An Enhanced Variational AutoEncoder Approach for the Purpose of Deblurring Bangla License Plate Images

Md. Siddiqure Rahman Tusher, Nakiba Nuren Rahman, Shabnaz Chowdhury, Anika Tabassum,
Md. Akhtaruzzaman Adnan, Rashik Rahman*, Shah Murtaza Rashid Al Masud*
Department of Computer Science and Engineering
University of Asia Pacific, Dhaka, Bangladesh

***Abstract*—Automated License Plate Detection and Recognition (ALPDR) is a well-studied area of computer vision and a crucial activity in a variety of applications, including surveillance, law enforcement, and traffic management. Such a system plays a crucial role in the investigation of vehicle-related offensive activities. When an input image or video frame travels through an ALPDR system for license plate detection, the detected license plate is frequently blurry due to the fast motion of the vehicle or low-resolution input. Images of license plates that are blurred or distorted can reduce the accuracy of ALPDR systems. In this paper, a novel Variational AutoEncoder(VAE) architecture is proposed for deblurring license plates. In addition, a dataset of obscured license plate images and corresponding ground truth images is proposed and used to train the novel VAE model. This dataset comprises 3788 image pairs, in which the train, test, and validation set contains 2841, 568, and 379 pairs of images respectively. Upon completion of the training process, the model undergoes an evaluation procedure utilizing the validation set, where it achieved an SSIM value of 0.934 and a PSNR value of 32.41. In order to assess the efficacy of our proposed VAE model, a comparison with contemporary deblurring techniques is presented in the results section. In terms of both quantitative metrics and the visual quality of the deblurred images, the experimental results indicate that our proposed method outperforms the other state-of-the-art deblurring methods. Therefore, it enhances the precision and dependability of an ALPDR system.**

***Keywords*—Image deblur; bangla license plate deblur; Variational AutoEncoder (VAE); computer vision**

I. Introduction

With the increase in the number of vehicles on the road, violations of traffic laws such as racing through red lights, leaving the scene of an accident, and kidnapping escalated. As a result, ALPDR systems have been extensively developed and applied to a variety of intelligent traffic systems [1]–[4]. Unfortunately, despite the fact that drive recorders and surveillance cameras perform significantly better than in the past, license plates of vehicles are frequently blurred due to fast-moving vehicles, camera shakes during the exposure period, and other factors. Multiple variables contribute to the blurring and distortion of license plate images. The initial variable is the local environment. For instance, the impacts of intense illumination, precipitation, and weather may increase the likelihood of blurring. The second variable consists of the vehicle's motions. For instance, when vehicles run red lights, they are frequently traveling at a very high rate of speed; consequently, the pictures taken tend to be blurry. The surveillance system serves as the final variable. Due to the fact that surveillance cameras are frequently positioned at higher elevations, far from the car, the captured image has a reduced resolution, leading to poor image quality. Blurred license plate images can significantly reduce the accuracy of ALPDR systems. Therefore, deblurring of the license plate images is a crucial step towards achieving reliable ALPDR systems.

Starting from deblurring images using methods such as dihedral group [5], image deblurring techniques have significantly advanced [6] in recent years. On the basis, of high-frequency residual image learning, the authors of [7] proposed a two-phase deblurring algorithm for restoring blurred images of dynamic scenes. The method proposed in [8] defines a new regularization term that incorporates both intensity and gradient assumptions and provides an efficient and convergent solution for deblurring license plate images. Convolutional Neural Networks (CNNs) have been extensively utilized [9]–[11] alongside Generative Adversarial Networks (GANs) [8], [12] to generate sharp license plate images. However, these approaches demand a large amount of training data, as with a small dataset the model may not converge and can lead to a less generalized model. To our knowledge, there are a handful of works related to Bangla license plate deblurring.

The purpose of this study is to develop and establish a state-of-the-art Bangla license plate deblurring system that can work in real-time. In order to accomplish this we propose a novel VAE architecture with a custom loss function for Bengali license plate image deblurring. The proposed VAE network is a generative model capable of learning the underlying distribution of training data and producing new samples based on the learned distribution. Thus, it can perform well even if it is trained on a small dataset. However, there's a lack of publicly available Bangla license plate dataset for deblurring. Thus, we created a new balanced and generalized dataset consisting of 3788 pairs of images that were used to train, test and validate the model. 75% of the data belong to the train set, whereas 15% and 10% data belong to the test and validation sets, respectively. The proposed VAE model achieved an SSIM score of 0.934 and a PSNR score of 32.41. To assess the efficacy, we recreated state-of-the-art deblurring models [13], [14] and trained them on our dataset. The evaluation demonstrates that our method outperforms state-of-the-art deblurring techniques in terms of both quantitative metrics and image quality. In addition, we demonstrate the

efficacy of our procedure using actual blurred license plate images in this paper.

The most significant contributions of this study are as follows:

- Established a novel VAE network for the deblurring of Bangla license plate images.
- Created a new dataset for Bangla license plate deblurring.
- Provided quantitative and visual performance analysis of the proposed model with other available image deblurring models.

The subsequent sections of the paper are structured in the following manner: Section II encompasses the literature review. The following Section III, provides a comprehensive summary of the dataset. Section IV presents a detailed description of the proposed methodology. Section V presents the results and comparisons of this research. The conclusive remarks can be found in Section VI. The residual content comprises references.

II. Literature Review

The field of deblurring license plate images is a highly intriguing area of research. Despite the limited amount of research on the deblurring of Bangla license plate images, a small number of studies have successfully developed methods that can be applied in real-world scenarios. Nevertheless, this area of research remains largely unexplored. The following segment delineates the progression of scholarly inquiry pertaining to the restoration of clarity in license plate images.

Fang *et al.* [8] introduced a deblurring methodology that integrates gradient priors and intensity through a novel regularization technique. In addition, the authors determined that the binarization threshold of the image is a significant factor in distinguishing between blurred and clear images. CNNs have been employed in the past for the purpose of image denoising [9]–[11], as well as for super-resolution [15], [16]. The study conducted by the author of [9] aimed to address the issue of image blurring in traffic surveillance. To achieve this objective, the author developed a customized CNN model. This model was designed to take a blurry image as input and generate a clear image as output, with the license plate content being easily discernible. While the methods proposed by [8], [9] were deemed partially satisfactory, they were found to be lacking in terms of accuracy measures and evaluations.

Qingbo *et al.* [17] introduced a new approach for estimating blur kernels by integrating linear uniform convolution and the angle of the image. A scheme was proposed utilizing sparse representation to detect the blur kernel resulting from rapid vehicular movement. The authors analyze the coefficients of the sparse representation of the reconstructed image to ascertain the orientation of the kernel. They estimate the length of the motion kernel using the Radon transform in the Fourier domain. However, the dataset that the authors used only had 240 images, which limited its ability to include all aspects of real-world scenarios.

Orest *et al.* [6] proposed an innovative approach for image deblurring that employs a Generative Adversarial Network (GAN) architecture. The researchers utilized the GoPro and Kohler datasets in order to train their model. Despite the fact that the dataset consists of arbitrary blurred images, the proposed approach demonstrates the ability to deblur, blurry vehicle images. The model proposed by the author attained a Peak signal-to-noise ratio (PSNR) score of 26.10 and a structural similarity index measure (SSIM) score of 0.816. The proposed model's inference time is 0.85 seconds. Subsequently, the author introduced an enhanced approach in [14]. The authors put forth a sophisticated GAN framework, which achieved a PSNR score of 29.55 and an SSIM score of 0.93. They concluded that the novel architecture outperformed their previous model. Furthermore, the inference time of the model under consideration is a mere 0.35 seconds. In any case, the duration required for inference in both of the aforementioned studies can be considered unsuitable for the practical application of ALDPR systems.

The researchers of [18] proposed a GAN architecture to address the task of image deblurring. The authors additionally employed the GoPro and Kohler dataset for the purposes of training and assessing their model. A PSNR score of 29.32 and an SSIM score of 0.93 show that the network the authors propose is significantly effective in the task of deblurring vehicle images. However, it is worth noting that the inference time associated with this network is relatively high, at 0.64s. A sparse regularization model for vehicle image deblurring was proposed by the author of [19], utilizing the statistical distribution characteristics of the image. The author presents a concise analysis of the statistical distributions of vehicle images and concludes that the gradient histogram of the ground truth follows the Hyper-Laplacian distribution. The model exhibited exceptional performance in comparison to contemporary techniques for deblurring vehicle images. The PSNR and SSIM values are 28.2 and 0.99 respectively. However, the impact of the model parameter p on the deblurring quality is not extensively discussed in the paper. Hence, additional research may be necessary to explore the optimal value of p for diverse license plate images.

Hiroki *et al.* [12] presented a technique for achieving high-quality image deblurring. The method employs a Discrete Cosine Transform (DCT)-based loss function to maintain texture and mitigate ringing artifacts in the resulting image. The authors' proposed model exhibits reduced computational complexity in comparison to alternative methods that employ multi-scale architecture. The proposed method involves a comparison of the frequency domain of the deblurred image with the ground truth image via DCT. This approach effectively mitigates block noise and ringing artifacts, while simultaneously preserving the deblurring performance. The empirical findings indicate that DeblurDCTGAN exhibits better results compared to conventional techniques. Their proposed model achieved PSNR and SSIM values of 28.84 and 0.93, respectively. Moreover, DeblurDCTGAN exhibits a comparatively quicker runtime of 0.28s per pair in contrast to alternative techniques, rendering it a more efficient alternative for the purpose of image deblurring. The Enhanced Super Resolution Generative Adversarial Network (ESRGAN) was employed by the authors of [20] to enhance the quality of license plate images that were originally of low quality. ESRGAN has undergone training using a dataset comprising license plate images, with the aim of acquiring the ability to perform image deblurring and

upscaling. Subsequently, the generated high-resolution images are employed to achieve precise recognition of license plates. However, The size of the dataset utilized in the training of ESRGAN is comparatively limited, potentially impeding the model's ability to generalize to license plate images beyond those included in the dataset. Additionally, the optical character recognition accuracy for the reconstructed images is lower than that of the ground truth images, indicating that there is still room for improvement for the generator model. Finally, the paper did not provide a detailed description of the results obtained using the ESRGAN.

Despite notable progress made in the domain of image deblurring in recent years, a number of constraints remain prevalent in previous research endeavors, as indicated earlier. A significant constraint lies in the challenge of precisely simulating the blurring mechanism, which can exhibit substantial variability contingent upon factors such as vehicle motion, lens, and scene attributes. Furthermore, the majority of previous studies concerning license plate deblurring or image deblurring have concentrated on restoring clarity to comparatively uncomplicated scenes or images, and may not possess the capability to address more intricate or demanding scenarios. One additional constraint pertains to the computational cost linked with deblurring techniques, which may pose a significant hindrance for real-time implementations. In general, although previous studies on the deblurring of license plates or images have made noteworthy advancements in enhancing the quality of deblurred images, there remain various obstacles that require attention to enhance the robustness, efficiency, and efficacy of these approaches.

III. Dataset Description

This section discusses the dataset's curation and preprocessing. There is a scarcity of publicly accessible datasets for deblurring purposes, as very little research has been conducted on the deblurring of Bangla license plates. Therefore, a new dataset is proposed for this investigation. Initially, a license plate-related dataset was proposed in [4]. As shown in Fig. 2, their proposed dataset was for license plate detection and recognition. Therefore, we accumulated and expanded this dataset. Afterward, the license plate was extracted from these images, and a dataset containing 3788 image pairs of license plates was created, samples of which are depicted in Fig. 3. Each image of a clear license plate I_i is passed through a blur function, resulting in a blurred license plate B_i. The blur function is denoted by the Eq 1.

$$B_i = I_i \circledast F + N + R + S \tag{1}$$

A dataset that is unbalanced and insufficient can cause the VAE network to produce low-quality images, which can hinder its ability to generalize to new data. A balanced dataset is crucial for training VAE since it helps to ensure that the network is trained properly and is capable of generalizing adequately to new data. The distribution of the dataset among the train, test, and validation set is shown in Fig. 1, where the train, test, and validation sets, respectively, include 2841, 568, and 379 pairs of images. The validation set is used to assess the model post-training but is concealed from the model during training.

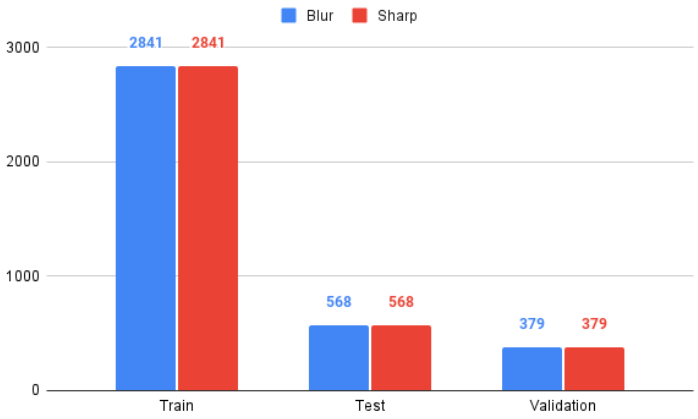

Fig. 1. Data distribution.

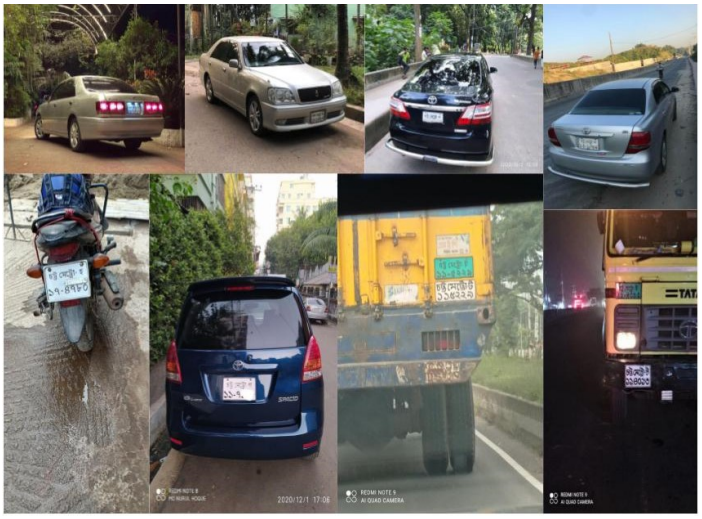

Fig. 2. Sample images from the dataset of [4].

IV. Proposed Method

Traditional approaches for deblurring license plate images [8], [12], [20], function well. However, issues with compute power, inference time, loss of details in the generated image, and the likelihood that noise may recur in the resulting image continue to exist. In this paper, a novel approach is proposed to address the issue of license plate image blurring. The proposed method utilizes the capabilities of a Variational AutoEncoder (VAE) to achieve this objective. The proposed approach involves developing a novel VAE network that has the capability to acquire a low-dimensional representation of the underlying image data. This representation can be leveraged to effectively eliminate blur and distortion from license plate images, while simultaneously preserving the texture and details of the input image. The objective of the model is to restore a clear image of a license plate denoted as I_i, given a blurred image B_i as the input. It is noteworthy that no information pertaining to the blur kernel is unseen by the model in the training, testing, and validating phases.

Fig. 3. Sample images of the proposed dataset.

A. VAE Architecture

The proposed VAE network incorporates a prior distribution on the latent vector Z, which follows a Multi-Variate Gaussian distribution. In contrast to mapping the image onto a point space, the encoder of VAE maps the image onto a normal distribution. The encoder employs a Convolutional Neural Network (CNN) as a dimensionality reduction model to effectively map intricate training data onto the latent space Z while maintaining appropriate statistically significant attributes.

Fig. 4 depicts the comprehensive structure of the proposed VAE network. The encoder receives a visually blurred input image denoted as X, from which it generates two latent vectors, ZU (mean) and ZS (variance), that serve as the distribution parameters learned during the training process. The ultimate latent vector Z is obtained through the utilization of a Multi-Variate Gaussian distribution, which samples from ZU and ZS. Subsequently, the vector Z is transmitted to the decoder module, where it is subjected to the inverse transformation of the encoding process. This results in the generation of the predicted image Y, which represents a restored and clear version of the original blurry input image. The proposed VAE network encompasses a specific region that is centered on the mean and has a magnitude equivalent to the standard deviation. This provides the decoder with a greater amount of data, enabling it to generate an image that closely resembles the ground truth image.

B. Encoder-Decoder

This section outlines the definition of the Encoder and Decoder components of the network. The encoder-decoder model is presented in Fig. 6 for a comprehensive understanding. The input image's dimensions for the encoder are (None, 128, 128, 3). The encoder architecture comprises three convolutional blocks, each of which incorporates Conv2D alongside the ReLu activation function and Batchnormalization. The final layer employs a Flatten operation to transform the feature matrix, which has dimensions of (16, 16, 256), into a vector with a size of 65536. Subsequently, the Flatter layer's outcome is transmitted to two distinct dense layers with the aim of obtaining latent vectors ZU and ZS. The encoder network acquires the ability to establish a mapping between the input data and ZU and ZS, which are anticipated to conform to a normal distribution. Subsequently, the ZU and ZS vectors are transmitted to a sampling block that employs a Lambda layer. The utilization of the lambda layer proves to be advantageous in the implementation of bespoke functions that are not inherently incorporated as standard functions within TensorFlow[1]. Equation 2 presents a bespoke Lambda layer function that accepts ZU and ZS as input and generates the Z vector as output during the training stage. The utilization of a basic sampling technique may result in a bottleneck within the backpropagation process. To address this concern, a reparameterization technique[2] is employed. This technique enables the loss to propagate backward through the mean and variance nodes, which are deterministic while segregating the sampling node by introducing a non-deterministic parameter $epsilon$, which is drawn from a standard normal distribution. This property ensures that Z is deterministic.

$$Z = ZU + ZS^2 * \epsilon \tag{2}$$

The two computation graphs depicted in Fig. 5 illustrate the original sampling block in (a) and its reparameterized form in (b). The blue nodes in the diagram depict the deterministic nodes, specifically the input, and weights, whereas the red nodes represent the stochastic nodes. Throughout the training process, the input image undergoes a mapping procedure resulting in the derivation of two latent variables, namely ZU and ZS. Subsequently, a vector Z is sampled from the aforementioned variables. The stochastic sampling process employed in this operation renders Z a random node, thereby creating an obstacle due to the inability of gradients to backpropagate through the sampling layer, owing to its stochastic nature. Consequently, the parameters ZU and ZS are unable to acquire knowledge. The process of backpropagation necessitates that the nodes exhibit determinism in order to facilitate the iterative propagation of gradients. The proposed VAE network introduces a reparameterization technique to tackle the aforementioned concern, whereby the stochastic node Z is transformed into a deterministic node. The utilization of ϵ enabled the preservation of the stochasticity of the entire system while allowing the ZU and ZS vectors to function as the learnable parameters of the network.

The decoder network receives the output of the encoder (vector Z), which has a dimension of (None, 2). The initial

[1]https://www.tensorflow.org/api_docs/python/tf/keras/layers/Lambda
[2]https://gregorygundersen.com/blog/2018/04/29/reparameterization/

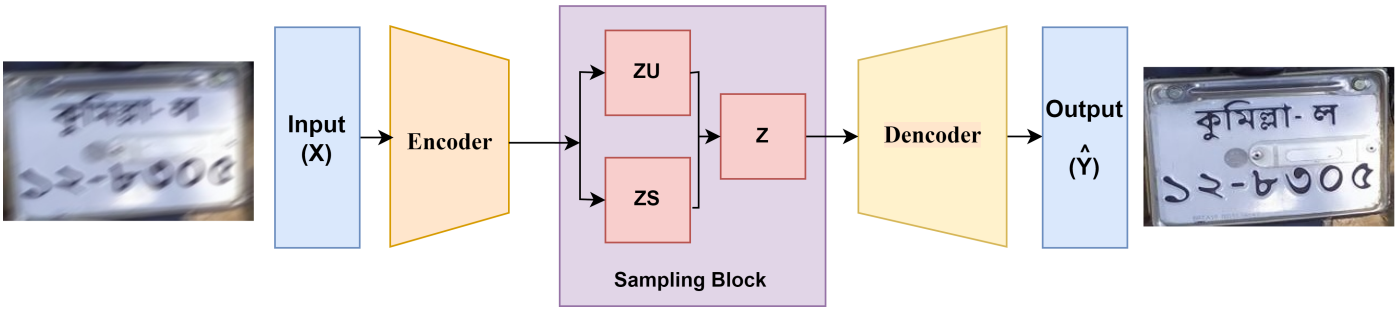

Fig. 4. Overview of the proposed VAE network.

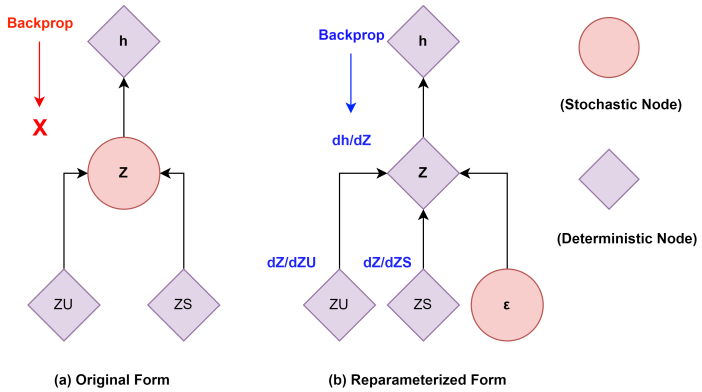

Fig. 5. Reparameterized form of custom sampling block.

dense layer is assigned a filter value of 65536, which corresponds to the product of 16, 16, and 256. Thus to make the vector into a 2D matrix a reshape layer is utilized with the shape value of (16,16,256). This value is derived from the flattened output of the preceding encoder. The architecture consists of four inverse convolutional blocks, each of which incorporates Conv2DTranspose with Relu activation function and Batchnormalization. The final convolutional block within the decoder exhibits a filter value of 3, resulting in an output image shape of (None, 128, 128, 3). Therefore, a clear image of the license plate is obtained.

C. Loss Function

The latent vector Z conforms to a Gaussian distribution with unit variance and is accountable for the minimization of the reconstruction loss denoted as L_r. The final loss function of the VAE that has been suggested is a combination of the reconstruction loss and Kullback–Leibler Divergence (KLD) loss, which is represented as a weighted sum in Equation 3. Both loss functions are optimized in this context. The reconstruction loss serves to guarantee that the resulting image is comparable to the ground truth image, while the KLD loss ensures that the latent variables are in proximity to the standard normal distribution. The KLD metric quantifies the divergence between the latent vector Z (sampled from ZU, and ZS) and the unit normal distribution γ. The parameters of the encoder and decoder are represented by σ and η in Equation 3.

$$L_{total}(\sigma, \eta, y) = L_r(\sigma, \eta, \hat{y}) + KLD[ZU, ZS, \gamma] \quad (3)$$

Equation 4 refers to the calculation of L_r, where P refers to the total number of pictures in a training batch, y is the ground truth image, and $f_\sigma(g_\eta(y)$ is the reconstructed image. The reconstruction loss is calculated by determining the difference between each pixel of the ground truth image and the corresponding pixel of the reconstructed output image. This difference is then squared and averaged across the entire batch of the data.

$$L_r(\sigma, \eta, y) = \frac{1}{P}\sum_{j=1}^{P}(y_j - f_\sigma(g_\eta(y_j))^2 \quad (4)$$

The Variational Autoencoder (VAE) is trained with the objective of minimizing the Kullback-Leibler Divergence (KLD) between the latent vectors and γ. In the event that the encoder produces a vector Z that deviates significantly from a standard normal distribution, the KLD loss function will impose a greater penalty. The KLD serves as a regularization technique to ensure adequate diversity within vector Z. During the computation of the KLD, it is customary to assign the parameter ZS to the natural logarithm of the variance. The mathematical expression for KLD is represented by Equation 5. Through the application of the logarithmic function to the ZS, the network is constrained to produce output values within the domain of natural numbers, as opposed to solely positive values. This facilitates more seamless illustrations of the underlying space.

$$KLD[ZU, ZS, \gamma] = -0.5 * \sum_{j=1}^{p} 1 + log(ZS_j^2) - ZU_j^2 - e^{log(ZS_j^2)} \quad (5)$$

V. Result and Analysis

The objective of this research is to design a Variational Autoencoder (VAE) architecture that exhibits high performance in the task of restoring clarity to the input image of license plates. The images in the dataset represent a variety of perspectives and environmental conditions. In addition to images of average to excellent quality, the dataset also contains rotated, tilted, pixelated, and occluded images. Therefore, this dataset is highly suitable for extrapolating the deblurring issue. This section provides an analysis of the results obtained from the proposed VAE model. Furthermore, a comparative analysis is presented between the proposed VAE model and prominent deblurring models such as DeblurGan [13] and DeblurGanV2 [14].

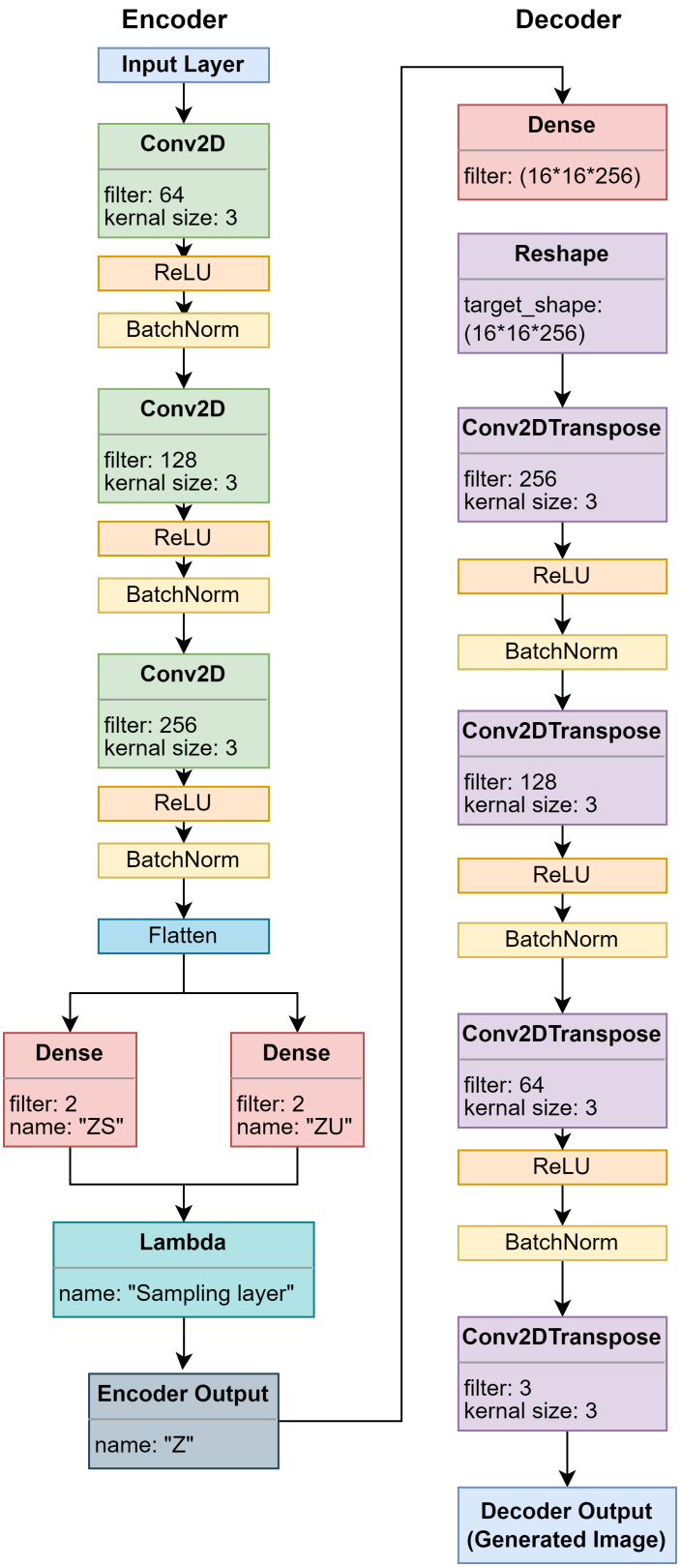

Fig. 6. Proposed encoder and decoder in the VAE network.

A. Experimental Setup

Throughout the model training and testing process, the hardware setup consisted of an Intel Core i7 10700K Central Processing Unit (CPU), a total of 32 GB of DDR4 RAM, and an Nvidia GTX 1070 GPU with 8 GB of memory. The VAE model was developed utilizing the Tensorflow 2.11 deep learning framework in conjunction with the Python 3.9 programming language.

B. Evaluation Metrics

1) PSNR: The Peak Signal-to-Noise Ratio (PSNR) is a metric employed to assess the fidelity of an image that has been restored. The process involves a comparison between the initial image and the restored image, followed by the computation of the ratio between the highest attainable value of the original signal and the value of the interference that impacts the quality of the restoration. The quality of a restoration of an image is positively correlated with the PSNR value, such that a higher PSNR value indicates superior image quality. Due to the broad dynamic range of signals, the PSNR is often expressed in decibels, a logarithmic scale. The PSNR is calculated using Equation 7. In Equations 6 and 7, the symbols p and q respectively denote the matrix values of the ground truth image and the restored image. The indices i and j represent the row and column of the matrices, while MAX_g represents the maximum value present in the original data. The primary constraint of this metric pertains to its exclusive reliance on numerical comparison, without considering any other factors of the human vision system, such as the structural similarity index (SSIM).

$$MSE = \frac{\sum_0^{i-1}\sum_0^{j-1}||p(i,j)-q(i,j)||^2}{i*j} \tag{6}$$

$$PSNR = 20\log_{10}(\frac{MAX_g}{\sqrt{MSE}}) \tag{7}$$

2) SSIM: The Structural Similarity Index Measure (SSIM) is a prominent metric employed to evaluate the degree of similarity between two images. In contrast to conventional metrics such as MSE or PSNR, which concentrate exclusively on discrepancies at the pixel level, SSIM incorporates both the structural characteristics and perceptual attributes of images. The Structural Similarity Index (SSIM) assesses three fundamental elements of image similarity, namely luminance (L), contrast (C), and structure (S). L, C, and S are calculated using Equation 8, Equation 9, and Equation 10 respectively. The method operates by evaluating the resemblance of matching localized image patches and subsequently calculating a weighted mean over the entire image. The SSIM value obtained falls within the range of 0 to 1, where a value of 1 denotes a high degree of similarity. The SSIM metric shown in Equation 11 is designed to overcome certain drawbacks of conventional metrics, including susceptibility to noise and variations in image resolution. The incorporation of human visual perception and structural information renders it more perceptually significant. The SSIM metric is highly advantageous in applications that involve image quality evaluation, image recovery, and image compression, as it is imperative to maintain perceptual quality in these tasks.

Fig. 7. Sample output of the proposed VAE network; where a) input, b) ground truth, and c) generated image by the proposed model.

$$L(n,m) = \frac{(2\mu_n\mu_m + v1)}{(\mu_n^2 + \mu_m^2 + v1)} \tag{8}$$

$$C(n,m) = \frac{2\sigma_n\sigma_m + v2}{\sigma_n^2 + \sigma_m^2 + v2} \tag{9}$$

$$S(n,m) = \frac{\sigma_{nm} + v3}{\sigma_n\sigma_m + v3} \tag{10}$$

$$SSIM(n,m) = L(n,m)^{\delta} * C(n,m)^{\lambda} * S(n,m)^{\omega} \tag{11}$$

The variables n and m represent the deblurred image and its corresponding ground truth image, respectively. The variables μ_n and μ_m represent the arithmetic means of the values of n and m, respectively. The variables σ_n^2 and σ_m^2 denote the variances of n and m, respectively, while σ_{nm} represents the covariance between n and m. The variables σ_n and σ_m represent the standard deviation of n and m respectively. The stabilization of the division with a weak denominator is achieved through the utilization of three parameters, namely $v1 = (U1 * L)^2$, $v2 = (U2 * L)^2$, and $v3 = v2/2$. Here, L denotes the dynamic range of the pixel values, while the values of $U1$ and $U2$ are assigned as 0.01 and 0.03, respectively.

C. Analysis of the VAE's Performance

Here Fig. 7 illustrates a) the input blurred image of the license plate, b) ground truth in other words sharp image, and c) the output image of the proposed VAE model. From Fig. 7, it is evident that the proposed model is better able to retain and restore the texture in the deblurred image. It shows superior results in terms of visual representation. The images depicted in Fig. 7(a) exhibit a significant degree of blurriness, rendering the license plate information indiscernible. However, the model performs well in restoring the image and the informations are clearly visible. Along with restoring private vehicle license plate images (with white license plate background), the model is also capable of restoring public vehicle license plate images (with green license plate background). The 6^{th} image of Fig. 7(b) has a noise and ringing effect. However, the generated image is very clear and there is no existence of noise in it. Thus, in accordance with image deblurring the proposed model can also reduce noise in the input image. After training the model for 100 epochs we achieved such results due to a generalized dataset and careful tuning of the hyperparameters and loss function tuning of the proposed model.

To conduct a comparative analysis, we replicated the structural design of DeblurGan [13] and DeblurGanV2 [14], and subsequently subjected them to training using our suggested dataset. The visual comparison between the two models and the proposed VAE network is depicted in Fig. 8. The analysis reveals that the DeblurGan [13] and DeblurGanV2 [14] models exhibit unsatisfactory performance, as evidenced

by their inability to produce output images of comparable quality and texture to the original ground truth image. The VAE network proposed in the majority of cases exhibits superior performance in deblurring and effectively enhances the legibility of license plates, resulting in the identification of information contained within the plate image. The results depicted in Fig. 8 demonstrate that the model proposed in this study exhibits superior performance in generating license plate images that are deblurred, noise-free, and the semantic characteristics of the image are preserved.

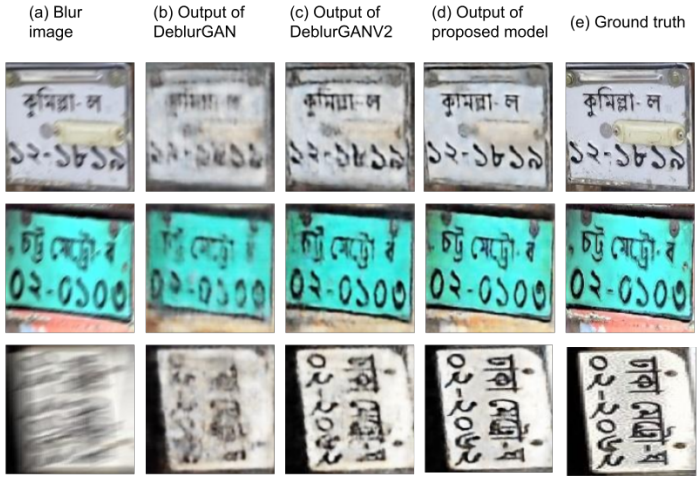

Fig. 8. Comparison with state-of-the-art-models; where a) input blurred image, b) Output of DeblurGan [13], c) Output of DeblurGanV2 [14], d) generated image by the proposed VAE model and (e) is the ground truth image.

The assessment of performance with respect to three primary metrics, namely Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Inference Time, is presented in Table I. The validation set of our proposed datasets is employed for the purpose of validating these models. Table I demonstrates that the PSNR and SSIM metrics of our model are significantly higher than those of the previous state-of-the-art models. The model under consideration attained a PSNR score of 32.41 and an SSIM score of 0.934. The VAE network proposed exhibits superior performance in terms of producing visually, semantically correct, and readable outcomes. Unlike alternative neural models, our model is not dependent on L2 distance within the pixel area, and therefore, it is not inherently optimized for the PSNR metric. The output, as depicted in Fig. 7, demonstrates the capability of the method to effectively mitigate the effects of camera shake and object movement-induced blur. Furthermore, the approach does not exhibit the typical artifacts that are commonly observed in kernel estimation techniques. Simultaneously, it exhibits the shortest inference time.

TABLE I. PERFORMANCE EVALUATION WITH RESPECT TO PSNR, SSIM, AND INFERENCE TIME

Performance Metrics	DeblurGan [13]	DeblurGanV2 [14]	Proposed VAE Network
PNSR	20.99	24.15	32.41
SSIM	0.504	0.563	0.934
Inference Time	0.85s	0.28s	0.16s

VI. CONCLUSION

This research introduces a new approach for enhancing the clarity of license plate images through the utilization of a novel Variational AutoEncoder(VAE) network. The experimental findings indicate that the proposed model surpasses the current state-of-the-art deblurring techniques, exhibiting a higher peak signal-to-noise ratio (PSNR) score of 32.41 and a structural similarity index measure (SSIM) score of 0.934. The proposed approach utilizes VAE to acquire a reduced-dimensional depiction of the image data, thereby facilitating the efficient elimination of blurriness and noise from license plate images. In future research, our suggested model may be further refined to minimize the duration of execution and enhance its efficacy, rendering it appropriate for real-time implementations. Despite the absence of a publicly accessible dataset for Bangla license plate deblurring, the model can be effectively trained on license plates from other countries to enhance its applicability. Furthermore, the model has the potential to be deployed on edge devices, such as smartphones or embedded systems, thereby facilitating license plate recognition applications in resource-constrained or remote settings.

ACKNOWLEDGMENT

We appreciate the financial support provided by the Institute of Energy, Environment, Research, and Development (IEERD, UAP) and the University of Asia Pacific

REFERENCES

[1] S. Du, M. Ibrahim, M. Shehata, and W. Badawy, "Automatic license plate recognition (alpr): A state-of-the-art review," *IEEE Transactions on circuits and systems for video technology*, vol. 23, no. 2, pp. 311–325, 2012.

[2] C. Gou, K. Wang, Y. Yao, and Z. Li, "Vehicle license plate recognition based on extremal regions and restricted boltzmann machines," *IEEE transactions on intelligent transportation systems*, vol. 17, no. 4, pp. 1096–1107, 2015.

[3] R. Rahman, T. S. Pias, and T. Helaly, "Ggcs: A greedy graph-based character segmentation system for bangladeshi license plate," in *2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*. IEEE, 2020, pp. 1–7.

[4] R. Rahman, A. F. Rakib, M. Rahman, T. Helaly, and T. S. Pias, "A real-time end-to-end bangladeshi license plate detection and recognition system for all situations including challenging environmental scenarios," in *2021 5th International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*. IEEE, 2021, pp. 1–6.

[5] H. H. A. Jassim, Z. M. Hussain, H. R. Shaaban, and K. B. Al-dbag, "Blurring and deblurring digital images using the dihedral group," *International Journal of Advanced Research in Artificial Intelligence*, vol. 4, no. 12, 2015. [Online]. Available: http://dx.doi.org/10.14569/IJARAI.2015.041204

[6] W.-Z. Shao, Y.-Y. Liu, L.-Y. Ye, L.-Q. Wang, Q. Ge, B.-K. Bao, and H.-B. Li, "Deblurgan+: Revisiting blind motion deblurring using conditional adversarial networks," *Signal Processing*, vol. 168, p. 107338, 2020.

[7] K.-H. Liu, C.-H. Yeh, J.-W. Chung, and C.-Y. Chang, "A motion deblur method based on multi-scale high frequency residual image learning," *IEEE Access*, vol. 8, pp. 66 025–66 036, 2020.

[8] J. Fang, Y. Yuan, W. Ji, P. Tang, and Y. Zhao, "Licence plate images deblurring with binarization threshold," in *2015 IEEE International Conference on Imaging Systems and Techniques (IST)*. IEEE, 2015, pp. 1–6.

[9] P. Svoboda, M. Hradiš, L. Maršík, and P. Zemcík, "Cnn for license plate motion deblurring," in *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016, pp. 3832–3836.

[10] M. Noroozi, P. Chandramouli, and P. Favaro, "Motion deblurring in the wild," in *Pattern Recognition: 39th German Conference, GCPR 2017, Basel, Switzerland, September 12–15, 2017, Proceedings 39*. Springer, 2017, pp. 65–77.

[11] P. Wieschollek, M. Hirsch, B. Scholkopf, and H. Lensch, "Learning blind motion deblurring," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 231–240.

[12] H. Tomosada, T. Kudo, T. Fujisawa, and M. Ikehara, "Gan-based image deblurring using dct discriminator," in *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2021, pp. 3675–3681.

[13] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "Deblurgan: Blind motion deblurring using conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8183–8192.

[14] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, "Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, pp. 8877–8886.

[15] R. Ran, L. Deng, T. Jiang, J. Hu, J. Chanussot, and G. Vivone, "Guidednet: A general cnn fusion framework via high-resolution guidance for hyperspectral image super-resolution." *IEEE Transactions on Cybernetics*, 2023.

[16] J. Fang, H. Lin, X. Chen, and K. Zeng, "A hybrid network of cnn and transformer for lightweight image super-resolution," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2022, pp. 1102–1111.

[17] Q. Lu, W. Zhou, L. Fang, and H. Li, "Robust blur kernel estimation for license plate images from fast moving vehicles," *IEEE Transactions on Image Processing*, vol. 25, no. 5, pp. 2311–2323, 2016.

[18] S. Zheng, Z. Zhu, J. Cheng, Y. Guo, and Y. Zhao, "Edge heuristic gan for non-uniform blind deblurring," *IEEE Signal Processing Letters*, vol. 26, no. 10, pp. 1546–1550, 2019.

[19] C. Zhao, Y. Wang, H. Jiao, J. Yin, and X. Li, "*l_p*-norm-based sparse regularization model for license plate deblurring," *IEEE Access*, vol. 8, pp. 22 072–22 081, 2020.

[20] N.-A. Alam, M. Ahsan, M. A. Based, and J. Haider, "Intelligent system for vehicles number plate detection and recognition using convolutional neural networks," *Technologies*, vol. 9, no. 1, p. 9, 2021.