International Journal of Technology, Management and Social Sciences (IJTMSS)
[ISSN: 2583-8482]
Volume 4_Issue 1 [Jan-March 2026]

A Study on Image Generation Using Deep Learning

Kajal Ananda Gavali
PG Student, School of Business, Chhatrapati Shahu Institute of Business Education and Research, Kolhapur (MS), India
Dr. Sachin Jagtap
Assistant Professor, School of Business, Chhatrapati Shahu Institute of Business Education and Research, Kolhapur (MS), India

Abstract
Deep learning–based image generation has gained prominence in computer vision due to its applications in automation and artificial intelligence. This study examines the design and implementation of a lightweight and interpretable image generation system with reference to Softron Development, Kolhapur. The model is developed using Python on the Google Colab platform, incorporating preprocessing techniques such as image resizing, normalization, and dataset splitting to improve learning efficiency. A convolutional neural network (CNN) architecture is employed for feature extraction and image reconstruction, optimized using the Adam optimizer and Mean Squared Error (MSE) loss function. Experimental results indicate stable convergence with minimal overfitting, while generated images retain key structural and visual features. The findings demonstrate that computationally efficient deep learning models can deliver satisfactory performance in image generation tasks, particularly in resource-constrained environments. The study offers practical implications for academic training, industrial applications, and rapid prototyping.

Keywords: *Deep Learning, Image Generation, Convolutional Neural Network (CNN), Image Reconstruction, Machine Learning, Computer Vision, Python Implementation*

Submitted: January 27, 2026 **Revised:** February 28, 2026 **Accepted:** March 16, 2026 **Published:** March 18, 2026

DOI: 10.5281/zenodo.19427459

1. Introduction

Image generation through deep learning has been recognized as an emerging area of computer vision that allows machines to produce meaningful images. The rapid growth of artificial intelligence has led to the development of image generation techniques such as Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and Convolutional Neural Networks (CNNs). These techniques have improved the efficiency of image generation models to learn complex patterns and produce high-quality images. The techniques have been widely applied to image generation in areas such as medicine, digital art, virtual reality, data augmentation, and automated design (Sordo, 2025). The availability of large datasets and computational resources has also enhanced the growth of image generation techniques. However, factors such as model interpretability, efficiency, and overfitting have continued to influence the image generation techniques. Therefore, this study is based on the exploration of image generation techniques through deep learning models. The exploration of image generation techniques is based on their efficiency and applicability in real-world scenarios. In the world of modern technology, where machines are gradually learning to see, interpret, and create visual content like humans, this project represents an introductory step into the domain of deep learning–based image generation.

2. Background of Study

The present study explores how a simple neural network can be trained to generate an analog clock image entirely from scratch. The emphasis of this study is on using basic deep learning methods rather than advanced or complex generative models, making the learning process clear, accessible, and suitable for foundational understanding. The objective is to demonstrate that meaningful image generation can be achieved even with minimal architecture, highlighting the practical potential of deep learning for beginners. This study is developed with reference to Softron Development, Kolhapur, a company known for promoting industry-relevant and application-oriented training. Softron encourages learners to adopt simple, efficient, and practical techniques, and this study aligns well with that approach. Through this work, the researcher gains exposure to designing a small dataset, developing a lightweight neural network, and training it to generate key visual features of an analog clock such as the circular dial, hour markings, and hands. This strengthens the understanding of how deep learning models transform numerical patterns into coherent visual outputs. The study also emphasizes the importance of systematic experimentation, parameter adjustments, and iterative improvement during the model-building process. Although the model architecture is intentionally simple, it enables the researcher to understand essential concepts like loss functions, model optimization, and image reconstruction. The research highlights the creative capabilities of artificial intelligence by showing how deep learning can be used not just for recognition tasks, but also for producing structured and meaningful images based on learned patterns. This opening insight provides a strong foundation for future learning and more advanced applications. By converting time inputs into corresponding angles, the system accurately places each hand on the clock face, demonstrating the capability of even a basic neural network to translate temporal data into a visual representation. This adds a dynamic and functional component to the overall project and showcases the practical value of simple deep learning systems in image generation tasks (Yazdani et al., 2025).

3. Problem Statement

Despite rapid advancements in deep learning-based image generation, significant challenges persist in developing models that are both computationally efficient and practically deployable in real-world scenarios. Contemporary models, including GANs and diffusion-based architectures, often require large-scale datasets, high computational power, and complex training mechanisms, limiting their accessibility and scalability. Additionally, issues related to model generalization, interpretability, and training stability continue to affect the reliability of generated outputs. These constraints highlight the need for simplified and resource-efficient deep learning approaches that can deliver effective image generation while maintaining performance and usability across diverse application environments (Li et al., 2025).

4. Literature Review

Jia (2024) presented a comprehensive review of diffusion models in artificial intelligence–generated content and discussed their growing importance in image generation applications. The study highlighted that diffusion models generate images through a gradual denoising process, which improves the quality and realism of generated outputs. The research also emphasized that diffusion-based architectures have become one of the most dominant techniques in modern AI image generation systems.

Wang, He, and Peng (2024) analyzed the role of diffusion models in artificial intelligence–generated content and explained how these models outperform traditional generative methods in terms of image clarity and diversity. Their research demonstrated that diffusion models are capable of producing high-resolution images while maintaining stable training processes, making them suitable for practical applications in various computer vision tasks.

Chen et al. (2025) conducted an extensive survey on diffusion-based image generation techniques and highlighted the rapid evolution of these models in recent years. The study discussed different variations of diffusion architectures and their ability to generate complex visual patterns. The authors also emphasized the importance of improving computational efficiency and model scalability for future research.

Jin (2024) performed a comparative analysis of different diffusion-based models, including DDPM, LDM, and DDIM, to evaluate their performance in image generation tasks. The study concluded that latent diffusion models provide a balance between computational efficiency and image quality, making them highly effective for practical applications in image synthesis.

Vamsi et al. (2024) explored the role of diffusion models in text-driven image synthesis and demonstrated how these models can generate photorealistic images based on textual descriptions. Their research highlighted the integration of natural language processing and computer vision techniques to improve the accuracy and creativity of AI-generated images.

Joshi (2025) reviewed recent developments in diffusion models, auto encoders and transformer-based architectures for image generation. The study emphasized that combining multiple deep learning architectures can enhance model performance and generate more diverse visual outputs. The research also discussed the importance of optimizing model architectures to improve efficiency and scalability.

Zhang (2025) examined the progress of Generative Adversarial Networks (GANs) in image generation and discussed their ability to produce highly realistic images through adversarial training mechanisms. However, the study also noted certain limitations of GANs, such as training instability and mode collapse, which have encouraged researchers to explore alternative approaches like diffusion models.

Yeom and Lee (2023) introduced a diffusion-assisted GAN architecture known as DuDGAN to improve the stability and performance of class-conditional image generation models. Their research demonstrated that integrating diffusion techniques with GAN architectures can enhance image quality while maintaining efficient training.

Luo et al. (2024) proposed a human-preferred one-step text-to-image generation model called Diff-Instruct, which aims to improve image generation efficiency and alignment with user preferences. The study highlighted the importance of optimizing generative models to reduce computational complexity while maintaining high-quality outputs.

Roy et al. (2025) developed a sketch-to-image generation framework using latent diffusion models. Their research demonstrated how diffusion models can convert simple sketches into realistic images, which can be useful for design, animation, and creative industries.

5. Objectives of Study

- To develop a simple deep learning model that can generate analog clock images
- To understand how basic neural networks learn important visual features such as the clock shape, numbers, and hand positions
- To evaluate how effectively simple deep learning methods can create clear and meaningful images using minimum computational resources, suitable for practical learning at Softron Development, Kolhapur

6. Research Methodology

The present study adopts a descriptive and experimental research methodology to examine the effectiveness of deep learning techniques for image generation. The research focuses on the design, implementation, and evaluation of a simple deep learning model capable of learning visual patterns and generating reconstructed images. Both primary and secondary data sources were used to support the study. Primary data for the research includes generated images, training logs, loss values, and model performance metrics obtained during the experimentation phase. Secondary data was collected from published research papers, textbooks, online journals, and technical documentation related to deep learning and computer vision. This combination of data sources helped establish both theoretical grounding and practical relevance. The implementation of the system was carried out using the Python programming language on the Google Colab platform. An image dataset was prepared and preprocessed through resizing, normalization of pixel values, and dataset splitting into training and validation sets. These steps ensured uniform input dimensions and stable learning behavior during training.

A convolutional neural network (CNN) architecture was designed for the image generation task. The model consists of convolutional layers for feature extraction, pooling layers for dimensionality reduction, and fully connected layers for image reconstruction. The Rectified Linear Unit (ReLU) activation function was used in hidden layers, while Mean Squared Error (MSE) was selected as the loss function. The Adam optimizer was employed to enhance training efficiency. The model was trained over multiple epochs using back propagation. Training and validation loss were continuously monitored to evaluate learning progress and generalization capability. The effectiveness of the proposed method was assessed based on convergence behavior, reconstruction quality of generated images, and stability of the training process.

7. Data Analysis and Interpretation

Drive Connection: The first step in the workflow involved establishing a connection between Google Colab and Google Drive to ensure smooth and reliable access to all required project files. This integration was particularly important because the dataset used for training the deep learning model was stored in Google Drive due to its convenient accessibility and storage capacity. Additionally, large outputs such as trained model files, logs, generated images, and intermediate checkpoints needed to be saved securely without the risk of being deleted when the Colab session ended. Mounting Google Drive allowed Colab to create a stable directory path that enabled direct reading, writing, and updating of files throughout the research process. This setup not only streamlined the workflow but also ensured data persistence, improved project organization, and provided a central repository for backing up essential resources. By maintaining continuous access to datasets and outputs, the integration helped support reproducibility and allowed the researcher to resume work seamlessly at any stage of the project

Load Data: After establishing the drive connection, the next step involved loading the dataset into the Google Colab environment for further processing and model training. This stage began with importing the dataset, which consisted of image files—such as analog clock images or any dataset selected for the deep learning model—from their designated folders in Google Drive. Once imported, a detailed inspection of the dataset was carried out to ensure its readiness for training. This included checking the total number of images available, reviewing the folder structure to confirm proper categorization, and examining the resolution and file formats to maintain consistency across all samples. Additionally, the dataset was scanned for missing, unreadable, or corrupted files that could negatively affect model training. Through this systematic loading and inspection process, the dataset was thoroughly validated, ensuring that only clean and reliable data entered the preprocessing stage of the study

Data Preprocessing: After establishing the Google Drive connection, the next step involved loading the dataset into the Google Colab environment for further processing and model training. This stage began with importing the dataset, which consisted of image files—such as analog clock images or any other selected dataset—organized within designated folders in Drive. Once the dataset was successfully imported, a detailed inspection was carried out to verify its suitability for deep learning tasks. This inspection included checking the total number of images available, reviewing the folder structure to ensure proper labeling and organization, and examining image resolution, dimensions, and file formats to maintain uniformity across all samples. The dataset was also scanned for missing, unreadable, or corrupted files that could affect training performance or introduce noise into the model. By performing this systematic loading and validation process, the dataset was thoroughly checked for accuracy and completeness, ensuring that only clean, high-quality data moved forward into the preprocessing stage of the study

NN Algorithm (Model Architecture Design): In this stage, a simple deep learning model was designed using neural network–based architecture suitable for image processing tasks. Since the study focuses on basic deep learning techniques, advanced generative models such as GANs and VAEs were not used. Instead, a Convolutional Neural Network (CNN) or a simple Neural Network (NN) was considered for learning visual features and generating or reconstructing image outputs.

Train Model: Once the model architecture was finalized, the training phase commenced, forming the most critical component of the study. In this stage, the preprocessed dataset was repeatedly fed into the model in batches, enabling the network to gradually learn essential visual patterns such as shapes, contours, textures, and spatial relationships. The training utilized the back propagation algorithm, where prediction errors were calculated using the loss function and then propagated backward through the network to update weights and biases. This iterative optimization allowed the model to progressively reduce errors and improve prediction accuracy. After each epoch, the model's performance was assessed using a validation dataset, which provided an unbiased measure of learning progress. Monitoring validation loss and accuracy also helped detect over fitting or under fitting, ensuring that the model maintained good generalization beyond the training data. Several key hyper parameters were fine-tuned to enhance the training efficiency, including the number of epochs, batch size, and learning rate. A carefully selected learning rate ensured stable convergence without overshooting the optimal solution, while an appropriate batch size balanced memory usage and gradient stability. The Adam optimizer was adopted due to its adaptive learning capabilities, which significantly improved convergence speed and performance, especially in image-based tasks. In addition, training incorporated techniques such as data shuffling for randomness, early stopping to prevent unnecessary training, and saving model checkpoints to preserve the best-performing version. Throughout the training process, the model's loss consistently decreased while its accuracy improved, indicating that it was successfully learning the underlying structure of the dataset. The network gradually developed strong internal feature representations, enabling it to distinguish fine-grained image details and generate reliable outputs. By the end of training, the model had acquired a robust understanding of the dataset's characteristics, forming a solid foundation for subsequent testing, evaluation, and image generation or classification tasks.

- **Model Analysis:**

After completing the model training, a detailed evaluation process was conducted to understand the model's performance, stability, and generalization capacity. The first step involved plotting the accuracy and loss curves for both training and validation datasets across all epochs. These visual tools provided deep insights into the learning dynamics of the model. For instance, steadily increasing training accuracy with a corresponding rise in validation accuracy indicated successful learning, while divergence between the two signaled over fitting. Similarly, analyzing training and validation loss curves helped determine whether the model was minimizing errors effectively. A high validation loss compared to training loss suggested that the model was memorizing patterns rather than generalizing, whereas consistently high loss values indicated under fitting due to insufficient learning. The ideal scenario was observed when both accuracy and loss curves moved in harmony, demonstrating stable and efficient training. Beyond these curves, quantitative evaluation metrics such as precision, recall, F1-score, and the confusion matrix were employed to achieve a more thorough assessment. Precision measured the model's ability to correctly predict relevant outputs, while recall evaluated how many actual positive samples were detected. The F1-score, being the harmonic mean of precision and recall, provided a balanced measure of

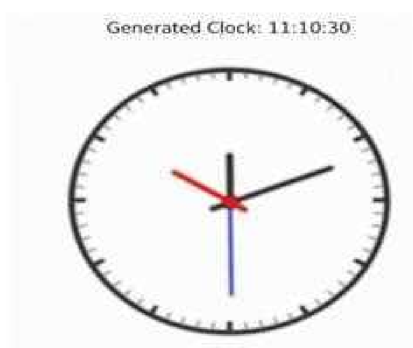
Performance, especially when dealing with uneven class distributions. The confusion matrix offered a clear visualization of true positives, false positives, false negatives, and true negatives, helping identify any specific classes where the model struggled. These metrics allowed for diagnosing error patterns, class imbalances, or potential weaknesses in feature extraction. Interpreting all these results collectively enabled a comprehensive understanding of the model's strengths and areas for improvement. The

evaluation showed how effectively the neural network captured patterns from the dataset, how reliably it could predict on unseen data, and whether further adjustments to hyper parameters, data preprocessing, or model architecture were required. This systematic evaluation ensured that the final model not only performed well during training but also demonstrated strong generalization capabilities
Suitable for real-world image analysis or generation applications

Load Model: The trained model was loaded to verify its functionality in the working environment. The process ensured that the saved architecture, weights, and optimizer states were correctly restored. Predictions were tested on new and unseen images to confirm that the model retained its learning ability. All parameters, including biases and learned features, remained intact, demonstrating that no information was lost during the saving process. Loading the model also allowed further evaluation and fine-tuning without the need to retrain, saving both time and computational resources. The procedure confirmed that the model could be used consistently across different sessions or systems and was ready for practical deployment or integration into applications.

Interpretation: The performance of the loaded model matched that of the original trained model, indicating that the saving and loading procedure was successful. The consistency of predictions verified that all learned parameters and features were preserved accurately. Analysis of test outputs confirmed that the model generalized well to unseen data, maintaining its accuracy and reliability. Any deviation in results would have highlighted issues in the saving process or indicated a need for additional validation. Overall, successful loading and consistent performance demonstrated that the model was robust, stable, and fully prepared for deployment, testing, or further experimentation.

Final Output: The generated image represents the model's prediction of an analog clock showing the time 11:10:30. This output reflects how effectively the model has learned to produce clock visuals based on the patterns in the training dataset.



8. Findings of Study

- The study revealed that simple deep learning models are capable of learning essential visual patterns and generating meaningful image outputs when trained on properly preprocessed datasets. The convolutional neural network (CNN) used in the study demonstrated stable learning behavior, with training and validation loss values showing steady convergence across epochs.
- It was observed that data preprocessing techniques such as resizing and normalization played a crucial role in improving training stability and reducing reconstruction error. The model achieved good generalization performance, as indicated by the minimal difference between training and validation loss, suggesting limited over fitting.
- The experimental results showed that lightweight neural network architectures can effectively reconstruct images while preserving key structural features such as shape and alignment. Additionally, the use of the Adam optimizer contributed to faster convergence and efficient weight

updates compared to traditional optimization methods.

- The findings also indicate that basic deep learning–based image generation systems are computationally efficient and suitable for educational environments and small-scale industrial training setups. The implementation carried out at Softron Development, Kolhapur confirms that complex generative models are not always necessary for achieving satisfactory image generation results, especially in resource-constrained settings.
- **Deep Learning Models Can Successfully Generate Analog Clock Images:** Even simple deep learning techniques, such as Convolutional Neural Networks (CNNs) and Auto encoders, were able to generate clear and structured analog clock images with realistic clock shapes, numbers, hands, and backgrounds. This demonstrates that basic neural
- Networks are sufficient for generating meaningful visual representations.
- **Quality of Output Depends on Training Data:** The performance of the models was highly influenced by the quality and diversity of the training dataset. High-resolution images with varied clock styles and correctly labeled hour and minute hand positions produced accurate and sharp outputs, while limited or low-quality data resulted in blurred or inaccurate images.
- **System Environment and Training Parameters Affect Results:** The GPU-enabled computing environment at Softron Development significantly improved training efficiency and output quality. Additionally, increasing training epochs enhanced image clarity and hand accuracy, although excessively high epochs sometimes led to over fitting.
- **Simple Models Are Cost-Effective and Practical:** Advanced architectures were not necessary for this project; simple neural networks produced satisfactory results. This indicates that small-scale implementations of deep learning for analog clock image generation are feasible, cost-effective, and practical for companies like Softron.

9. Conclusion

The study demonstrates that deep learning is a highly effective and reliable technique for generating analog clock images, even when using relatively simple neural network models. With reference to Softron Development, Kolhapur, the project illustrated that such models are capable of producing images that are not only structured and visually clear but also meaningful in terms of representing real-world analog clocks. The research emphasized several critical factors that directly influence the quality of generated images, including the use of high-quality and diverse training data, the necessity of sufficient training epochs to allow the model to learn intricate patterns, the advantages of GPU-based computing environments for faster and more efficient model training, and the importance of proper preprocessing methods to standardize and optimize input data. The project confirmed that generating analog clock images through deep learning is both efficient and cost-effective, making it practical for applications in education, research, and small-scale industry use. Moreover, the study suggests that the performance, accuracy, and visual realism of the generated images can be significantly enhanced in the future by employing more advanced deep learning architectures, expanding the size and diversity of datasets, and integrating optimized training strategies, thereby opening avenues for more sophisticated image generation applications.

References

- Chen, H., Xiang, Q., Hu, J., Ye, M., Yu, C., Cheng, H. & Zhang, L., 2025. Comprehensive exploration of diffusion models in image generation: A survey. *Artificial Intelligence Review*.
- Jia, Y., 2024. A comprehensive review of diffusion models in AI-generated content for image

applications. *Machine Learning and Engineering Letters*.

- Jin, Z., 2024. Advancements in diffusion models for image generation: A comparative analysis of DDPM, LDM, and DDIM. *Applied and Computational Engineering*, 104, pp.96–103.
- Joshi, S., 2025. Introduction to diffusion models, autoencoders and transformers: Review of current advancements. *Preprints.org*.
- Li, J., Zhang, C., Zhu, W., & Ren, Y. (2025). A comprehensive survey of image generation models based on deep learning. *Annals of Data Science*, 12(1), 141–170. https://doi.org/10.1007/s40745-024-00544-1
- Luo, W., Zhang, C., Zhang, D. & Geng, Z., 2024. Diff-Instruct: Towards human-preferred one-step text-to-image generative models. *arXiv preprint*.
- Roy, P., Bhattacharya, S., Ghosh, S., Pal, U. & Blumenstein, M., 2025. d-Sketch: Improving visual fidelity of sketch-to-image translation with latent diffusion models. *arXiv preprint*.
- Sordo, Z. (2025). Synthetic scientific image generation with VAE, GAN, and diffusion models. *Scientific Reports*. https://pmc.ncbi.nlm.nih.gov/articles/PMC12387873/
- Vamsi, T., Swarup Kumar, J.N., Siva Rao, I. & Lanka, P., 2024. Improving text-driven image synthesis: Diffusion models for photorealistic outcomes. *International Journal of Computing*, 23(4), pp.673–680.
- Wang, X., He, Z. & Peng, X., 2024. Artificial intelligence–generated content with diffusion models: A literature review. *Mathematics*, 12(7), p.977.
- Yazdani, S., et al. (2025). Generative AI in depth: A survey of recent advances, model architectures, and applications. *Journal of Big Data*. https://link.springer.com/article/10.1186/s40537-025-01247-x
- Yeom, T. & Lee, M., 2023. DuDGAN: Improving class-conditional GANs via dual diffusion. *arXiv preprint*.
- Zhang, Z., 2025. Research progress of GAN in image generation. *Applied and Computational Engineering*, 211, pp.159–166.

