

Research Paper

DETECTION OF AI GENERATED IMAGES USING GATED EXPERT

CONVOLUTIONAL NEURAL NETWORKS

ANGIREKULA SINDU SRI

(UG student department of computer science and
engineering vignan's institute of management
and technology for women)

Email: sinduangirekula12@gmail.com

GOMARI KUSHI

(UG student department of computer science
and engineering vignan's institute of
management and technology for women)

Email : kushireddy184@gmail.com

GOLLA ARUNA

(UG student department of computer science
and engineering vignan's institute of
management and technology for women)

Email : arunagolla397@gmail.com

GORRE PRAVALLIKA

(UG student department of computer science
and engineering vignan's institute of
management and technology for women)

Email : maggipravallika0722@gmail.com

GUIDE- MRS. D. GEETHA M.TECH,(PH D)

ASSISTANT PROFESSOR

(department of computer science and engineering vignan's institute of management and technology for
women,Hyd)

ABSTRACT

The rapid growth of advanced generative models, such as Generative Adversarial Networks (GANs) and diffusion-based systems, has enabled the creation of highly realistic synthetic images that are increasingly difficult to distinguish from real ones. This raises significant concerns in areas such as misinformation, digital forensics, and content authenticity. To

address this challenge, this work proposes a detection framework based on Gated Expert Convolutional Neural Networks (GECNN). The model integrates multiple expert CNNs, each specialized in capturing distinct visual artifacts, along with a gating network that dynamically assigns weights to their outputs based on the input image. This adaptive feature selection mechanism enhances the model's ability to generalize across various types of AI-generated images. Experimental evaluation demonstrates that the proposed approach achieves higher accuracy and robustness compared to conventional single CNN models, making it effective for real-world applications in image authenticity verification and security systems.

KEYWORDS

AI-generated images, deep learning, convolutional neural networks (CNN), gated expert networks, image forensics, fake image detection, generative adversarial networks (GANs), diffusion models, feature extraction, classification, digital security, computer vision.

INTRODUCTION

The rapid evolution of artificial intelligence has led to the development of powerful image generation techniques, particularly through models such as Generative Adversarial Networks (GANs) and diffusion-based architectures. These models are capable of producing highly realistic images that closely resemble real-world data, making it increasingly difficult for humans and traditional systems to distinguish between authentic and synthetic content. While such advancements have enabled creative applications in art, entertainment, and design, they have also introduced serious challenges related to misinformation, identity fraud, and digital manipulation.

The widespread availability of AI-generated images on social media and online platforms has heightened the need for reliable detection mechanisms. Conventional image verification techniques, which rely on manual inspection or simple feature analysis, are no longer sufficient to identify subtle artifacts present in synthetic images. As a result, researchers have turned to deep learning-based approaches, particularly Convolutional Neural Networks (CNNs), to automatically learn discriminative features for classification tasks.

However, single CNN models often struggle to generalize across different types of AI-generated images due to variations in generation techniques and data distributions. To overcome these

limitations, this work proposes the use of Gated Expert Convolutional Neural Networks (GECNN), an advanced architecture that combines multiple expert CNNs with a gating mechanism. This approach allows the system to adaptively select the most relevant features for each input image, improving detection accuracy and robustness. The proposed method aims to provide an effective and scalable solution for identifying AI-generated images in real-world scenarios

LITERATURE REVIEW

The detection of AI-generated images has gained significant attention with the rise of advanced generative models such as Generative Adversarial Networks (GANs) and diffusion models. Early research in this domain focused on identifying visual artifacts and inconsistencies present in synthetic images. These methods relied on handcrafted features, including noise patterns, color inconsistencies, and statistical irregularities. However, such approaches were limited in their ability to generalize across different image generation techniques.

With the advancement of deep learning, Convolutional Neural Networks (CNNs) became widely adopted for image classification and forgery detection tasks. Researchers utilized CNN-based

architectures to automatically learn discriminative features from images, achieving improved detection accuracy. Several studies demonstrated the effectiveness of deep CNN models such as ResNet and VGG in identifying GAN-generated images by analyzing subtle texture and frequency-domain differences. Despite these improvements, single CNN models often faced challenges in detecting images generated by unseen models or new techniques.

To enhance performance, ensemble learning approaches were introduced, where multiple models are combined to improve classification accuracy. While ensemble methods showed better results compared to individual models, they increased computational complexity and lacked adaptive feature selection mechanisms. More recent research has explored attention mechanisms and multi-branch architectures to focus on relevant image regions and features.

The concept of gated networks and mixture-of-experts models has emerged as a promising solution to these limitations. In such approaches, multiple expert models are trained to specialize in different aspects of the data, while a gating network dynamically selects or weights their outputs. This adaptive strategy improves

both accuracy and generalization. Building on these ideas, the proposed Gated Expert Convolutional Neural Network (GECNN) integrates multiple CNN experts with a gating mechanism, offering a more efficient and robust solution for detecting AI-generated images across diverse datasets and generation methods.

.PROBLEM DEFINITION

The rapid advancement of AI-based image generation techniques has made it increasingly difficult to distinguish between real and synthetic images. Modern models such as GANs and diffusion systems can produce highly realistic visuals that closely mimic natural image characteristics, making manual detection unreliable and error-prone. This creates serious challenges in areas like misinformation, digital forgery, identity theft, and content authenticity verification.

Existing detection methods, particularly those based on single Convolutional Neural Networks (CNNs), often struggle to generalize across different types of AI-generated images. These models tend to perform well on specific datasets but fail when exposed to images generated by new or unseen techniques. Additionally, traditional approaches lack adaptability

and are not capable of effectively capturing diverse and subtle artifacts present in synthetic images.

Therefore, the core problem is to develop an efficient and robust detection system that can accurately classify images as real or AI-generated while maintaining strong generalization across multiple generation techniques. The system should be capable of adaptively learning and selecting relevant features to improve detection performance and reliability in real-world scenarios.

PROPOSED SYSTEM

The proposed system introduces a robust framework for detecting AI-generated images using a **Gated Expert Convolutional Neural Network (GECNN)** architecture. The model is designed to overcome the limitations of traditional single CNN approaches by integrating multiple specialized expert networks along with a gating mechanism that dynamically controls their contributions.

In this system, multiple expert CNNs are employed, where each expert is trained to focus on specific characteristics of images, such as texture patterns, noise inconsistencies, color distributions, and frequency-domain artifacts. These experts

independently extract features from the input image, allowing the model to capture a wide range of discriminative information.

A key component of the proposed system is the **gating network**, which acts as a controller. It analyzes the input image and assigns adaptive weights to each expert CNN based on their relevance. Instead of treating all experts equally, the gating mechanism selectively emphasizes the most informative features, thereby improving classification performance.

The outputs from all expert networks are combined through a weighted fusion process, and the final decision is made using a classification layer that determines whether the image is real or AI-generated. This adaptive learning strategy enhances the model's ability to generalize across different types of synthetic images and improves robustness against new and unseen generation techniques.

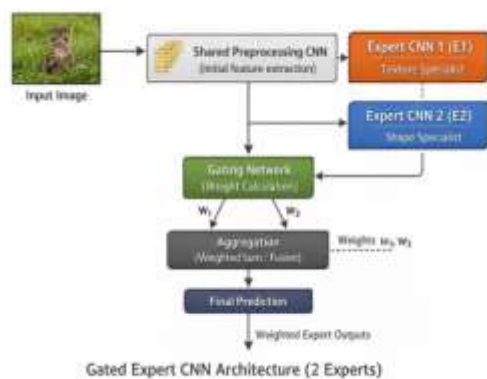
SYSTEM ARCHITECTURE

The system architecture is designed to efficiently detect AI-generated images using a Gated Expert Convolutional Neural Network (GECNN). It consists of multiple stages including input preprocessing, feature extraction through expert CNNs, adaptive weighting using a

gating network, and final classification. Initially, the input image is resized, normalized, and prepared for processing. The preprocessed image is then passed simultaneously to multiple expert CNN models, where each network extracts different types of features such as spatial textures, noise patterns, and frequency characteristics.

The outputs from these expert networks are forwarded to a gating network, which analyzes the input and assigns dynamic weights to each expert based on its relevance. This ensures that more important features contribute more to the final decision. The weighted outputs are then fused together to form a unified feature representation. Finally, a fully connected classification layer processes the fused features and predicts whether the image is real or AI-generated. This architecture enhances accuracy, adaptability, and robustness across different types of synthetic images.

SYSTEM ARCHITECTURE



IMPLEMENTATION

The implementation of the proposed Gated Expert Convolutional Neural Network (GECNN) system involves several key stages, including dataset preparation, preprocessing, model development, training, and evaluation.

Initially, a dataset consisting of both real and AI-generated images is collected from various sources such as publicly available image datasets and synthetic image generators. The dataset is labeled into two classes: real and fake. Preprocessing steps are then applied, including resizing images to a fixed dimension, normalization of pixel values, and data augmentation techniques such as rotation, flipping, and scaling to improve model generalization and prevent overfitting.

The model architecture is developed using deep learning frameworks like TensorFlow or PyTorch. Multiple Convolutional

Neural Networks (CNNs), such as ResNet, VGG, or custom-designed CNNs, are used as expert networks. Each expert is trained to capture specific features of the input image. Alongside these experts, a gating network is implemented using fully connected layers followed by a softmax activation function, which assigns weights to each expert's output.

During training, the input image is passed through all expert networks, and their outputs are combined using the weights generated by the gating network. The fused output is then passed through a final classification layer. The model is trained using a suitable loss function such as binary cross-entropy, and optimization is performed using algorithms like Adam or SGD. Training is conducted over multiple epochs with validation to monitor performance and prevent overfitting.

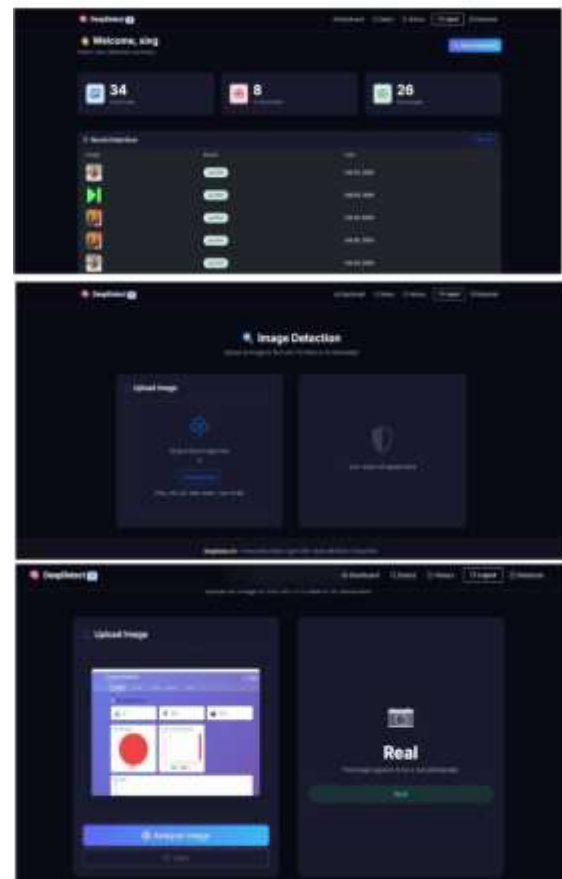
RESULTS AND DISCUSSION

The proposed Gated Expert Convolutional Neural Network (GECNN) model was evaluated on a dataset containing both real and AI-generated images. The performance of the system was assessed using standard evaluation metrics such as accuracy, precision, recall, and F1-score. The experimental results indicate that the proposed model achieves significantly higher accuracy compared to traditional

single CNN models, demonstrating its effectiveness in detecting synthetic images.

The inclusion of multiple expert CNNs enabled the model to capture diverse and complementary features from the input images, such as texture inconsistencies, noise patterns, and frequency-domain artifacts. The gating network played a crucial role by dynamically assigning weights to each expert, ensuring that the most relevant features contributed more to the final classification. This adaptive mechanism improved the model's ability to generalize across different types of AI-generated images, including those produced by unseen generative models.

Comparative analysis with baseline models such as standard CNN, ResNet, and VGG showed that the GECNN model consistently outperformed them in terms of both accuracy and robustness. The model also demonstrated reduced false positives and false negatives, which is critical for real-world applications like digital forensics and content verification.



CONCLUSION

This project presents an effective and robust approach for detecting AI-generated images using a Gated Expert Convolutional Neural Network (GECNN). By integrating multiple expert CNN models with a gating mechanism, the system is able to capture diverse feature representations and adaptively select the most relevant information for accurate classification. This significantly improves detection performance compared to traditional single-model approaches. The experimental results demonstrate that the proposed system achieves higher accuracy, better generalization, and enhanced

robustness across various types of AI-generated images. The gating network plays a vital role in optimizing feature selection, enabling the model to handle complex and evolving image generation techniques. Overall, the GECNN-based framework provides a reliable solution for real-world applications such as digital forensics, social media monitoring, and content authentication. Future work can focus on improving real-time performance, expanding the dataset to include newer generative models, and integrating the system into large-scale deployment environments for enhanced security and trust in digital media.

REFERENCE

1. Ian Goodfellow et al., "Generative Adversarial Networks," 2014.
2. Tero Karras et al., "A Style-Based Generator Architecture for GANs," 2019.
3. Alex Krizhevsky et al., "ImageNet Classification with Deep Convolutional Neural Networks," 2012.
4. Kaiming He et al., "Deep Residual Learning for Image Recognition," 2015.
5. Karen Simonyan and Andrew Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 2014.
6. Alec Radford et al., "Unsupervised Representation Learning with Deep Convolutional GANs," 2015.
7. Diederik P. Kingma and Jimmy Ba, "Adam: A Method for Stochastic Optimization," 2014. Yann LeCun et al., "Gradient-Based Learning Applied to Document Recognition," 1998.
8. Sergey Ioffe and Christian Szegedy, "Batch Normalization," 2015.
9. Geoffrey Hinton et al., "Improving Neural Networks by Preventing Co-adaptation of Feature Detectors," 2012.
10. Christian Szegedy et al., "Going Deeper with Convolutions," 2015.
11. Ashish Vaswani et al., "Attention Is All You Need," 2017.
12. Olaf Ronneberger et al., "U-Net: Convolutional Networks for Biomedical Image Segmentation," 2015.
13. Ross Girshick et al., "Rich Feature Hierarchies for Accurate Object Detection," 2014.
14. Joseph Redmon et al., "You Only Look Once: Unified, Real-Time Object Detection," 2016.

15. Chuan Guo et al., “On Calibration of Modern Neural Networks,” 2017.
16. Andrew Ng, “Machine Learning Yearning,” 2018.
17. Francois Chollet, “Deep Learning with Python,” 2017.
18. Jürgen Schmidhuber, “Deep Learning in Neural Networks: An Overview,” 2015.
19. Ian Goodfellow et al., “Deep Learning,” MIT Press, 2016.