Techno-Science Research Journal 12 (1) (2024) P 6-11

Content list available at ITC

Techno-Science Research Journal

Journal Homepage: http://techno-srj.itc.edu.kh/

Text Image Reconstruction and Reparation for Khmer Historical Document

Chanchen Pork[1*], Dona Valy[2] , Sokkhey Phauk[3]

[1]*Department of Information and Communication Engineering, Institute of Technology of Cambodia, Russian Federation Blvd., P.O. Box 86, Phnom Penh, Cambodia*

[2]*Research and Innovation Center, Institute of Technology of Cambodia, Russian Federation Blvd., P.O. Box 86, Phnom Penh, Cambodia*

[3]*Department of Applied Mathematics and Statistics, Institute of Technology of Cambodia, Russian Federation Blvd., P.O. Box 86, Phnom Penh, Cambodia*

Received: 07 August 2023; Accepted: 05 September 2023; Available online: June 2024

Abstract: *This research focuses on preserving Cambodia's historical Khmer palm leaf manuscripts by proposing a text-image reconstruction and reparation framework using advanced computer vision and deep learning techniques. To address the preservation, Convolutional Neural Networks (CNN) and Generative Adversarial Networks (GAN) are employed to fill in the missing patterns of characters in the damaged images. The study utilizes the SleukRith Set [1], which consists of 91,600 images divided into two parts: 90,600 training images and 1,000 test images. Each image contains a single character of the Khmer palm leaf script. The training images are intentionally degraded into three different variants, each subjected to three levels of degradation (level 1, level 2, and level 3). To assess the performance and compare the effectiveness of the Convolutional Neural Networks (CNN) and Generative Adversarial Networks (GAN) models in the proposed framework, various evaluation metrics were employed. These metrics include Mean Square Error (MSE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index (SSIM). By evaluating the results of both models based on these metrics, it was observed that the GAN model consistently outperformed the CNN model in terms of MSE, PSNR, and SSIM. The GAN model achieved lower MSE values, higher PSNR values, and higher SSIM values compared to the CNN model, indicating its superior performance in image reconstruction and preservation of the original text.*

Keywords: Text-image, Image processing, Image reconstruction, Deep learning, Model

1. INTRODUCTION

The study of historical documents is an important aspect of the protection and exploration Cambodia's artistic heritage. Still, numerous of these records have been compromised over time, performing in the loss of precious information and making it delicate for scholars and experimenters to gain a complete literal record. Working on the preservation of Cambodia's artistic heritage is not only a remarkable endeavor but also a great challenge for researchers in this field.

In recent years, the field of image reconstruction is an active research area in computer vision which has made significant advancements driven by the rapid development of deep learning techniques [2]. The problem of reconstructing and repairing text images of historical Khmer documents arises from the need to recover and preserve the original content of these documents. This includes repairing damaged or missing parts and reconstructing lost information. Because many of these documents are written in a complex script known as the Khmer palm leaf script, the task of reconstructing text images is complicated and presents unique challenges in terms of character recognition and repair [1].

Valy et al. [6] highlighted the immense complexity of the Khmer language, which is renowned worldwide due to its extensive use of visually similar symbols in its alphabet. This inherent complexity presents a great challenge for researchers and historians who are eager to explore this invaluable cultural

* Corresponding author: Chanchen Pork
E-mail:chenpork.chen@gmail.com; Tel: +855-86 559 951

heritage. This complexity not only poses a great challenge for researchers and historians but also presents a hurdle in training machine learning models to accurately reconstruct the text from these images. Specialized methods and techniques need to be developed to overcome these challenges and effectively preserve these documents. Several previous investigations have explored different methodologies and datasets in the development of reconstruction models for text images in Khmer manuscripts. Raha et al. [7] focused on the restoration of historical document images using CNN. The objective is to denoise and restore old hand-writing documents, which often suffer from degradation, damage, or deterioration over time. They utilized an auto-encoder method to learn a compact representation of the historical document images. The encoder component of the network maps the input images into a lower-dimensional latent space, while the decoder component reconstructs the images from this latent representation. By incorporating the auto-encoder technique into their methodology, they can demonstrate the effectiveness of this approach for the restoration of historical document images. Isola et al. [4] introduced pix2pix or conditional GAN (cGAN) framework for image-to-image translation. It focuses on paired datasets and utilizes a generator network conditioned on input-output pairs to generate realistic output images. The framework is trained adversarially with a discriminator network to provide feedback on the generated images. The pix2pix/cGAN method achieves impressive results in various image translation tasks, such as converting sketches to images or transforming aerial maps to satellite images. It contributes to the advancement of GAN-based image translation techniques in the field of computer vision and image processing. Wang et al. [8] addressed the limitations of using Mean Square Error (MSE) and Peak Signal-to-Noise Ratio (PSNR) as single metrics for image quality evaluation. They introduced a new metric called Structural Similarity (SSIM) that incorporates local luminance, contrast, and structural information to better align with human perception of image quality. The authors highlighted the deficiencies of MSE and PSNR, emphasizing the importance of considering perceptual factors in image evaluation. Their work introduced SSIM as a valuable alternative that improves image quality assessment by capturing perceptual aspects beyond simple pixel-wise differences.

In this research, our focus revolves around three pivotal components of state-of-the-art methods in image reconstruction: CNN, GAN, and Evaluation Matrices including MSE, PSNR and SSIM. We delve into various techniques of data processing that have been tailored for data training, incorporating diverse combinations of training instances, epochs, and degradation levels for each method. The objective is to develop an optimized output model capable of achieving the highest score in image reconstruction tasks, utilizing the aforementioned evaluation matrices that allow us to quantitatively evaluate the extent of similarity between the generated images and the original image, aiding in the selection of the best-performing model.

2. METHODOLOGY

2.1 Dataset

In the context of machine learning, datasets function as the building blocks upon which algorithms and models are trained. They provide the necessary examples, allowing models to discern patterns, relationships, and features within the data. The significance of a well-constructed dataset cannot be overstated, as it directly influences the performance and accuracy of the resultant machine learning models. The SteukRith set, in particular, offers a window into the intricate world of Khmer palm leaf characters [1]. Each literal image within the dataset is a grayscale representation, measuring 48 by 48 pixels. This standardized format ensures consistency and enables to harness the potential of the dataset effectively.

Data augmentation is a critical technique in machine learning. It involves creating variations of original dataset to improve the generalization and robustness of the model. One common approach is data degradation, where synthetic "degraded" images are created from the original images. Fig. 1 displays three distinct variants of degraded datasets: random line erasing, random square erasing, and random ellipse erasing. These techniques introduce randomness and controlled modifications to the pixel values of specific areas within the images, simulating real-world scenarios of imperfect or incomplete data.

- Random Line Erasing - involves randomly selecting a starting point and an angle for a line that spans across the image. The pixel values along this line are modified to a certain predefined value (e.g., 255), effectively "erasing" the line from the image.
- Random Square Erasing - a random pair of coordinates defines the top-left and bottom-right corners of a square region within the image. All pixel values within this square region are changed to the predefined value. This technique imitates situations where portions of the image are obscured, blocked, or damaged by objects.
- Random Ellipse Erasing - similar to the square erasing technique, random coordinates determine the bounding box of an ellipse within the image. Pixel values within the ellipse are modified to the predefined value. This technique adds more complexity to the degradation, as ellipses simulate irregularly shaped occlusions or damage.

Each of these degradation techniques can be adjusted using three different levels of severity such as:

- Level 1 (low level) - the degradation might be subtle, affecting only a small portion of the image and introducing minor imperfections.
- Level 2 (medium level) - a larger area is degraded, and the changes to pixel values are more pronounced, resulting in more significant modifications to the image.
- Level 3 (high level) - the degradation is intense, covering a substantial portion of the image and making it challenging for the model to recognize the original content.

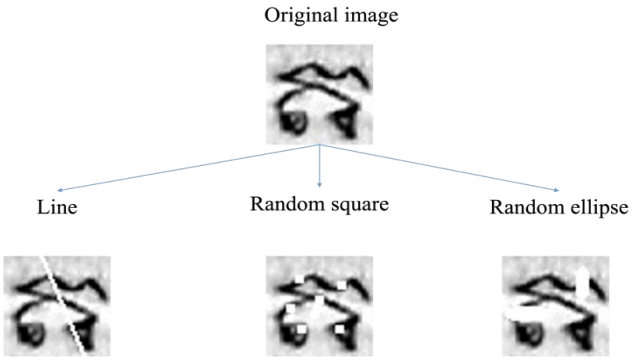


Fig. 1. Dataset degradation variants .

2.2 Methods

The first method, which is Convolutional Neural Networks (CNN) excels at feature extraction and object recognition, making them highly suitable for reconstructing historical document images and text. The architecture of CNN, consisting of convolutional, pooling, and fully connected layers, enables the learning of hierarchical features that can capture the intricacies of different characters and writing styles found in historical documents [3]. Fig. 2 demonstrated the designed architecture of CNN model called auto-encoder which consists of two networks encoder and decoder, each network has four convolving layers and two pooling layers. Encoder performs feature extraction by convolving input image with a set of 3x3 and 2x2 kernel sizes for max pooling to reduce spatial dimensions. Decoder performs reverse process by taking the input with shape (12, 12, 32) and performs a series of transposed convolutions to upsample the spatial dimensions until it reaches the final output shape of (48, 48, 1). Each transposed convolutional layer increases the spatial dimensions while reducing the number of channels (filters) in the output. This model will be trained and minimized the loss function using the Adam optimizer [5].

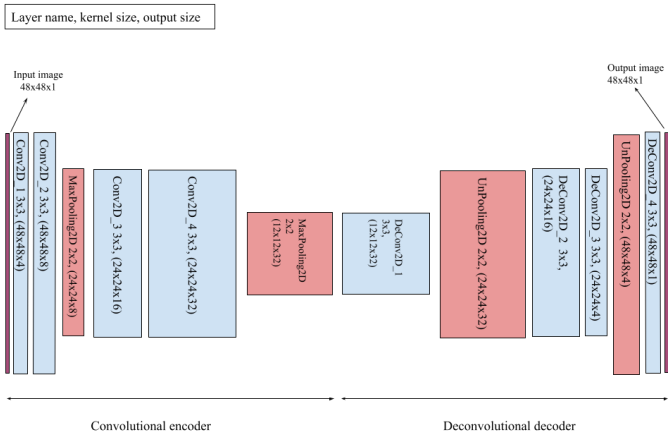


Fig. 2. Auto-encoder architecture of CNN.

The second method, GAN consists of two neural networks, a generator and a discriminator - which collaborate competently to generate realistic-looking images. The generator model produces synthetic images, while the discriminator distinguishes between real images from the dataset and fake images generated by the generator. This process leads to the generation of high-quality reconstructed images, making GAN an ideal choice for tasks such as inpainting and super-resolution in historical document restoration. The Pix2Pix model belongs to the category of conditional GAN (cGAN) [4], where the output image generation depends on a specific input, in this case it's a degraded image. The discriminator model undergoes direct training using both real and generated images, while the generator model does not (Fig. 4). Instead, the generator model is trained iteratively with the assistance of the discriminator model. Its training involves minimizing the loss predicted by the discriminator for generated images labeled as "real". Furthermore, the generator is updated to minimize the L1 loss [9] or mean absolute error between the generated image and the target image. Fig. 4 tells that generator model is defined by connecting the encoder and decoder components. The encoder takes an input image and progressively downsamples it to extract high-level features. It consists of multiple encoder blocks, each containing a convolutional layer, optional batch normalization, and a leaky ReLU activation function. The decoder takes the encoded representation and upsamples it to generate the output image. It consists of multiple decoder blocks, each containing a transposed convolutional layer, batch normalization, optional dropout, skip connections, and a ReLU activation function. Same as CNN, Adam optimizer algorithm is used to determines how the model will update its parameters during training to minimize the specified loss function. Likewise, binary cross-entropy is commonly used for binary classification tasks, which is appropriate for the discriminator's role [10].

2.3 Evaluation Matrices

The purpose of the evaluation is to measure the difference between the resulting image and the original image and evaluate the resulting image's visual quality. These matrices below are commonly used in terms to evaluate the quality of the model by scoring the output image to the original.

- Mean Squared Error (MSE) - calculates the average of the squared errors between the intensity or color values of each pixel in the reconstructed image and the corresponding pixel in the original image. A lower MSE value suggests a closer resemblance between the images.

$$MSE = \frac{\sum_{M,N}[I_1(m,n)-I_2(m,n)]^2}{M*N} \qquad \text{(Eq. 1)}$$

*Where M and N are the numbers of rows and columns in the input images.

- Peak Signal-to-Noise Ratio (PSNR) - helps overcome the limitation of MSE by incorporating a normalization factor and considering the ratio of signal power to noise power. PSNR is

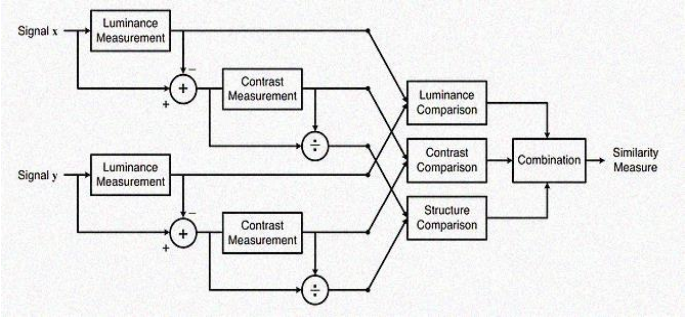


Fig. 3. SSIM-System.

expressed in decibels (dB) and provides a relative measure of image quality. Higher PSNR values indicate better quality.

$$PSNR = 10\log_{10}\left(\frac{R^2}{MSE}\right) \qquad \text{(Eq. 2)}$$

*Where R is the maximum fluctuation in the input image data type.

- Structural Similarity Index (SSIM) - Fig. 3 describes the SSIM-System it considers three main components: luminance, contrast and structure. The SSIM index ranges from 0 to 1, where a percentage value of similarity between the images. Higher SSIM values indicate greater similarity, while lower values suggest more noticeable differences.

2.4 Experimental Setup

Conducting experiments with the TensorFlow [11] environment on the dataset of 91,600 images, divided into 90,600 training images and 1,000 test images, provides a valuable opportunity to explore the impact of different training setups on the performance of both CNN and GAN models. For the CNN model, training for 200 epochs with a batch size of 100 per epoch offers the potential for deeper convergence and finer adjustments of the model's parameters. This extended training period can lead to improved accuracy and enhanced generalization on the test set. Differently, the GAN model is trained with only 2 epochs and 1 batch size per epoch due to hardware limitations. It is a prudent approach to ensure efficient use of available resources. Given the complexity of GAN and the computational intensity of adversarial training, it is common to adapt the training duration according to hardware capabilities. Each model weight is saved for every variant of degraded data type when training complete.

3. RESULTS AND DISCUSSION

Through our testing task, we have obtained compelling results that demonstrate the superior performance of Generative Adversarial Networks (GAN) compared to Convolutional Neural Networks (CNN) in image reconstruction. The output scores obtained from our evaluation as presented in Table 1. is the average value of all test images the score of comparison between the output image to its original based on each degrade level to all degrade variants. It provides strong evidence that GAN performs remarkably better than CNN across all metrics (PSNR, MSE, SSIM) for image reconstruction. This reinforces the notion that GAN is a highly effective and promising approach for this specific task, offering superior image reconstruction results with improved visual fidelity and accuracy compared to traditional CNN-based methods. Additionally, we have noticed that CNN is better at reconstructing images of recently trained degrade variant compared to the variant that were trained first.

The difference in performance between the CNN and GAN models can be attributed to their training setup and architectural variations also, the effectiveness of one-to-one mapping with a batch size of 1 might be emphasized. Significantly, the GAN's discriminator has a more sophisticated architecture these factors likely influenced the GAN's ability to produce better image reconstructions compared to the CNN as showing in Fig. 5. Further investigation is needed to explore the impact of different training configurations and discriminator architectures on the models' performance.

4. CONCLUSIONS

The research conducted in this study has unveiled highly promising outcomes within this domain, attributed to the adept incorporation of advanced deep learning methodologies. Through rigorous experimentation, these techniques have showcased their effectiveness in rectifying flawed images, reinstating absent or corrupted text, and elevating the global quality of historical documents - as exemplified by the outcomes of the present approach. By channeling these enhancements, encompassing optimized data processing, expanded training datasets, and the integration of supplementary techniques, our conviction is fortified that this approach holds the potential to attain even more precise and streamlined outcomes in the realm of text image reconstruction and restoration. In doing so, it stands poised to play a pivotal role in safeguarding invaluable cultural artifacts for generations to come.

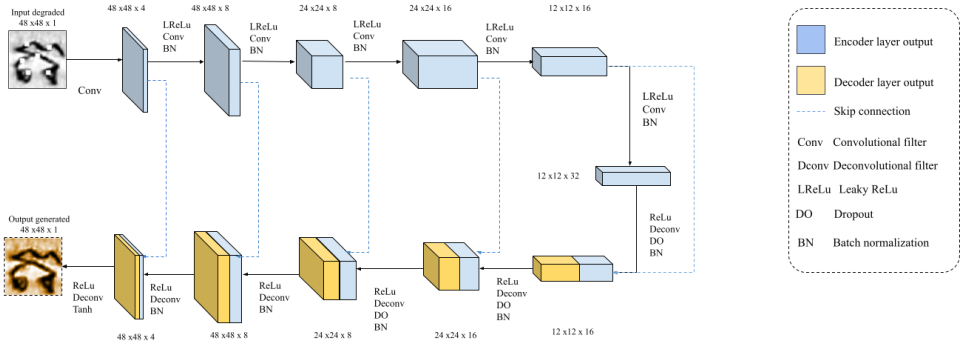
ACKNOWLEDGMENTS

This research study is supported by Cambodia Higher Education Improvement Project (Credit No. 6221-KH).

Table 1. Experiment result average scores of all degraded variants by each level and each model.

Degrade level	PSNR(dB)		MSE		SSIM(%)	
Model	CNN	GAN	CNN	GAN	CNN	GAN
1	22.97	**31.80**	369.24	**64.48**	89.34	**98.23**
2	22.11	**29.70**	449.29	**98.77**	87.82	**97.65**
3	21.53	**28.09**	509.61	**139.77**	86.02	**96.74**

(a)



(b)

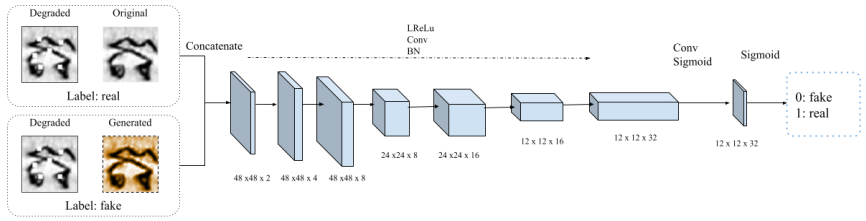


Fig.4. GAN architectures: (a) Generator model, (b) Discriminator model.

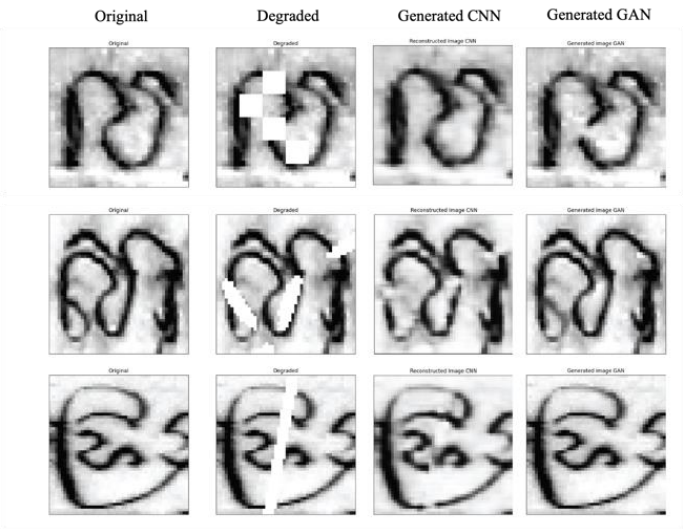


Fig. 5. Comparison of three simple images generated by both models. Each image represented a degrade variant of level 2.

REFERENCES

[1] Valy, D., Verleysen, M., Chhun, S. & Burie, J. C. (2017). A New Khmer Palm Leaf Manuscript Dataset for Document Analysis and Recognition: SleukRith Set. In 4th International Workshop on Historical Document Imaging and Processing. https://doi.org/10.1145/3151509.3151510

[2] Liu, Po-Yu, and Edmund Y. Lam. "Image reconstruction using deep learning" (2018) in arXiv preprint arXiv:1809.10410,2018.

[3] Zhang, K., Zuo, W., Gu, S., & Zhang, L. (2017). Learning Deep CNN Denoiser Prior for Image Restoration. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2017.300

[4] Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). Image-to-Image Translation with Conditional Adversarial Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2017.632

[5] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in arXiv preprint arXiv:1412.6980, 2014.

[6] Valy, D., Verleysen, M., Chhun, S., & Burie, J. C.(2018, August). Character and Text Recognition of Khmer Historical Palm Leaf Manuscripts. 2018 16th International Conference on Frontiers in Handwriting Recognition. (ICFHR). https://doi.org/10.1109/icfhr-2018.2018.00012

[7] Raha, P., & Chanda, B. (2019). Restoration of Historical Document Images Using Convolutional Neural Networks. 2019 IEEE Region 10 Symposium (TENSYMP). https://doi.org/10.1109/tensymp46218.2019.8971112

[8] Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image Quality Assessment: From Error Visibility to Structural Similarity. IEEE Transactions on Image Processing, 13(4), 600–612. https://doi.org/10.1109/tip.2003.819861

[9] Zhao, O. Gallo, I. Frosio and J. Kautz, "Loss Functions for Image Restoration With Neural Networks," in IEEE Transactions on Computational Imaging, vol. 3, no. 1, pp. 47-57, March 2017. doi: 10.1109/TCI.2016.2644865.

[10] Ruby, Usha & Yendapalli, Vamsidhar. (2020). Binary cross entropy with deep learning technique for Image classification. International Journal of Advanced Trends in Computer Science and Engineering. 9. 10.30534/ijatcse/2020/175942020.

[11] https://www.tensorflow.org/