Modified Base Autoencoder and Variational Autoencoder for Denoising Images in CIFAR-10 and MNIST Datasets

Jenan A. Alhijaj[1], Rana J. AL-Sukeinee[2], Asaad A. Alhijaj[1], Noor M. Al-Moosawi[1], Raidah S. Khudeyer[1]
[1]Department of Computer Science, College of Computer Science and Information Technology, University of Basrah, Basrah, Iraq
[2]Department of Physics, College of Science, University of Basrah, Basrah, Iraq
E-mail: jenan.alkereem@uobasrah.edu.iq, rana.jabbar@uobasrah.edu.iq, asaad.abdulhassan@uobasrah.edu.iq, almoosawinoor2@gmail.com, raidah.khudayer@uobasrah.edu.iq

Keywords: PSNR, SSIM, denoising image, autoencoder, variational autoencoders

Received: June 9, 2025

With the increasing volume of digital images, we must increase the quality of images for accuracy and visible applications, and we need ways to reduce the image noise while keeping important features such as edges, corners, and sharp details. In recent years, deep learning algorithms have become more significant for solving image denoising problems because they can simulate complex image patterns. This paper compares the performance of modified base AutoEncoders (AEs) and Variational Autoencoders (VAEs) models for image denoising in CIFAR-10 for color images and MNIST for grayscale images datasets. Our proposed modification to base AE and VAE architectures consists of changes in the encoder and decoder layers of feature extraction and reconstruction abilities, resulting in improved denoising performance. To simulate real-world image damages, data preparation involved normalization and the injection of Gaussian noise (0.5 for MNIST and 0.5 for CIFAR-10). With batch normalization and UpSampling2D layers with sigmoid outputs, the encoder-decoder architecture guaranteed the accuracy of spatial reconstruction, while VAE combined MSE with KL divergence for latent regularization, and AE optimized MSE reconstruction loss. The two models' performance was evaluated using important essential metrics: the Structural Similarity (SSIM) and the Peak Signal to Noise Ratio (PSNR). In both datasets, the results indicate that the VAE model outperforms the AE model in terms of image quality. The CIFAR-10 color dataset was given an SSIM of 0.954 and a PSNR of 32.86 dB, whereas the MNIST grayscale dataset provided an SSIM of 0.951 and a PSNR of 24.44 dB to the modified VAE model. In addition, the CIRAF-10 dataset achieved an SSIM of 0.891 and a PSNR of 27.72 dB, whereas the MNIST dataset was given an SSIM of 0.883 and a PSNR of 29.06 dB from the AE model. This study addresses how AE and VAE architectures differ in denoising performance across dataset complexities and the principles for optimal model selection.

Povzetek: Spremenjeni modeli VAE in AE so bili primerjani pri odstranjevanju šuma iz slik, pri čemer je VAE na podatkovnih zbirkah MNIST in CIFAR-10 dosegel boljšo kakovost rekonstrukcije.

1 Introduction

The study of deep learning, a branch of machine learning, has brought about a new phase in the advancement of neural networks [2]. A crucial component of neural networks, an autoencoder efficiently compresses (encodes) input data to its most basic components before using this compressed representation to reconstruct (decode) the original input. The reduction of the features from high dimensions to low dimensions leads to an enhanced quality of the digital images [1]. Autoencoder has several applications, such as data compression and image denoising [3]. In addition, abnormality Detection in industrial operations. In addition, Autoencoders can be used in recommendation systems that can be used for personalized suggestions and potential representations of users [4].

Similar to classic autoencoders but using a probabilistic framework, variational autoencoders (VAEs) are a kind of generative model; it's use generative models to produce new data in the form of variants of the input data that the models are trained on [5]. Neural networks known as VAEs take in input data and transform it into its latent representation, which consists of hidden layers for the variables mean and variance. then, the original data is recreated by converting these two hidden layers into sampled latent vectors [6]. VAEs have various applications like Latent Variable Modelling, Data Augmentation, Image Generation, and Anomaly Detection. Moreover, VAEs can produce new data points enabled by their probabilistic structure, which is especially helpful for tasks like image denoising [7].

To reduce noise while maintaining important image information, image denoising is a basic task in computer

vision, medical imaging, surveillance systems, and image processing [8]. The noise can be brought about by various elements like transmission faults, bad lighting, and sensor noise. Prior to denoising, a noisy image is first obtained and then preprocessed to enhance its characteristics and reduce artifacts. Particularly, in high-noise situations, traditional denoising approaches battle to reduce noise while maintaining fine features and which can be brought due to various elements like transmission faults, bad lighting, and sensor noise [9]. Deep learning, in particular, has become one of the most effective methods for image denoising. These methods can automatically adjust to various noise types and levels by directly learning noise patterns from data, which improves denoising performance [10].

The main aim of the research is to determine the best autoencoder architecture for image denoising by systematically comparing deterministic Autoencoders (AE) and probabilistic Variational Autoencoders (VAE) across datasets with different levels of complexity (e.g., simple grayscale MNIST versus complex color CIFAR-10). The main hypotheses suggest that VAEs will surpass AEs on complex datasets because of their probabilistic latent space and regularization, resulting in higher PSNR and SSIM; while AEs will perform better in pixel-level accuracy (PSNR) on simple datasets but will have lower perceptual quality (SSIM). Objectives, including a quantitative evaluation of denoising performance through PSNR/SSIM metrics, sampling of PSNR-SSIM to data complexity, and the practical model selection guidelines, favor AEs for uniform data that requires pixel precision and VAEs for complex data that necessitates structural fidelity.

2 Related works

In 2020[11], Li et al. This study used grayscale image datasets and the Matrix-variate Variational Auto-Encoder (MVVAE) technique. It uses matrix neural networks (MVMLPs) for the encoder and decoder, a matrix Gaussian distribution for the latent variable, and direct matrix modelling of 2D image data (input, hidden, latent, and output variables are matrices). According to experimental results, MVVAE outperforms VAE

In [12], Pawar et al. described a method for achieving autoencoders with image processing and deep learning algorithms to minimize the noise in images. The primary goal of using an autoencoder to eliminate noise is to preserve originality because autoencoders use the backpropagation process instead of the conventional methods. The method covered in the paper is dependable, effective, compatible with a wider range of devices, interconvertible, and applicable to any signal.

In [13], Ranganath et al. A multi-stage Gaussian noise reduction technique utilizing autoencoders and recurrent neural networks (RNNs) is presented in the study. A modified version of the CIFAR-10 dataset was used for their experiments, where every image was changed to a single-channel grayscale. MSE and SSIM are the main performance measures for the two denoising techniques. RNN that uses autoencoders at each denoising stage and a convolutional autoencoder. The RNN network reduces noise by mimicking a multi-stage process as it learns to denoise increasingly on images during training.

Patil et al. [28], provided an article that focuses on the assessment of classification accuracy for different image noise levels, particularly about image denoising methods that combine spatial filters and autoencoders. Three different datasets were used: MNIST, CIFAR-10, and a medical dataset (mini-MIAS), which were used to assess the model and obtain a comparative analysis of the denoising pipeline's performance. At different noise levels, the AE pipeline demonstrated reduced classification accuracy on the mini-MIAS medical dataset. This analysis is essential for creating efficient image denoising techniques because it shows how noise impacts image quality and how well various methods can counteract these effects.

In [14], P.Venkataraman et al., this research used Convolutional Autoencoder, Basic Autoencoder, and Variational Autoencoder to denoise the image. There are two types of losses in the variational autoencoder: latent loss and generative loss. The convolutional and basic autoencoders, on the other hand, only have one loss parameter. In this project, autoencoders are implemented using Keras as the backend and TensorFlow as the frontend.

In [15], Ranjan et al., this research presents a method for denoising grayscale photographs in the MNIST (Handwritten Digits) dataset using Convolutional Neural Networks (CNNs). Performance is measured using PSNR (dB) for denoised images. The authors intend to work on real image denoising techniques; the method has not been tested on high-resolution or complex, real-world color images.

Muazzez Buket DARICI et al. [16], presented a comparison between traditional methods like Gaussian and Salt & Pepper and deep learning methods such as deep convolutional denoising autoencoders (CDAE) to denoise facial pictures. The denoising performance of autoencoders and conventional techniques has been estimated using each computational time and accuracy criterion. The experiment results showed that the autoencoders performed better based on the Adam optimizer for eliminating both kinds of noise in comparison with PSNR and SSIM measurements.

Another study in 2023[17], S. R. Tusher et al., proposes a new Variational Autoencoder (VAE) for deblurring license plates. Corresponding to the experimental conclusions, the suggested system approach achieves improved results than the other cutting-edge deblurring techniques. Consequently, it improves the ALPDR system's accuracy and reliability.

In 2024[19], Younus FAROOQ and Serkan SAVAŞ Introduced work to denoise from image medical images using a CNN-based denoising autoencoder. This work uses datasets like the SIIM-medical-images and Covid19 radiography database. The results obtained based on the

RMSE and PSNR criteria have proven to be superior to the basic method.

Kimleang Kea et al. [20] focused on reducing image noise through the use of approach of quantum convolutional autoencoder (QCAE). In this work, the researchers used the electrical circuit instead of the latent space autoencoder's representative , and a higher structural similarity index (SSIM) value with less training loss. The suggested QCAE approach performed better than its conventional equivalent.

The previous studies do not compare AE/VAE under the same conditions and lack metric consistency. In this study for AE/VAE denoising, we used the MNIST dataset for grayscale images and the CIFAR-10 for color images. We produced one evaluation of AE and VAE on both grayscale MNIST and color CIFAR-10, as well as performance metrics based on standard metrics (PSNR/SSIM).

The summary of the related works is in Table 1.

Table 1: Summarization of related works of image denoising

I.	Study	Technique	Dataset / Color	Metrics	Remarks
1	Li et al. [11] (2020)	VAE	MNIST / Grayscale	PSNR:17.253dB SSIM: 0.7938 NMSE: 0.1025	low value of PSNR, SSIM metric
2	Pawar et al. [12] (2020)	Autoencoders	MNIST / Grayscale	PSNR:29.54 dB MSE: 72.26	No SSIM metric
3	Ranganath et al. [13] (2021)	Autoencoders	CIFAR-10 / Color	MSE: 0.001446 SSIM: 0.9478	No PSNR metric; Complex training
		RNN		MSE: 0.001394 SSIM:0.9485	
4	Patil, et al. [28] (2021)	Autoencoders	MNIST / Grayscale	PSNR: 7. 17 - 21.94 dB	No SSIM metric; Poor performance
			CIFAR-10 / Color	8.12-19.30 dB	
5	P.Venkataraman et al. [14] (2022)	Basic AE, AE, Convolutional AE	CIFAR-10 / Color	VAE: Accuracy: 0.143	No PSNR, SSIM metrics
6	Ranjan, et al. [15] (2022)	CNN	MNIST / Grayscale	PSNR: 21.65 dB	No SSIM metric
7	M. B. Darici et al. [16] (2023)	Convolutional Denoising AE (CDAE)	Autism Image/ Color	PSNR:23.09 dB SSIM: 0.62	Low SSIM metric; Specific application
8	S. R. Tusher et.al.[17] (2023)	Variational autoencoder	Blurred license plate images / Color	SSIM:0.934 PSNR:32.41	Focus on text and fails in natural images
9	Y. Faarooq et al. [19] (2024)	(CNN) denoising autoencoder	Medical (CXR, CT) / Grayscale	PSNR: 7.82% - 42.37% RMSE: 0.65% - 10.77%	PSNR percentage (nonstandard); Medical focus
10	K. Kea et al. [20] (2024)	Quantum Convolutional AE (QCAE)	MNIST/ Grayscale	SSIM:0.75	Low SSIM metric

3 Base autoencoder

Since the autoencoder's goal output is the same as its input, it has gained a lot of interest in the domain of image processing. To remove noise from the source data and visualize high-dimensional data, removing duplicate data to enhance performance, minimizing storage usage as well as speeding up the time required for the same

calculation, therefore it is used in data compression [21]. An Encoder, Decoder as well as a Bottleneck layer, are basic components of autoencoder architectures that do not rely on labeled training data; therefore, autoencoders are not regarded as a supervised learning technique. Encoder: These layers take the raw Bottleneck. The most important part of the network is the module that saves the compressed knowledge representations. input data and compact into a lower-dimensional representation (latent space). This part reconstructs the original data by decoding the latent representation. It is converting the encoded representation back to the dimensions of the original input. The objective of the hidden layers is to rebuild the original input by progressively increasing their dimensionality. The reconstructed output is produced by the output layer and should preferably be as close to the input data as is practical [22].

4 Variational autoencoder

The encoded latent space and the variety of use cases that their probabilistic encoding provides are a distinct way for VAEs from other autoencoders. In addition,
the latter's ability to provide a range of data or continuous data in the latent space, known as a variational encoder, helps to create new images or data. The first step, the encoder network transforms an input into a latent space by mapping it, which is a probability distribution [18]. This is a Gaussian distribution that is multivariate. After that, VAEs take a sample of the latent space in a manner that is differentiable in relation to the encoder's parameters.
Finally, the Decoder Network compressed the latent space to reconstruct the input data [23].

5 Methodology

5.1 Datasets

This study uses two popular datasets, CIFAR-10 and MNIST, downloaded from Kaggle, to demonstrate advanced image denoising capabilities using optimized autoencoders. The CIFAR-10 dataset covers ten classes, each with 60,000 color images, each coming in (32x32x3) pixel sizes. Ten thousand images were included in the test set and fifty thousand in the training set [25]. In addition, 70,000 images of the grayscale handwritten numbers (0 - 9), each measuring (28x28x1) pixels, are part of the MNIST collection. from this dataset, 10,000 images were allocated for testing, and 60,000 for training [24].

5.2 Denoising autoencoder (AE)

An unsupervised learning method creates new data through two processes: encoding and decoding. In the encoding stage, the input is collapsed into the latent space and then reconstructed by the decoder using the information from the latent variables. Consequently, the two encoders and decoders can be assumed to as two symmetric neural networks, Figure 1. Autoencoders are useful for extracting low-level information for use in machine learning or deep models. Autoencoders can be employed for a wide range of applications, including dimension and noise reduction. This study intends to make use of the autoencoder's noise reduction capability.

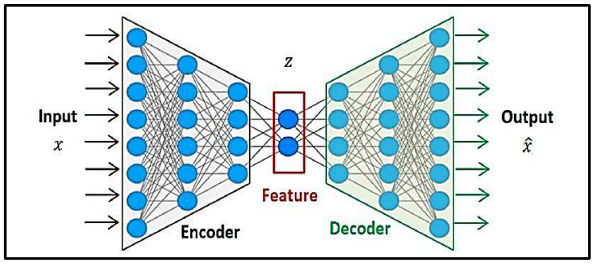


Figure 1: AutoEncoder(AE) architecture [26].

5.3 Variational autoencoder (VAE)

Neural networks transform input data into a latent representation with hidden layers for mean and variance. These two hidden layers are then translated into sampled latent vectors, which are used to rebuild the original data. In VAE, the loss function includes both reconstruction loss (rating how well the output matches the input) and a regularization term (Kullback-Leibler divergence) to guarantee the learned latent space follows a desirable distribution. Figure 2 depicts a variational autoencoder architecture, the encoder samples the input and transforms it to a latent space representation. The decoder uses this representation to reassemble the input image.

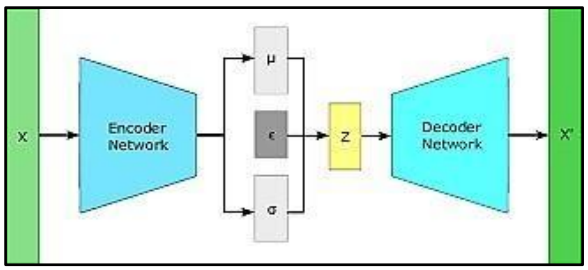


Figure 2: Variational AE (VAE) structure [27].

5.4 Proposed method

The proposed method consists of major pipeline steps: The first step involves loading the images of the dataset, which is either MNIST (grayscale) or CIFAR-10 (color). The dataset is split into training and testing subsets. Afterwards, preprocessing takes place: the images are resized and their pixel values are normalized. To imitate a genuine image as it manifests in reality, Gaussian noise was injected; the ideal levels of Gaussian noise (σ) vary among the datasets, since grayscale images can withstand higher σ than color images because of the interdependence of channels in RGB data, where the color spaces show a reduced tolerance for noise. Data augmentation techniques were utilized to improve the dataset's diversity and robustness. The diagram flowchart in Figure 3, illustrates the proposed method.

The core architectural selection is whether to choose a standard Autoencoder (AE) or a Variational Autoencoder

(VAE). The model's encoder layers gradually decrease the dimensionality of the noisy input, compressing their information and extracting crucial features while eliminating superfluous noise. In AE, the bottleneck consists of a latent space vector, whereas in VAE, it comprises a probabilistic representation where the latent vector Z is sampled via reparameterization. The decoder layers aim to produce a clean output by reconstructing the image from this compressed representation. During the training phase, the reconstructed images are assessed against the original input, and the effectiveness of the trained model is measured using PSNR and SSIM metrics. The end product is a denoised image from which the model has learned to remove the Gaussian noise that was added previously.

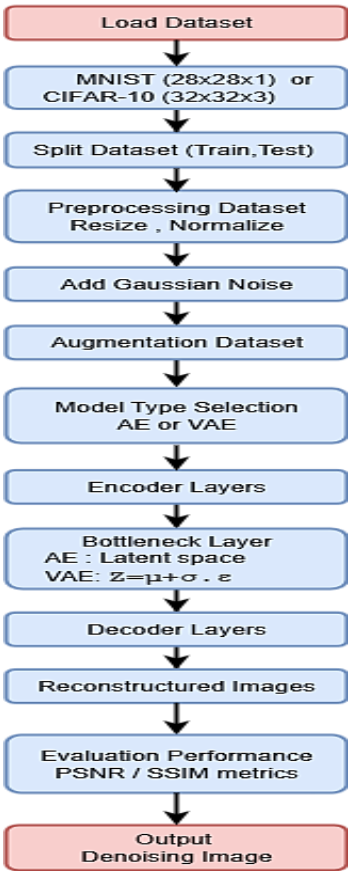


Figure 3: Flowchart of proposed procedures for AE and VAE models with MNIST and CIFAR-10.

5.5 Training procedure

The Autoencoder (AE) and Variational Autoencoder (VAE) models were trained by first splitting the dataset into training and testing sets (80,20) and then normalizing the by rescales input pixel values from their original range of (0,255) to a standardized range of (0,1). Gaussian noise (σ=0.2 for CIFAR-10 images and σ=0.5 for MNIST images), with (0 mean, 1 standard deviation) for both datasets, was added to the original images to simulate real-world noise.

All source codes and model training were run on the Kaggle platform. This makes NVIDIA TESLA P100 GPUs available for free. The models utilize TensorFlow versions (2.16.1), Keras versions (3.3.3), and Python programming language libraries (version 3.10.14).

The model architecture, which includes encoder and decoder components, works together to learn and reconstruct data. The encoder compresses the input data into a lower-dimensional latent representation, retaining crucial properties while eliminating irrelevant information. In AEs, the latent representation is a fixed vector, whereas in VAEs, the encoder generates Gaussian distribution parameters (mean and variance), enabling the latent space to be interpreted probabilistically. The decoder will reconstruct the original input using this latent representation; in AEs, it does so directly from a fixed latent vector, but in VAEs, the learner first samples from a Gaussian distribution. Probabilistic methods in VAEs allow for systematic sampling and the generation of additional data points. Equation (1) determines the sample z [29]:

$$Z = \mu + \sigma_{\bullet} \varepsilon \, , \, \epsilon \sim \mathcal{N}(0,I) \quad \ldots (1)$$

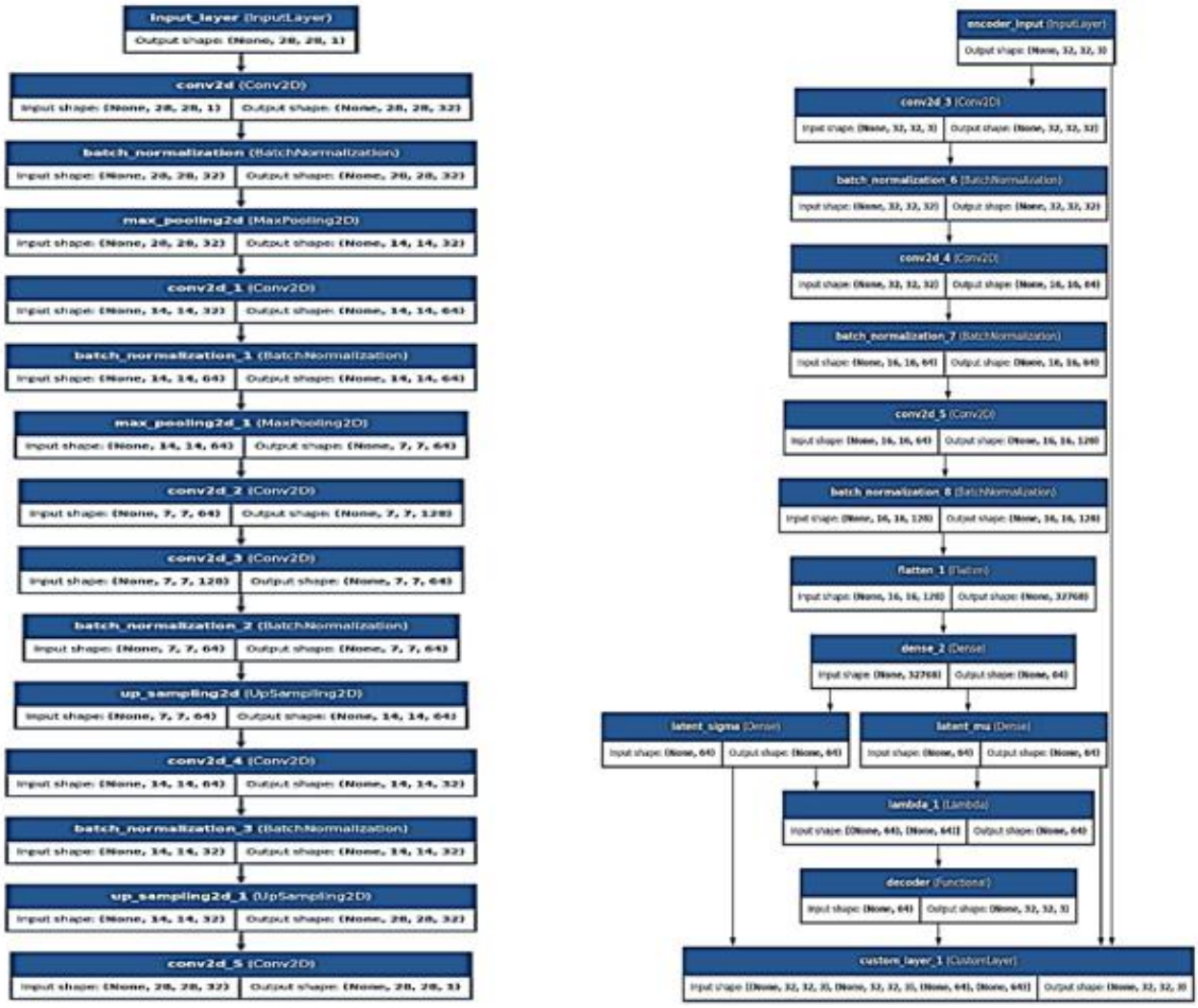
This outlines how to sample; the essential principle underlying variational autoencoders (VAEs) is to start with a Gaussian distribution in latent space. Where Z is the latent variable, μ is the encoder's mean, σ is the standard deviation, and ε is a random variable (mean 0, variance 1) obtained from the standard normal distribution. Many latent representations are improved as an outcome of the model's ability to consist of randomness in the sampling procedure according to this equation. The VAE model can keep flexibility while effectively discovering the latent space and constructing new data points that are similar to the training data.

There is only one loss parameter in base autoencoders; the mean squared error (MSE) is used as a suitable loss function to evaluate the variation between the input and the reconstructed output. The two components of the variational autoencoder loss function are the latent loss and reconstruction loss (Kullback-Leibler or KL) variance. Reconstruction loss measures the model's capability to recover input data from latent representations and is computed using mean squared error. The KL deviation term indicates how close a known latent distribution is to the previous distribution, and a normal distribution promotes accurate reconstruction and a well-organized latent.

In terms of model construction, both AE and VAE utilize the same 3 layers convolutional encoder (with 32, 64, and 128 filters, 3×3 kernels), and ReLU activation and batch normalization layers, but they differ in their latent space, with thayers of the layers of latent_mu and latent_sigma: AE relies on a fixed 64D vector, whereas VAE uses probabilistic sampling with Z parameterization. Images are symmetrically reconstructed by the decoders of both models using UpSampling2D layers, convolutional blocks with (128, 64, and 32 filters), and a sigmoid output to ensure valid

pixel ranges. Model configuration utilizes the Adam optimizer (lr=0.01) with divergence-aware loss functions: MSE for AE and a combination of MSE and KL divergence for VAE. The training loop iterates for 100 epochs (with batch sizes of 32, 64, and 128), performing forward passes, computing loss, executing backpropagation, and updating weights for each batch, while applying batch normalization after each convolution to ensure stable learning.

Upsampling is an important technique used in the decoder component to reassemble the original input from the compressed latent representation. The UpSampling2D layer was used to double the height and the width of the feature maps, which is necessary for maintaining the spatial structure of the image data. Upsampling layers followed by convolutional layers in decoders allow the model to refine features and generate high-quality reconstructions. Allowing the model to gradually reconstruct the image from a low-dimensional latent space to the original input dimensions to generate outputs that look like the original input data in both AEs and VAEs. The model reconstructs the input in a forward pass, computes the loss by comparing the reconstruction to the original data, and then changes the model weights via backpropagation. Validation was performed after each epoch to check performance and minimize overfitting, utilizing the regularization approach and batch normalization layers to stabilize and expedite training, which improves generalization. The flow diagram of the top-performing AE-MNIST and VAE-CIFAR-10 model architecture is illustrated in Figure 4.



(a) (b)

Figure 4: The diagrams of the top-performing AE-MNIST and VAE-CIFAR-10 model architecture (a) AE-MNIST, (b) VAE-CIFAR-10.

6 Results and discussion

6.1 Evaluation metrics

This study uses the SSIM and PSNR as assessment metrics for the comparison of the performance of denoising Autoencoders (AE) and Variational Autoencoders (VAE) on images from the MNIST and CIFAR-10 datasets.

PSNR value, obtained from Equation (2) [30]:

$$PSNR=20\times\log_{10}(MAXI / (\sqrt{MSE})) \quad \ldots (2)$$

where MAXI is the image's maximum pixel value; the MAXI for most PNG images is 255.

Mean Squared Error (MSE), is the difference between the original and processed images; Equation (3)[16].

Where:

$$\mathbf{MSE} = \frac{1}{m \cdot n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [O(i,j) - D(i,j)]^2 \quad \ldots (3)$$

MSE, or mean square error, is used in Equation (3). RMSE is based on MSE. O is the original image matrix. D represents the image matrix after the denoising process. All PSNR values were acquired in units dB, which stands for decibels.

The SSIM value is obtained from Equation (4) [30]:

$$SSIM(x,y) = (2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2) / ((\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma^2 y + C_2)) \quad \ldots (4)$$

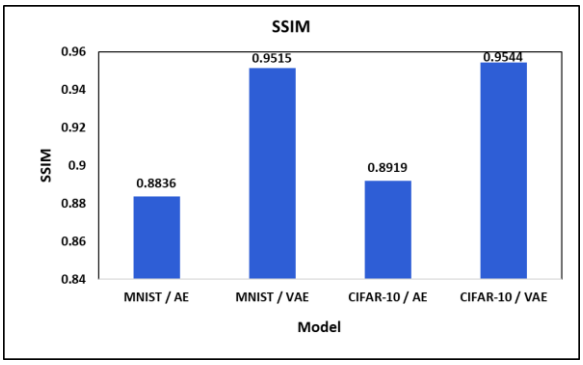
μ_x and μ_y represent the average values of x and y, respectively, $\sigma^2 x$ and $\sigma 2y$ denote the variances of images x and y, respectively, σ_{xy} indicates the covariance between images x and y, while C1 and C2 are constants used to prevent division by zero.

The VAE, with its probabilistic latent space, achieves higher SSIM scores (0.951 for MNIST, 0.954 for CIFAR-10), showing better structural detail preservation, although the AE performs well in PSNR of MNIST, with a value (29.06 dB). On CIFAR-10, VAE outperforms the AE in both SSIM and PSNR (32.86 dB), as seen in Table 2.

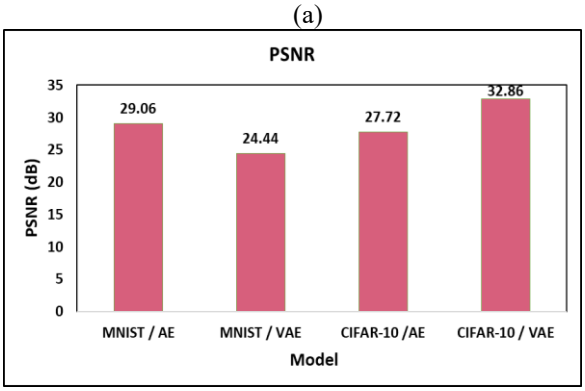
Table 2: Average SSIM and PSNR results of AE and VAE on the MNIST and CIFAR-10 datasets

Model	Dataset	SSIM	PSNR (dB)
Autoencoder (AE)	MNIST gray images	0.883	29.06
	CIFAR-10 color images	0.891	27.72
Variational Autoencoder (VAE)	MNIST gray images	0.951	24.44
	CIFAR-10 color images	0.954	32.86

The visualization of the comparison analysis between AE and VAE performance across the MNIST and CIFAR-10 datasets for SSIM and PSNR metrics is illustrated in Figure 5.



(a)



(b)

Figure 5: AE and VAE performance of MNIST and CIFAR-10 datasets for SSIM and PSNR metrics (a) SSIM, (b) PSNR.

Architecture is important, where VAEs perform better in states where probabilistic modeling is required due to data complexity. The dataset complexity determines model selection; AE deterministic reconstruction performs well on uniform grayscale digits, and AE is suitable for basic data MNIST and reaches pixel precision.KL regularization and probabilistic latent modeling are crucial components of VAEs' design for complex data (CIFAR-10). Regularization and KL divergence in VAEs reduce overfitting, which is important for the diversity of CIFAR-10's color and texture. While deterministic AEs may overfit to noise in complicated data, probabilistic modeling is better at capturing complex distributions, and VAEs' regularization helps in avoiding overfitting

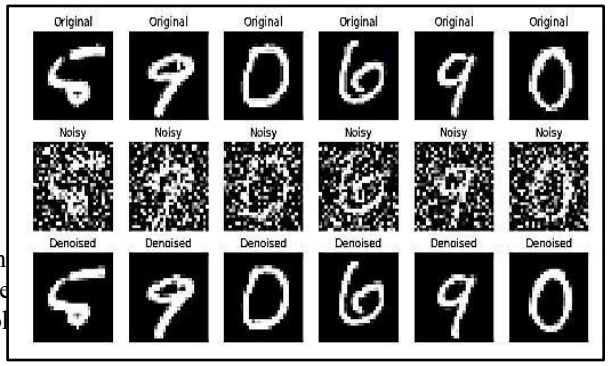
6.2 Discussion

The proposed VAE exceeds specialized models and establishes a new SOTA on CIFAR-10. The suggested AE is less appropriate for complicated data, yet competitive on MNIST. Setting a new SOTA for denoising on CIFAR-10, the suggested VAE exceeds previous studies such as [13] (SSIM: 0.9485) and [28] PSNR: 19.30 dB) by modelling complex color and texture

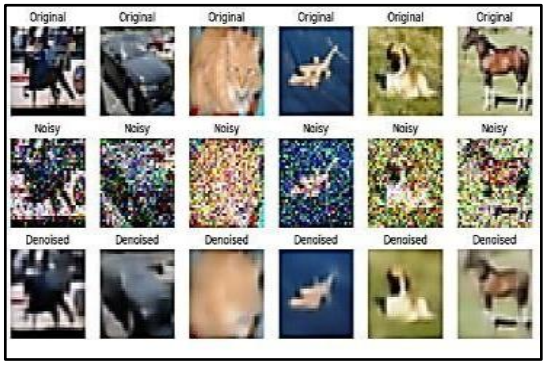
changes with a PSNR (32.86 dB) and SSIM (0.954) due to its probabilistic design. The VAE outperforms its actual performance (SSIM: 0.951) while [11] achieves (0.7938) despite lower PSNR owing to basic reducing, the AE excels in pixel-level accuracy (PSNR: 29.06 dB), approaching SOTA [12] (29.54 dB) and removing [11] (17.25 dB) on MNIST for pixel-level accuracy. AEs perform best on simple, uniform data (MNIST) for PSNR, while VAEs perform better on complex data (CIFAR-10) through regularization and latent distribution modelling. This reveals limitations in previous studies, such as inconsistent metrics [12], [15], [28], lack of SSIM, architectural flaws [11], [28], and specialized applications [13], [17].

The images in Figures 6 and 7 show the outcomes of applying AE and VAE to the CIFAR-10 and MNIST datasets. On MNIST, AE results in sharper digit reconstructions that retain noise artifacts, whereas VAE experiences over-smoothing. On CIFAR-10, VAE provides superior visual quality, vibrant color retention, and sharper edges. AE faces difficulties in this area, producing desaturated hues and indistinct outlines because of its deterministic bottleneck, while VAE demonstrates superior performance on intricate color data through the modelling of probabilistic distributions.

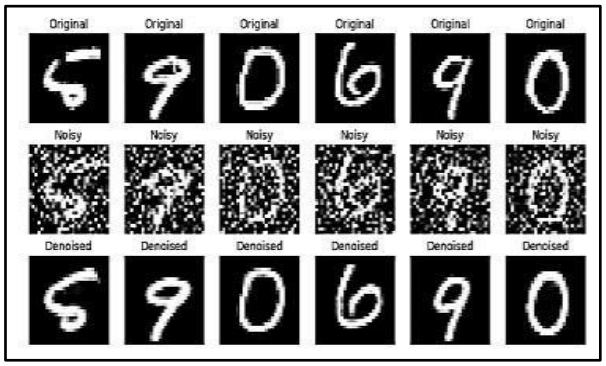
Th
pr
co



(a) AE-MNIST



(a) AE-CIFAR-10



(b) VAE-MNIST

Figure 6: Results of applying (a) AE and (b) VAE to the MNIST Dataset.

(b) VAE-CIFAR-10

Figure 7: Results of applying (a) AE and (b) VAE to the CIFAR-10 Dataset.

7 Conclusion and future work

The results show that VAE performs better for image denoising than other models, especially in dealing with structural similarity (SSIM) and complex datasets such as CIFAR-10. However, AE is effective at reducing pixel-level errors (PSNR) on simple datasets such as MNIST.

Reconstructing high-quality images with maintained structural details is made easier by the VAE's capacity to simulate the latent space distribution. The results presented demonstrate that in simple images, AE maintains pixel - level accuracy, but VAE is better at structural reconstruction for complicated data, suggesting it could be applied to computer vision applications where image clarity is important in training models. This study supports the creation of adaptive denoising systems between AE and VAE across various dataset

complexities, which is a crucial area for further research and improvement.

Future studies could look into combining the strengths of both approaches, such as employing hybrid models, fine-tuning hyperparameters, or refining the VAE's loss function to increase pixelwise SSIM and PSNR performance without sacrificing structural similarity.

References

[1] Alhijaj, J. A., & Khudeyer, R. S. (2023). "Integration of efficientnetb0 and machine learning for fingerprint classification." *Informatica*, *47*(5). https://doi.org/10.31449/inf.v47i5.4724

[2] R. J. AL-Sukeinee and R. S. Khudeyer, "Review: Deep Learning and Fuzzy Logic Applications," Engineering and Technology Journal, vol. 9, no. 2456-3358, pp. 4231-4240, 2024. https://doi.org/10.47191/etj/v9i06.09

[3] Malhotra, R., & Singh, P. (2023). "Recent advances in deep learning models: a systematic literature review". *Multimedia Tools and Applications*, *82*(29), 44977-45060. https://doi.org/10.1007/s11042-023-15295-z

[4] Berahmand, K., Daneshfar, F., Salehi, E. S., Li, Y., & Xu, Y. (2024). "Autoencoders and their applications in machine learning: a survey". *Artificial intelligence review*, *57*(2), 28. https://doi.org/10.1007/s10462-023-10662-6

[5] Ma, C. (2025). "Fusion of SP-VAE and IMP-VAE for Proxy Attack Detection in E-Commerce Systems". *Informatica*, *49*(8). https://doi.org/10.31449/inf.v49i8.6906

[6] R. WEI, C. GARCIA, A. EL-SAYED, V. PETERSON, and A. MAHMOOD, "Variations in Variational Autoencoders- A comparative evaluation". IEEE Access, vol. 8, pp. 153651-153670, 2020. https://doi.org/10.1109/access.2020.3018151

[7] Biswas, B., Ghosh, S. K., & Ghosh, A. (2019). "DVAE: deep variational auto-encoders for denoising retinal fundus image". In *Hybrid machine intelligence for medical image analysis* (pp. 257-273). Singapore: Springer Singapore. https://doi.org/10.1007/978-981-13-8930-6_10

[8] M. Sreeteish, A. Mohammed, S. S. Reddy and C. N. Sujatha, "Image De-Noising Using Convolutional Variational Autoencoders", *International Journal for Research in Applied Science & Engineering Technology (IJRASET),* vol. 10, 2022. https://doi.org/10.22214/ijraset.2022.44826

[9] A. Raj1, A. A. Jadhav, C. G. Bhimrao and D. J. Anil, "Image Denoising Using Python and Machine learning". *International Journal for Research in Applied Science & Engineering Technology (IJRASET),* vol. 11, 2023. https://doi.org/10.22214/ijraset.2023.53406

[10] S. Singh and D.R. Gupta,"Comparative Analysis of Deep Learning-Based and Traditional Methods for Image Denoising and Restoration", *International Journal for Research in Applied Science & Engineering Technology (IJRASET),* vol. 12, 2024. https://doi.org/10.22214/ijraset.2024.58979

[11] Li, J., Yan, H., Gao, J., Kong, D., Wang, L., Wang, S., & Yin, B. (2020). "Matrix-variate variational auto-encoder with applications to image processs ". *Journal of Visual Communication and Image Representation*, *67*, 102750. https://doi.org/10.1016/j.jvcir.2019.102750

[12] Pawar, Aashay." Noise reduction in images using autoencoders." *3rd International Conference on Intelligent Sustainable Systems (ICISS),* pp. 987-990, 2020. *https://doi.org/10.1109/iciss49785.2020.9315908*

[13] Ranganath, A., DeGuchy, O., Singhal, M., & Marcia, R. F. (2021, October). "Multi-stage gaussian noise reduction with recurrent neural networks", In *2021 55th Asilomar Conference on Signals, Systems, and Computers* (pp. 135-139). IEEE. https://doi.org/10.1109/ieeeconf53345.2021.9723266

[14] P. Venkataraman, "Image Denoising Using Convolutional Autoencoder", *arXiv preprint arXiv:2207.11771*, 2022. https://doi.org/10.48550/arXiv.2207.11771

[15] Ranjan, A., & Azeemuddin, S. M. (2022). "Image Denoising using Convolutional Neural Network". *Proceedings - 2022 4th International Conference on Advances in Computing, Communication Control and Networking, ICAC3N 2022*, 2315–2319. https://doi.org/10.1109/icac3n56670.2022.10074437

[16] M. B. DARICI and Z. ERDEM, "A Comparative Study on Denoising from Facial Images Using Convolutional Autoencoder",*Gazi University Journal of Science,* vol. 36, no. 3, pp. 11221138, 2023. https://doi.org/10.35378/gujs.1051655

[17] S. R. Tusher, N. N. Rahman, S. Chowdhury, A. Tabassum, A. Adnan. R. Rahman and S. R. Al Masud, "An Enhanced Variational AutoEncoder Approach for the Purpose of Deblurring Bangla License Plate Images", International Journal of Advanced Computer Science and Applications, vol. 14, no. 6, pp. 10-14569, 2023. https://doi.org/10.14569/ijacsa.2023.01406133

[18] Zou, L., & Zhang, M. (2024). "Variational autoencoder model combining deep learning and probability statistics: research and application". *Informatica*, *48*(22). https://doi.org/10.31449/inf.v48i22.6921

[19] Y. FAROOQ and S. SAVAŞ, "Noise Removal from the Image Using Convolutional Neural Networks-Based Denoising AutoEncoder " Journal of Emerging Computer Technologies, vol. 3, no. 1, pp. 21-28, 2024. https://doi.org/10.57020/ject.1390428

[20] K. Kea, W.-D. Chang, H. C. Park and Y. Han, "Enhancing a Convolutional Autoencoder with a Quantum Approximate Optimization Algorithm for Image Noise Reduction," *arXiv preprint arXiv:2401.06367,* 2024. https://doi.org/10.2139/ssrn.4719914

[21] R. Malhotra and P. Singh, "Recent advances in deep learning models: a systematic literature review,". Multimedia Tools and Applications, vol. 82, no. 29, pp. 44977-45060, 2023. https://doi.org/10.1007/s11042-023-15295-z

[22] Sewak, M., Sahay, S. K., & Rathore, H. (2020). "An overview of deep learning architecture of deep neural networks and autoencoders", *Journal of Computational and Theoretical Nanoscience*, *17*(1), 182-188. https://doi.org/10.1166/jctn.2020.8648

[23] A. Solera-Rico, C. S. Vila, M. Gómez-López, Y. Wang, A. Almashjary, S. T. M. Dawson and R. Vinuesa, "β-Variational autoencoders and transformers for reduced-order modelling of fluid flows," Nature Communications, vol. 15, no. 1, p. 1361, 2024. https://doi.org/10.1038/s41467-024-45578-4

[24] M. S. Rana, M. H. Kabir and A. Sobur, "Comparison of the Error Rates of MNIST Datasets Using Different Type of Machine Learning Model," North American Academic Research, vol. 6, no. 5, pp. 173-181, 2022. https://www.researchgate.net/publication/371378159

[25] M. Kundroo and T. Kim, "Demystifying Impact of Key Hyper-Parameters in Federated Learning: A Case Study on CIFAR-10 and FashionMNIST," IEEE Access, 2024. https://doi.org/10.1109/access.2024.3450894

[26] Alaghbari, K. A., Lim, H. S., Saad, M. H. M., & Yong, Y. S. (2023). "Deep autoencoder-based integrated model for anomaly detection and efficient feature extraction in IOT networks". *IoT*, *4*(3), 345-365. https://doi.org/10.3390/iot4030016

[27] C. J. Harvey, S. Shomaji, Z. Yao, M. I. and A. Noheria, "Comparison of Autoencoder Encodings for ECG Representation in Downstream Prediction Tasks," arXiv preprint arXiv:2410.02937,2024. https://doi.org/10.48550/arXiv.2410.02937

[28] Patil, A. P., Pramod, A., Harish, A., Singh, K., & Purushotham, K. (2021, January). "An approach to image denoising using autoencoders and spatial filters for Gaussian noise". In *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 454-458).IEEE. https://doi.org/10.1109/confluence51648.2021.9377166

[29] Fu, B. (2025). "Variational Autoencoder-based High-dimensional Feature Extraction for Economic Analysis of Power Cost Data". *Informatica*, *49*(25). https://doi.org/10.31449/inf.v49i25.8012

[30] Adeshina, A. M., Razak, S. F. A., Yogarayan, S., & Sayeed, S. (2025). "Measuring Fidelity of Steganography Approach in Securing Clinical Data Sharing Platform using Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM)". *Informatica*, *49*(11). https://doi.org/10.31449/inf.v49i11.5661