S. Maji, and D. Sheldon, "A Bayesian perspective on the deep image prior," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5443–5451.
- [58] *Deepimageprior-Github*. [Online]. Available: <https://github.com/DmitryUlyanov/deep-image-prior>



IBRAHIM H. EL-SHAL received the bachelor's degree (Hons.) in communications and electronics from the Faculty of Electronic Engineering, Menoufia University, in 2015. He is currently pursuing the master's degree with the Informatics Science Center, Nile University. In 2016, he got a Nine-Month Diploma Scholarship in wireless communication track with the Information Technology Institute (ITI), Ministry of Communications and Information Technology (MCIT). He has full-time employment as a Senior Supervisor Machine Learning Operations Engineer at Orange Egypt. He had four publications during his master's journey, which started in 2017. His research interests include computer vision, image processing, telecommunications, and machine/deep learning.



OMAR M. FAHMY (Senior Member, IEEE) received the bachelor's, master's, and Ph.D. degrees in communication engineering from Assiut University, in 2006, 2008, and 2012, respectively. In 2013, he joined the Signal and Image Processing Group, University of Vigo, as a Post-doctoral Fellow, where he conducted research in image forensics and image forgery detection. Since 2019, he has been working as a Senior Researcher at the Electrical Engineering Department, South Ural State University, Russia. He is currently an Associate Professor at the Electrical Engineering Department, School of Engineering, Badr University in Cairo. His research work results in 35 publications in international peer-reviewed journals and IEEE conferences in the field of computer vision, machine learning, and image processing. His research interests include image processing, image forensics, computer vision, machine learning, and their related areas.



MUSTAFA A. ELATTAR was born in Cairo, Egypt, in 1986. He received the bachelor's degree in systems and biomedical engineering from Cairo University, Cairo, in 2008, the master's degree in communication and information technology, in 2010, after conducting image analysis research for cardiac imaging, and the Ph.D. degree in biomedical engineering and physics, medicine from the Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands, in 2016, after developing preoperative planning framework for transcatheter aortic valve implantation. He joined the Medical Imaging and Image Processing Research Group, Nile University, Giza, Egypt, as a Research Assistant. In 2015, he joined The Netherlands Cancer Institute (NKI) as a Postdoctoral Fellow, where he conducted research for image-guided radiotherapy. In August 2017, he joined Nile University as an Assistant Professor at the Information Technology and Computer Science School. He is also leading the Medical Imaging and Image Processing Research Group focusing on incorporating deep neural networks in 2D and 3D medical image analysis contexts. On the professional side, he worked at the Research and Development Division, Diagnosoft Inc., 3mensio B.V., PieMedical N.V., and Myocardial Solutions Inc. In August 2018, he has founded and currently leading Intixel Company S.A.E., which focuses on serving medical imaging solution firms with its turnkey artificial intelligence solutions that achieve their business needs. He has more than 25 journal articles and conference publications.