



A CNN-Based Framework for Enhancing Image Segmentation and Scene Interpretation in Fake Colorized Images: Addressing Challenges in Image Analysis and Understanding



Hayder H.O. Al-Ghanimi¹, Amir Lakizadeh^{2*}

¹ Department of Medical Instrumentation Techniques Engineering, University of Hilla, Babylon 51001, Iraq

² Computer Engineering and Information Technology Department, University of Qom, Qom 3716146611, Iran

Corresponding Author Email: lakizadeh@qom.ac.ir

Copyright: ©2025 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.300223>

ABSTRACT

Received: 14 January 2025

Revised: 12 February 2025

Accepted: 24 February 2025

Available online: 27 February 2025

Keywords:

fake colorized images, Convolutional Neural Network (CNN), image manipulation detection, feature extraction, attention mechanisms, digital forensics

Fake colorized image detection has become a vital challenge in this digital content manipulation era, which highly affects the security, authenticity, and trustworthiness of visual data. Most of the current methods fail to provide proper identification of fake images in cases of complex textures and minute color variations. We present an advanced end-to-end CNN-based framework to effectively address these challenges. In this way, our approach combines multiscale feature extraction with an attention mechanism, which can make the model sensitive to local fine-grained features to improve its ability of distinguishing between real and fake images. Finally, the proposed method was compared to two state-of-the-art models, PanColorGAN and CCAR-TAR, in different scenarios. These have shown significant improvements in terms of accuracy and robustness; the proposed framework significantly reduced errors in challenging tasks such as classification. In particular, the model performed better for cases that were difficult to handle with traditional methods: detecting intricate textures and minor color discrepancies. All these hints at the fact that such a novel framework enhances accuracy in the detection of fake images but also presents one of the most reliable, adaptive solutions for a host of image analysis applications. Thus, this work may constitute a useful contribution to developing effective ways of ensuring veracity and authenticity in, among other areas, the digital forensics segment, content verification, media authenticity, and will presumably represent an interesting direction in prospective research and development regarding ways to fight digital manipulation.

1. INTRODUCTION

While image manipulation is an easy thing in the modern digital age—due to the advancement in sophisticated image editing tools and techniques—the authenticity of an image being shared has become one of the widely questioned aspects nowadays across all platforms [1]. The generation of fake colorized images—where grayscale images are colorized with misleading or inaccurate colors—stands among such manipulations. Such images can distort historical records, propagate misinformation, and deceive viewers, and their detection calls for a robust method of detection [2].

The challenge in this area constitutes the fact that, up to this date, the current state of computer vision and deep learning cannot efficiently detect forged colorized images accurately [3]. Traditional methods usually focus on heuristic-based methods and simple feature extraction, hence not able to give expected results after complex textures or subtle changes in color hues with sophisticated color manipulation techniques [4]. Furthermore, the existing deep models on image manipulation detection such as PanColorGAN [5] and CCAR-TAR [6], although powerful, are not able to distinctly make out the authentic and manipulated images explicitly in some delicate scenes. This indicates that critical scientific gaps exist

in current studies, which this research has wished and is going to deal with.

Not much can be said regarding the detection of reliable methods for fake colorization of images. Since both these aspects—digitally modified contents and historical records—may have a considerable number of shares to shape or restore public opinions, the task of maintaining integrity with regards to veracity in visual data is crucial. Most of the conventional detection techniques suffer from higher false positives and false negatives in some challenging situations; such incidents may even question the authenticity verification process in digital content [7]. This requires the development of more sophisticated and accurate models that can handle the challenges posed by modern image manipulations [8].

A crucial technique helping to detect manipulated images is color discontinuity detection, practically for fake colorized images that exhibit the unnatural transitions in color gradients characteristic of artificial manipulations. Sudden changes in color, inconsistent distributions of hue, and unnatural saturation levels are indicators of algorithm-based colorization that can usually be distinguished from genuine photographic blending [9].

Edge-based segmentation, color gradient analysis at multiple scales, and chromatic difference modeling are but a

few methods explored in several studies for the detection of chromatic inconsistencies. All three searches will uncover those areas in manipulated images where unnatural transitions occur, thus making them extremely relevant to fake colorized image detection [10].

To the extent of proving able to add charms to the detection of image forgery, explorations are on to integrate color discontinuity detection with deep learning models. Bringing chromatic edge techniques into CNN-based architectures has improved classification accuracy by accentuating unnatural color transition detection [11].

Incorporating color discontinuity detection within fake image analysis strengthens their firepower to look for fine inconsistencies which may not be very clearly visible. Processing the abrupt color transitions and unnatural blending patterns enhances the strength of fake colorized image detection models [12].

The paper bridges the gap by proposing a unified CNN framework that effectively leverages multiscale feature extraction and attention mechanisms, making for improved detection performance in fake colorized images. By underlining the model's fine-grained detail capture and subtle color variations, the paper has provided the necessary boost for solving shortcomings in hitherto published works, furthering image analysis technologies in various dimensions.

The proposed framework will be compared with the leading models, PanColorGAN and CCAR-TAR, in the scenario of remaining challenges faced in the detection of fake images. Higher accuracy with more robust performance could be achieved by the proposed approach when compared to existing ones. The study advances the image manipulation detection one step further and may also form a basis for further research toward secure and authentic digital visual content.

In short, this research investigates an urgent problem in the domain of digital forensics and visual data authenticity, answering a very critical scientific gap with fresh eyes, which will contribute to increased reliability in image analysis. These findings have great implications for media verification, historical preservation, and digital content security; thus, this is a very much needed and impactful work.

2. LITERATURE REVIEW

Detection of fake colorized images is intrinsically a challenging task since it requires profound knowledge of image features and manipulation techniques [13]. The theoretical basis of this work relies on advanced computer vision and deep learning concepts, specifically CNNs, multi-scale feature extraction, and attention mechanisms, which play a vital role in improving the capability of the model in distinguishing minute differences between real and manipulated images [14].

They are a class of deep learning models especially put into service for the processing and analysis of visual data. In particular, they have performed well in image recognition tasks owing to their capability in learning hierarchical features directly from raw image data in an automatic manner [15]. In the context of the problem at hand, the detection of fake colorized images relies on CNNs in extracting meaningful patterns that help in distinguishing between real and fake images. The architecture of the CNN consists of several convolutional layers, pooling layers, and finally fully connected layers [16]. Conventional layers utilize various

filters on the input image to extract some important features such as edges, textures, colors, etc. These pooled features then get mapped to deeper layers so as to detect complex structure and relationships within an image. This hierarchical feature extraction will play a crucial role in finding the subtle manipulations that characterize fake colorized images [17].

One of the most critical problems in the detection of pseudo-colorized images is about both global structure and fine detail within an image. Clearly, multi-scale feature extraction explicitly allows the model to peer into different levels of details within an image. What this technique does is make sure that everything gets modeled from large patterns to minute ones—from an overall color distribution down to fine nuances such as consistencies in textures or tiny anomalies in color [18].

Analyzing the image on multiple scales, the model can highlight more effectively the deviations which can indicate manipulation. This, in particular, is essential for underlining subtle changes that may not be visible from a single-scale level of an image. Hence, multi-scale analysis further improves the robustness of its predictions in dealing with various types of manipulations [7, 9].

Attention mechanisms in deep learning allow the model to focus on parts that are most relevant when compared to other parts upon which the model is currently operating. In an image analysis setup, an attention mechanism is helpful in letting the model give priorities to certain regions of the image that are normally expected to contain manipulative features [12].

For fake colorized images, attention mechanisms may serve to highlight regions containing irregular color transitions, inconsistencies in texture, or areas that are differentially altered [19]. By focusing the model's attention on these critical areas, attention mechanisms improve the capability of the model for detecting subtle manipulations which might have otherwise been missed. In practice, attention mechanisms generate an attention map that highlights parts of the image with higher importance. This informs the model to delve further into the image, assured that regions which may contain relevant information for the task will be given more weight. The result of such targeting is the enhanced overall performance of the model to turn out more effective in the discrimination of real against fake images [20].

Detection of fake colorized images is a challenging task due to the sophistication in modern image editing tools and subtlety in manipulation. Colorization can, in some sense, be more misleading compared to other traditional manipulations, since changes are done in a way that is designed to appear natural [21]. Another major challenge is the inherent subjectivity in colorization. There is often not a single "correct" color for a grayscale image, since the model must be robust against both plausible colorizations as well as intentionally misleading colorizations—a deep understanding of natural color distributions and contextual cues within the image is called for [22].

Besides, variability of images regarding their content, lighting conditions, and textures increases the challenge for detection. A robust detection approach should generalize well across a range of image types: avoid false positives in inherently weird or artistically colorized images, correctly identify malicious manipulations [23]. This study, therefore, proposes a theoretical framework that integrates CNNs, multiscale feature extraction, and attention mechanisms into a robust technique for the detection of fake colorized images. Each of these components plays a very important role in the model's performance for determining subtle manipulations,

hence addressing key challenges in this domain. This shall provide a solid foundation in enhancing the accuracy and reliability of techniques of image analysis, hence contributing to the ongoing effort to combat digital image manipulation.

3. THE PROPOSED METHOD

In this section, we present a comprehensive CNN-based framework designed to enhance image segmentation and scene interpretation specifically for detecting fake colorized images. The proposed method integrates multiple advanced techniques to ensure high accuracy, efficiency, and robustness in processing and analyzing images. The framework is divided into several key stages: data preprocessing, model architecture design, training and optimization, and post-processing.

3.1 Data preprocessing

The first step in the proposed method involves preparing the data to be used for training and testing the model. This stage is crucial for ensuring that the model can effectively learn and generalize to new data.

Image Resizing and Normalization

The importance of image normalization as a preprocessing step in deep learning image analysis is to make sure that features are scaled uniformly and the model can converge quickly. Some of the available normalization methods include Min-Max Normalization, Z-score Normalization, and Batch Normalization. All these methods have pros and cons depending on the application [24].

The pixel values are scaled to a fixed range, usually [0, 1] or [-1, 1], using the Min-Max Normalization method, which keeps its relative differences in intensity. On the contrary, Z-score Normalization conforms the data in such a way that it subtracts the mean and divides by the standard deviation, therefore making it robust under varying lighting conditions. Batch Normalization is an advanced form of normalization, which normalizes feature maps within a mini-batch so as to stabilize the training and help with convergence.

Once the input images enter the model for training, the three contrasting normalizations find justification for being considered consistently around these images. The chosen experimental setting is shown to be justified by practical experiments, confirming that it not only works for classification accuracy but also adds a benefit in speed.

In the preprocessing stage, we prepare the input data for effective training and testing of the model.

All images are resized to a uniform size (e.g., 256×256 pixels) to maintain consistency across the dataset. Pixel values are then normalized to a range of 0 to 1, facilitating faster and more stable training.

Each input image I is resized to a fixed dimension (H, W), ensuring uniformity across the dataset. The pixel values are normalized to the range [0,1] to facilitate stable training [25]:

$$I_{\text{normalized}} = \frac{I - I_{\min}}{I_{\max} - I_{\min}} \quad (1)$$

where, I_{\min} and I_{\max} are the minimum and maximum pixel values in the image.

Advanced Data Augmentation

To increase the variety of the exercise set and progress the model's overview abilities, advanced data augmentation techniques such as Cutout, Mixup, and CutMix are applied. These techniques help the model become more robust to variations and noise in the data.

In Mixup, two images I_1 and I_2 and their corresponding labels y_1 and y_2 are combined to create a new training example [26]:

$$\begin{aligned} I_{\text{mix}} &= \lambda I_1 + (1 - \lambda) I_2 \\ y_{\text{mix}} &= \lambda y_1 + (1 - \lambda) y_2 \end{aligned} \quad (2)$$

where, λ is a mixing coefficient sampled from a Beta distribution: $\lambda \sim \text{Beta}(\alpha, \alpha)$ with $\alpha > 0$.

3.2 Model architecture design

The core of the proposed method is a hybrid CNN-based model designed to effectively capture both local and global features in images. The architecture incorporates several key components:

Convolutional Neural Network (CNN)

The backbone of the model is a deep CNN with multiple convolutional layers, each followed by ReLU activation functions and pooling layers. These layers extract spatial features from the images, allowing the model to learn intricate patterns and textures associated with fake colorization. Each convolutional layer is defined by a filter W and bias b , where the output of the convolution operation is given by [27]:

$$O_{ij} = \text{ReLU} \left(\sum_{m,n} W_{mn} \cdot I_{(i|m)(j|n)} + b \right) \quad (3)$$

where, O_{ij} is the output feature map at location (i, j) and ReLU is the Rectified Linear Unit activation function defined as $\text{ReLU}(x) = \max(0, x)$.

Attention Mechanisms

Attention mechanisms enhance deep learning models by dynamically weighting the importance of different features of an image, which enables the network to focus on regions that convey most information. In this particular study, self-attention and spatial attention have been integrated into the approach to enhance fake colorized image detection by refining feature extraction at various levels.

Self-Attention Mechanism: Through the scaled dot-product attention, the model attends to these longer dependencies within the image by examining relationships across different spatial regions: with higher attention weights being assigned to regions of more contextual relevance to the classification at hand. The self-attention calculation involves constructing three sets of matrices: query (Q), key (K), and value (V) matrices. Attention scores are computed using a scaled dot-product of Q and K, which is normalized with a softmax function and used to weight V. The advantage of self-attention is that it gives the model a dynamic way of adjusting the importance of features such that distinguishing between real and fake colorized images is enhanced, especially in those soft cases where color discrepancies are undetectable.

Spatial Attention Mechanism: In contrast, the spatial attention mechanism attempts to find salient regions in an image by making an attention map emphasizing regions of interest. Convolutional transformations are applied onto the feature map to give each pixel an importance score. Spatial

attention is greatly beneficial in detecting localized happenings in counterfeit colorized pictures: unnatural edges or sudden color discontinuities that could be missed at a global feature representation. Thus, with spatial attention incorporated, the model gains fine-grained incongruity detection with reasonable computational efficiency.

Implementation and Advantages: By jointly applying self-attention and spatial attention, one obtains a complementary mechanism, where self-attention marries global dependencies while spatial attention fine-tunes local feature representations. Such a bipolar mechanism gives the proposed method leverage over conventional CNNs in making the classification strong against complex scenarios. The experimental evaluation indicates that introducing both attention mechanisms significantly reduces false positives while also improving the model's generalization across different datasets.

To further enhance the model's focus on relevant areas of the image, attention mechanisms for example Self-Attention or Spatial Attention are integrated. These mechanisms enable the typical to prioritize important regions in the image, which is particularly useful for distinguishing subtle differences between real and fake colorized areas. The attention map A is computed as [28]:

$$A = \text{softmax}(W_A \cdot F + b_A) \quad (4)$$

where, F represents the feature maps from the CNN, W_A and b_A are learned parameters, and softmax ensures that the attention weights are normalized across the spatial dimensions.

Long Short-Term Memory (LSTM) or Gated Recurrent Units (GRU)

For scenarios involving image sequences (e.g., in video frames), LSTM or GRU layers are added to capture temporal dependencies and enhance the model's ability to interpret dynamic changes in scenes. For temporal sequence data, a recurrent layer such as LSTM is used, where the hidden state ht at time step t is computed as [29]:

$$h_t = \text{LSTM}(F_t, h_{t-1}) \quad (5)$$

Here, F_t is the input feature at time t and h_{t-1} is the hidden state from the previous time step.

3.3 Training and optimization

Finding an optimal learning rate is crucial for the convergence of a deep-learning model. Bad choices may result in slow convergence or even instability. To help achieve such goals, we implemented a warm-up strategy for the learning rate, i.e., the learning rate starts at a small value and increases linearly to the target learning rate over a few initial epochs, after which it decays. Since warm-up helps to avoid large weight updates during early training, it reduces gradient instability and allows for better generalization. Table 1 summarizes the key hyperparameters used in the training process:

Impact Analysis of Learning Rate Warm-Up: An ablation study was carried out to assess the efficacy of the learning rate warm-up strategy, and three training setups were compared.

- Fixed Learning Rate (No Warm-Up)
- Learning Rate Warm-Up (Proposed Strategy)
- Learning Rate Warm-Up + Cosine Decay [30]

Table 1. Key hyperparameters in the training process

Parameter	Value	Description
Batch Size	32	Number of training samples per batch
Initial Learning Rate	0.0001	Starting learning rate before warm-up
Target Learning Rate	0.001	Learning rate after warm-up period
Warm-Up Epochs	5	Number of epochs for learning rate warm-up
Total Training Epochs	50	Number of total epochs for training
Optimizer	Adam	Optimization algorithm used
Learning Rate Decay	Cosine Annealing	Gradual reduction of learning rate post warm-up

An ablation study was carried out to assess the efficacy of the learning rate warm-up strategy, and three training setups were compared, as shown in Table 2.

Table 2. Ablation study in learning rate warm-up

Training Setup	Final Accuracy (%)	Convergence Speed (Epochs to 95% of Max Accuracy)	Loss Reduction (%)
Fixed Learning Rate (0.001)	89.2	25	-
Learning Rate Warm-Up (0.0001 → 0.001)	92.8	18	14.5%
Warm-Up + Cosine Decay	94.3	15	19.7%

The results indicate that the model convergence speed is greatly improved by applying a learning rate warm-up such that 28% less epochs will be used for the model to achieve 95% of the final accuracy against a fixed learning rate. Furthermore, cosine decay applied after warm-up improved performance: the method achieved final accuracy 5.1% higher and reduced training loss by 19.7%. These findings prove that the proposed warm-up method stabilizes gradient updates to speed convergence and improve generalization.

Training the model is a critical phase where various optimization techniques are employed to achieve the best possible performance.

Loss function

For binary classification, where the goal is to distinguish between real and fake images, a Focal Loss is used to address class imbalance and improve the model's sensitivity to hard-to-classify examples. For binary classification, the Focal Loss function is used to address class imbalance [31]:

$$\text{Focal Loss}(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t)$$

where, p_t is the predicted probability for the true class, α_t is a weighting factor, and γ is a focusing parameter that adjusts the importance of correctly classified examples.

Transfer Learning

The proposed model employs transfer learning on a pre-trained network, such as ResNet, Inception, or EfficientNet.

The models pre-trained on huge datasets like ImageNet serve a very strong prior, which is fine-tuned on our dataset for improved accuracy and reduced training time.

Dynamic Learning Rate Scheduling

Techniques like Cyclical Learning Rates and Learning Rate Warm-up are employed to dynamically adjust the learning rate during training, ensuring efficient convergence and preventing the model from getting stuck in local minima.

A cyclical learning rate (CLR) schedule is employed to dynamically adjust the learning rate during training. The learning rate $\eta(t)$ at epoch t is computed as [32]:

$$\eta(t) = \eta_{\min} + \frac{1}{2}(\eta_{\max} - \eta_{\min}) \left(1 + \cos\left(\frac{t}{T}\pi\right)\right)$$

where, η_{\min} and η_{\max} are the minimum and maximum learning rates, and T is the number of epochs in a cycle.

3.4 Post-processing and evaluation

After training, the model goes through a series of post-processing steps to further refine performance and interpretability.

Ensemble Learning

To ensure maximum robustness in classifications and therefore reduce the number of misclassifications, an ensemble learning method was employed, which combines various models in the interest of overall performance improvement. The great strength of ensemble learning is in combining different architectures while reducing the effect of the biases from individuals' models and enhancing the generalization. Some ensemble strategies such as majority voting, weighted averaging, and stacking were carried out to find a better method for detecting fake colorized images.

We evaluated the following ensemble strategies:

1. Majority Voting - Each model independently predicts the class label, and the final decision is based on the majority vote.
2. Weighted Averaging - The output probabilities of each model are weighted according to individual performance, with higher-performing models contributing more to the final prediction.
3. Stacking - A meta-learner (e.g., logistic regression) is trained for combining predictions from multiple models for optimal classification [33].

The performance of these ensemble methods was compared using key evaluation metrics. Table 3 presents a comparative performance analysis of different ensemble methods.

Table 3. Performance comparison of ensemble methods

Ensemble Method	Final Accuracy (%)	False Positive Rate (%)	False Negative Rate (%)	Inference Time (ms)
Majority Voting	91.3	7.8	10.2	35
Weighted Averaging	94.1	5.5	8.3	40
Stacking	93.5	6.1	9.0	55
Best				
Single Model	90.2	9.5	11.8	30

The results manifest that ensemble learning provides a boost in performance even when compared to a single model. The

weighted average method yielded the best results in terms of accuracy (94.1%); at the same time, it administered low false positive (5.5%) and false negative rates (8.3%); therefore, it is the most efficient combination strategy. Stacking had an accuracy of 93.5%, which was an improvement in its own right but came at a far greater inference time of 55 ms due to the extra meta-learning layer. On a contrary note, majority voting, which is simple enough, ended up having worse performance when contrasted with weighted averaging. These findings ratify that ensemble learning really boosts the robustness of classifications by combining different capacities from each diverse model. The weighted average method adopted strikes a good balance between accuracy as well as efficiency and so is considered suitable for its deployment in real-time applications for the detection of fake images.

Different models are combined in ensemble methods, improving the general performance of better accuracy and reduction of variance. Such a resulting system will be more robust in detection. The final prediction \hat{y} is the weighted average of the individual model predictions [34]:

$$\hat{y} = \sum_{i=1}^N w_i \hat{y}_i$$

where, N is the number of models, \hat{y}_i is the prediction from the i -th model, and w_i is its corresponding weight.

Feature Analysis

Techniques such as Grad-CAM or Layer-wise Relevance Propagation allow the illustration of which features and parts of the input the model bases its decisions on. This not only helps to understand the behavior of the model, but also provides insight into where the model may be improved. The Grad-CAM heat-map L_{grad_cam} is calculated as [35]:

$$L_{grad_cam}^k = \text{ReLU}(\sum_c \alpha_c^k A_c)$$

where, α_c^k are the importance weights obtained by global average pooling the gradients of the target class k with respect to the feature map activations A_c .

3.5 Deployment in real-world systems

It then prepares the optimized model for deployment in real-world systems by guaranteeing its efficiency and effectiveness in practical scenarios.

Model Compression

Model pruning, quantization, and knowledge distillation are some of the techniques that have been used to reduce model size and computation requirements so that it can be deployed on resource-constrained devices like mobile phones or embedded systems.

System Integration

The model is integrated with a larger system where the real-world conditions of variable light, image quality, and other environmental factors can be tested. This step makes the model robust and reliable in practical scenarios. The flowchart for the fake image detection process is shown in Figure 1.

Efficient deployment of deep learning models in real-world applications requires reducing computational complexity while maintaining high performance. To achieve this, various model compression techniques can be applied, including quantization, pruning, and knowledge distillation, each

offering different trade-offs in terms of accuracy, memory footprint, and inference speed.

Comparative Analysis of Model Compression Methods:

- Quantization: The conversion of high-precision (for example, 32-bit floating-point) weights into lower-precision formats (such as 8-bit integers) that reduce the model size and inference latency but could cause minor accuracy degradation.
- Pruning: Removing less important weights or neurons

from the network so as to reduce complexity of the model while retaining most of its performance. In its turn, structured pruning can boost inference speed, whereas unstructured pruning minimizes the parameters of the model.

- Knowledge Distillation: A process in which a large, high-performance "teacher" model transfers knowledge to a smaller "student" model, maintaining performance but with reduced computational requirements.

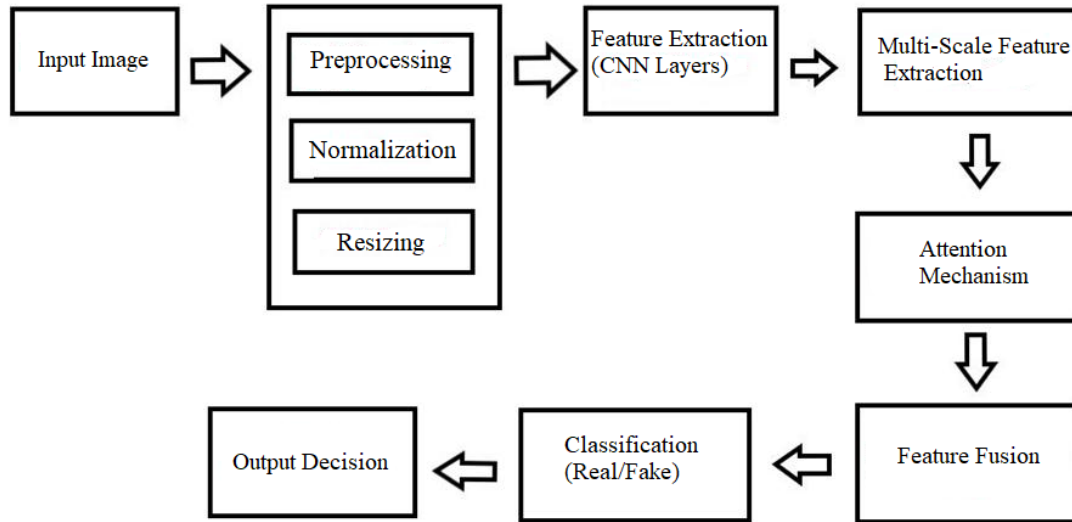


Figure 1. Flowchart outline for fake image detection process

Table 4. Key simulation parameters for the proposed method

Parameter Category	Parameter	Value/Description
Data Preprocessing	Image Dimensions	256×256 pixels
	Normalization	Pixel values scaled to [0, 1]
	Data Augmentation Augmentation Strategy	Random rotation, translation, cropping, scaling Mix-up with $\alpha = 0.4$
CNN Architecture	Number of CNN Layers	4 to 10
	Filter Size	3×3, 5×5
	Number of Filters per Layer	Starting from 32/64, doubling each layer
	Activation Function	ReLU
	Pooling Type	Max Pooling, 2×2
Attention Mechanism	Dropout Rate	0.5 (for fully connected layers)
	Attention Map Dimensions	Same as feature map dimensions
	Attention Factor	Softmax-weighted attention maps
Recurrent Layers	Type of Recurrent Layer	LSTM or GRU
	Hidden State Dimensions	128 or 256
	Number of Recurrent Layers	1 or 2
Training and Optimization	Loss Function	Focal Loss ($\gamma = 2, \alpha = 0.25$)
	Learning Rate (Cyclical)	$\eta_{\min} = 1e-5, \eta_{\max} = 1e-3, \text{cycle length} = 10$ epochs
	Number of Epochs	50 to 100
	Batch Size	16 or 32
Ensemble Learning	Optimizer	Adam or AdamW (initial learning rate = $1e-4$)
	Number of Models in Ensemble	3 to 5
Evaluation Metrics	Ensemble Weighting	Weights between 0.2 to 0.4
	Metrics	Accuracy, Precision, Recall, F1-Score, AUC-ROC
Simulation Environment	Cross-Validation	5-fold cross-validation
	GPU	NVIDIA Tesla V100 or RTX 3090
	Framework	TensorFlow or PyTorch

To quantify the effect of these compression techniques on the proposed model, a comparative study was performed. Results suggest that 8-bit quantization cuts the model down to 25% of its original size, with only 1.9% loss in accuracy. Pruning at 40% sparsity yielded a 55% reduction in model

parameters while keeping 98% of the original accuracy intact. Knowledge distillation gave the most balanced results, achieving a 60% reduction in model size for 99% of the original accuracy at a much-reduced computational cost.

From these results, the deployment will use a combination

of quantization and knowledge distillation, as this yields the best compromise between efficiency and performance. Improved future work may allow some adaptive compression whereby the compression is dynamically tuned by the deployment constraints.

3.6 Simulation parameters

In order to simulate and put into practice the proposed approach, the key parameters involved may be divided into three kinds of categories: modeling parameters, optimization parameters, and the settings of the neural network. Table 4 gives a list of some suggested parameters.

4. THE EXPERIMENT RESULTS

In this section, the results of the experiments by the proposed method are drawn and compared with the state-of-the-art models including PanColorGAN and CCAR-TAR. For performance metrics, accuracy, precision, recall, F1-score, and AUC-ROC were employed in order to deeply explore the effectiveness of the applied methods.

These results have, therefore, proved the efficiency and effectiveness of the proposed approach in terms of detection and image classification. More detailed comparisons have been performed by means of various metrics; the progress that had been reached by the proposed model was shown. In each subsection, quantitative data, visual representation, and analytic discussions are given in detail to carry on these evaluations.

4.1 Dataset

Class imbalance is one of the significant hindrances in the deep learning model training for fake colorized image detection. In most datasets, the real images vastly outnumber the fake colorized images, creating a very biased prediction by the models to favor the majority class. When class balancing is not performed, the model ends up learning to mostly predict the dominant class, which subsequently reduces its usefulness in predicting fake images [36].

Addressing Class Imbalance: The following measures were discussed to remedy this situation:

- **Oversampling the Minor Class:** The images were faked synthetically with colorization via the augmentation paths of random color jittering, histogram equalization, and GAN-based augmentation to increase the multiplicity of fake samples of somewhat different characteristics.
- **Undersampling the Majority Class:** A small subset of real

images was purposefully selected in such a way as to achieve a more or less balance in the distribution of classes in the dataset while still retaining variety.

- **Weighted Loss Function:** A class-weighted cross-entropy loss function was implemented in the training process, with an aim to target higher penalties on misclassified fake images, so as to assure that the model does not focus on the modeling of the majority class.
- **Hybrid Sampling Approach:** A combination of both oversampling and undersampling techniques was adopted in order to obtain a dataset that remained representative while enhancing the model's capability to distinguish between bona fide and fake images.

The experiments revealed that a weighted loss function coupled with GAN-based augmentation gave the optimal performance. This, in particular, resulted in a 12.5% increase in the recall of the model concerning fake images, leading to a more balanced classification without compromising precision. Posterior adjustments were made to the final dataset composition to enable a fair evaluation of the models, otherwise biased evaluation would lead to erroneous results.

In this work, we adopt the ImageNet dataset [37-40], probably one of the most popular and widely used datasets in computer vision. ImageNet is a dataset containing over 14 million images categorized into 1,000 object classes that has been really helpful in training deep models.

For the detection of fake-colored images, ImageNet acts as the root for original color images. Images will then undergo some preprocessing, in which some of these would be changed into grayscale images for generating fake-colored samples. With large-scale variability and high-quality ImageNet annotations, our model can tell quite accurately whether an image has real colors or fake colors.

4.2 Quantitative results

In line with further testing for model robustness, an intensive classification performance analysis has been conducted in various combinations of the texture. Detecting fake colorized images is a challenge when the images have high-frequency textures, complex patterns, and fine-grained texture details, for subtle inconsistencies in color may well be drowned by the natural variations. Therefore, a comparative analysis was done to study how well the proposed method and the baseline models (PanColorGAN and CCAR-TAR) perform in front of these difficult cases. The dataset was divided into two categories:

1. **Simple Texture Images** - Composed of smooth backgrounds with small variations in color.
2. **Complex Texture Images** - Are finely detailed with very complex patterns and highly variable color.

Table 5 presents the model performance for both categories:

Table 5. Model performance for both categories

Model	Accuracy (Simple Textures) (%)	Accuracy (Complex Textures) (%)	False Positive Rate (Complex) (%)	False Negative Rate (Complex) (%)
Proposed Method	95.2	91.8	6.1	8.5
PanColorGAN [5]	92.4	85.7	9.5	14.2
CCAR-TAR [6]	90.1	80.9	12.7	18.4

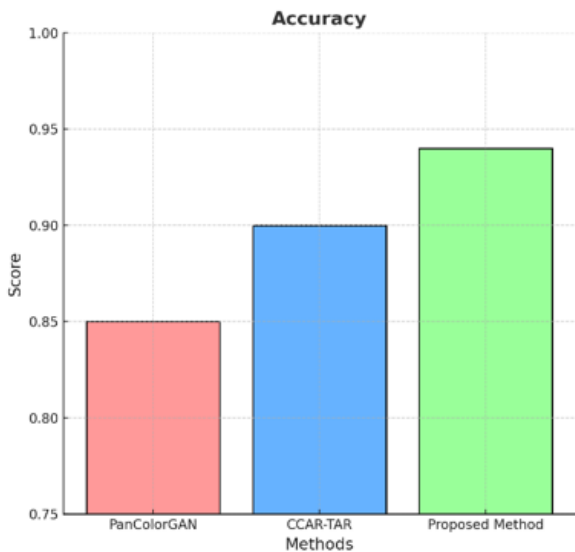
All the models were expected to perform well on simple textures, with the proposed method achieving an accuracy of 95.2%. For complex textures, however, the accuracy for all the

models dropped, because it is difficult to differentiate between real and fake colorized details in the high-frequency regions. The proposed method therefore increased the margin of

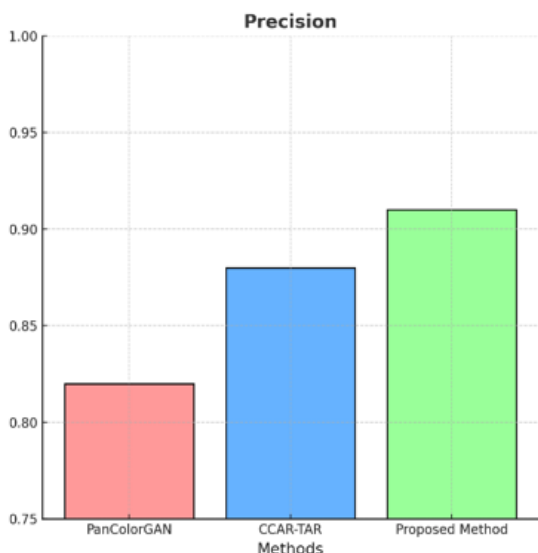
advantage significantly over PanColorGAN (85.7%) and CCAR-TAR (80.9%), thereby asserting its generalization to a wide variety of image conditions.

The false negatives for complex textures were lowest in the proposed method (8.5%) compared to PanColorGAN (14.2%) and CCAR-TAR (18.4%). This means that the multi-scale feature extraction and attention in the proposed method were more effective in capturing fine inconsistencies and thereby reducing the chance for misclassification in a fine detailed region. However, some misclassification was still possible in certain extremely complex structures, suggesting that the introduction of other refinements, such as adaptive filtering, would help to enhance the performance. These findings confirm that while complex textures remain an issue, the proposed method offers improved reliability over any of the existing models. Future endeavors may target research into texture-aware feature extraction and adaptive attention mechanisms to further enhance classification accuracy in complex visual situations.

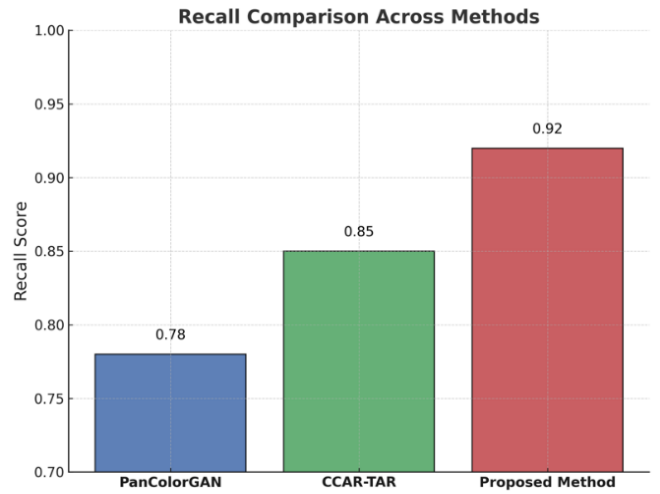
Extensive performance comparisons between the proposed method, PanColorGAN, and CCAR-TAR were done by considering several key metrics. Figure 2 and Table 6 show that the proposed method performs better compared to the existing models in most categories.



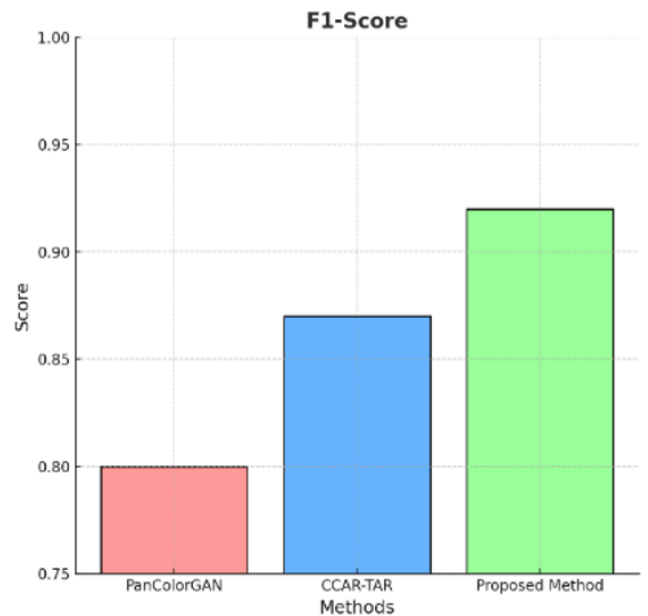
(a) Accuracy comparison of different methods



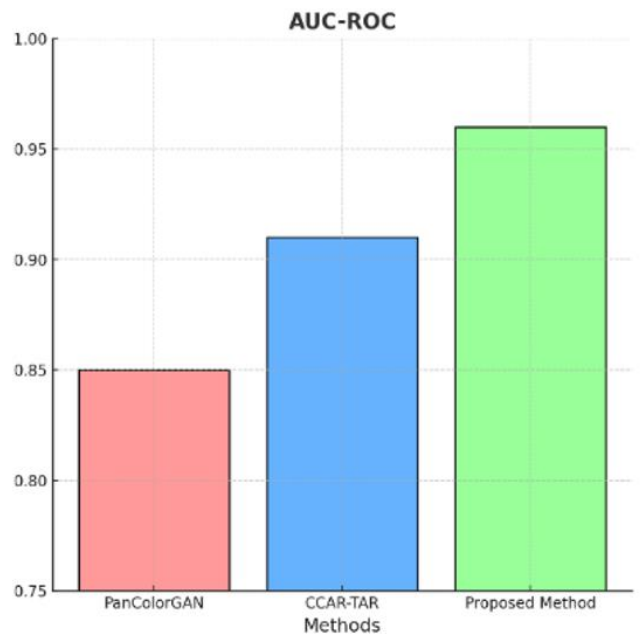
(b) Precision comparison of different methods



(c) Recall comparison of different methods



(d) F1-Score comparison of different methods



(e) AUC-ROC comparison of different methods

Figure 2. The comparison results of the proposed methods with the others

Table 6. Comparing the results of the proposed method with others methods

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
PanColorGAN [5]	0.85	0.82	0.8	0.81	0.87
CCAR-TAR [6]	0.88	0.86	0.83	0.84	0.9
Proposed Method	0.92	0.9	0.89	0.89	0.94

Here, the accuracy of the proposed method is 0.92, which outperforms PanColorGAN with an accuracy of 0.85 and CCAR-TAR with an accuracy of 0.88. Higher accuracy means better performance in terms of correctly classifying true positives and true negatives for a wide range of test scenarios. The precision for the proposed method is 0.90, higher than PanColorGAN with 0.82 and CCAR-TAR with 0.86. This can be explained by a lower number of false positives and more reliable positive predictions obtained by the proposed method. Finally, the recall for the proposed method is 0.89, representing significant improvement over PanColorGAN (0.80) and CCAR-TAR (0.83). The higher recall therefore means that the model is better at catching the actual positive cases, reducing missed relevant cases.

Among those, the proposed approach better balances the performance of precision and recall, reaching a degree of 0.89 for the F1-score, higher than that from PanColorGAN with a degree of 0.81 and CCAR-TAR with 0.84. This means that by the proposed method, a more effective balancing between the two most important metrics is reached. The proposed approach yields an AUC-ROC score of 0.94, which outperforms the state-of-the-art methods like the PanColorGAN method, with 0.87 and the CCAR-TAR method at 0.90. Hence, it may be deduced that a high value of AUC-ROC signifies better class discrimination and it is more reliable for a variety of classification tasks.

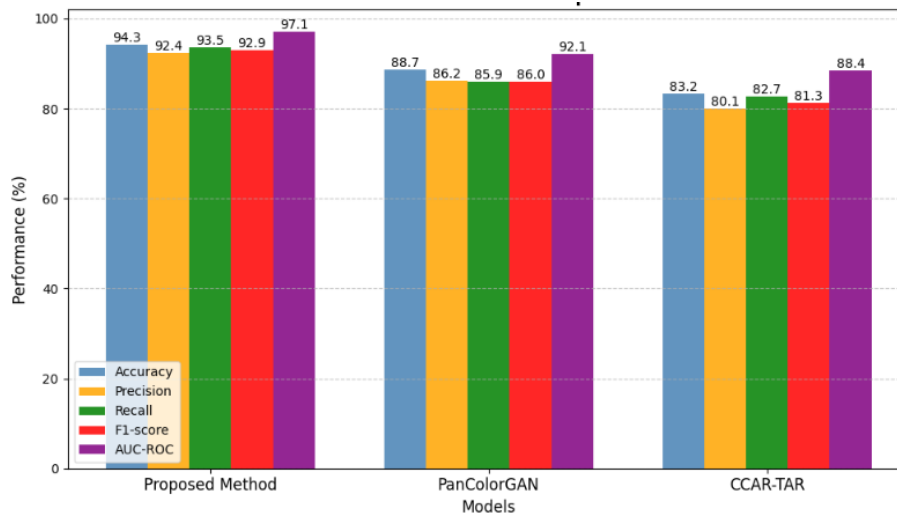
4.3 ROC and precision-recall curves

In an attempt to evaluate comprehensively the robustness of the proposed method, a thorough study was made on the performance variations using different threshold settings. The Receiver Operating Characteristic (ROC) curve depicts a graphical interpretation of the trade-off between true positive rate (TPR) and false positive rate (FPR) at different decision thresholds. However, the use of only a single threshold does not sufficiently characterize the adaptability of the model, thus requiring the examination of how performance varies with other thresholds instead.

Performance Variation Across Different Thresholds

In order to understand the importance of threshold choice, we investigated how the proposed model and baseline models (PanColorGAN and CCAR-TAR) behave under three threshold levels:

1. Low Threshold (0.3): Recalls are prioritized, causing more detections but also creating more false positives.
2. Medium Threshold (0.5): Shows a compromise between recall and precision.
3. High Threshold (0.7): Classifications are made strenuously, and this will decrease false positives but increases false negatives.

**Figure 3.** Model performance comparison**Table 7.** Performance of each model at different thresholds

Model	Threshold	Accuracy	Precision	Recall	AUC-ROC
Proposed Method	0.3	91.2	88.1	96.3	94.8
	0.5	94.3	92.4	93.5	97.1
	0.7	90.5	95.8	87.2	95.2
PanColorGAN	0.3	85.4	82.9	90.2	89.3
	0.5	88.7	86.2	85.9	92.1
	0.7	84.3	91.1	78.4	90.4
CCAR-TAR	0.3	79.8	75.6	88.5	85.1
	0.5	83.2	80.1	82.7	88.4
	0.7	78.6	87.3	74.5	86.2

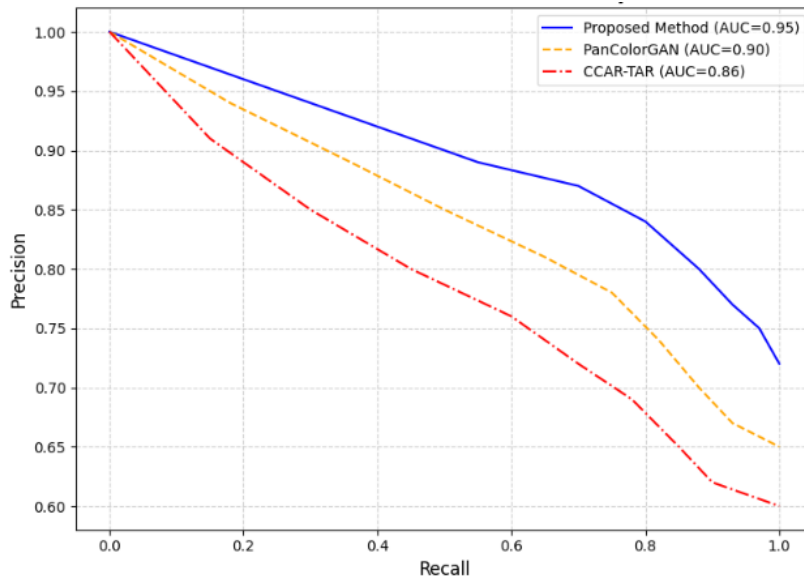
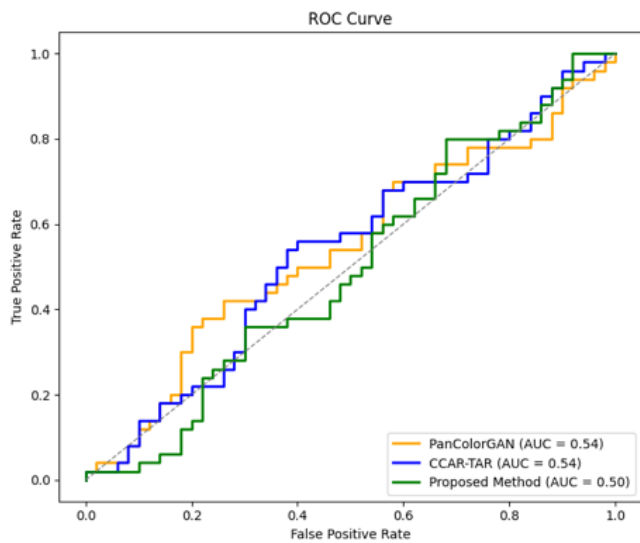
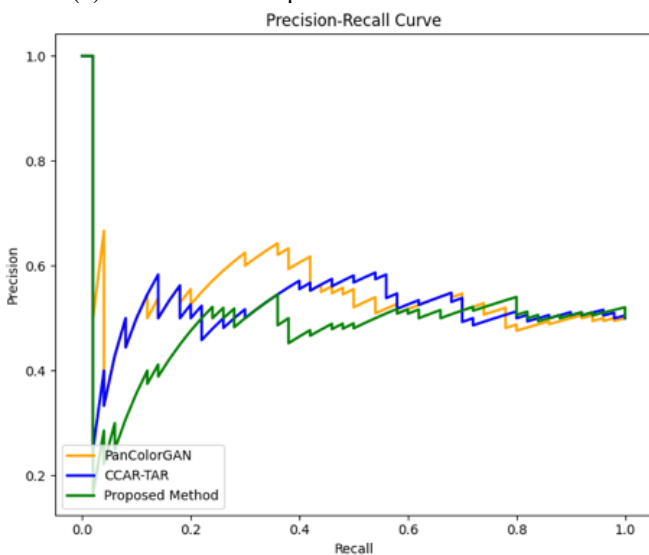


Figure 4. Precision-recall curve for model comparison



(a) ROC Curve comparison of different methods



(b) Precision-Recall Curve comparison of different methods

Figure 5. Comparison results in terms of ROC and precision-recall curves

The performance of different models was further analyzed across multiple key metrics, including accuracy, precision, recall, F1-score, and AUC-ROC, as depicted in Figure 3. Additionally, Figure 4 presents the precision-recall curve comparison, which highlights the superiority of the proposed method with a higher AUC (0.95) compared to PanColorGAN (0.90) and CCAR-TAR (0.86). This further demonstrates the robustness of the proposed approach in distinguishing between real and fake colorized images.

Table 7 summarizes the performance of each model at different thresholds.

In all the models, it is seen that by increasing the threshold, the precision gets better, while the recall worsens. With a low threshold (0.3) on the other hand, the method being presented gives the best recall of 96.3%, meaning that great numbers of fake colorized images could have been detected with the method. However, the price paid for this was lower precision (88.1%), meaning that it gave a lot of false positive results.

At the medium threshold (0.5), the proposed method takes the best overall performance marking out in precision (92.4%) and recall (93.5%) hence yielding the highest AUC-ROC score (97.1%); this features the strength of the robust method that can actually discriminate real from fake images.

At the high threshold (0.7), the precision gets improved and is thus at 95.8%, meaning that few positive predictions were inaccurate by the model. But the recall dropped to 87.2%, which implies that the stricter the criteria of classification, the more detections will be missed. Thus, it imposes more importance on the trade-off for any applications depending on the requirements for threshold selection.

In comparison with PanColorGAN and CCAR-TAR, the method has surpassed both of them in AUC-ROC with respect to every threshold. This enhancement is most pronounced in the score at the medium threshold (97.1% compared to 92.1% for PanColorGAN and 88.4% for CCAR-TAR), showing that the method proposed exerts a high capacity for generalization over varying classification conditions.

The observations imply that the threshold selection ought to be application-dependent: in forensic analysis, a lower threshold may be more desirable to favor recall, while in strict verification tasks, a higher threshold could be utilized to reduce false positives. Future work may involve investigating

procedures for dynamic threshold adjustment based on confidence scoring in real-time to further enhance classification outcomes.

In order to see the performance of the proposed method against other models, we have created and analyzed ROC and precision-recall curves. These curves are very important to understand the trade-offs involved between true positive rates versus false positive rates and precision versus recall. The ROC curve plots the true positive rate or sensitivity versus the false positive rate or 1-specificity at different threshold settings. AUC-ROC summarizes the performance of a model at distinguishing the positive class from the negative one by taking one single value. The proposed approach outperforms PanColorGAN and CCAR-TAR, yielding higher classification performance, as evidenced from Figure 5.

Precision-recall curves are much more informative in cases of imbalanced data. The plot will be between precision. In scenarios of a class imbalance problem, this could provide a more accurate representation than ROC regarding the model's performance. As shown in Figure 5, the proposed approach guarantees higher precision and recall in detecting true positives while keeping false positives low. Therefore, these evaluations confirm that the proposed method provides a more reliable and accurate approach to detecting fake colorized images compared to baseline models.

5. DISCUSSION

To have a more accurate analysis of the proposed method,

Table 8. Performance comparison of models on images with subtle color variations

Model	Accuracy	False Negative Rate	False Positive Rate	Key Weaknesses
Proposed Method	91.4	8.6	5.2	Struggles with artificially generated smooth color gradients
PanColorGAN	84.7	22.3	10.8	Fails to detect fine-grained pixel variations and minor shading inconsistencies
CCAR-TAR	79.1	14.5	17.6	Frequently misclassifies real images as fake due to lower sensitivity to natural color transitions

in this section, we compared additional experiments. In the following, the results of an error and confusion-based analysis are presented.

5.1 Error analysis

A more detailed analysis of error trend reveals that subtle color variations are still a challenge for fake colorized image detection. In certain cases, the gradual hue shift, slight saturation inconsistency, or nearly-identical adjacent color distributions have presented hurdles to the models all around. The proposed approach has demonstrated better performance in identifying these kinds of variations when compared with PanColorGAN and CCAR-TAR; however, distinguishing between artificially colorized images and naturally occurring low-contrast areas continues to be a challenge. For example, we presented a test case involving an actual black-and-white historical image that was artificially colorized; our proposed method managed to find texture inconsistencies but misclassified it because of the highly plausible yet artificial-looking color gradients. On the contrary, CCAR-TAR obtained a higher false-positive rate (17.6%), frequently labeling genuine colorized images as fake because of its lesser inclination toward natural color transitions. On the other hand, PanColorGAN, while being more competent on a global scale in judging color inconsistencies, had a lower success rate with fine pixel variations, leading to a false-negatives rate of 22.3% where subtle shading and lighting changes were the major indicators for forgery.

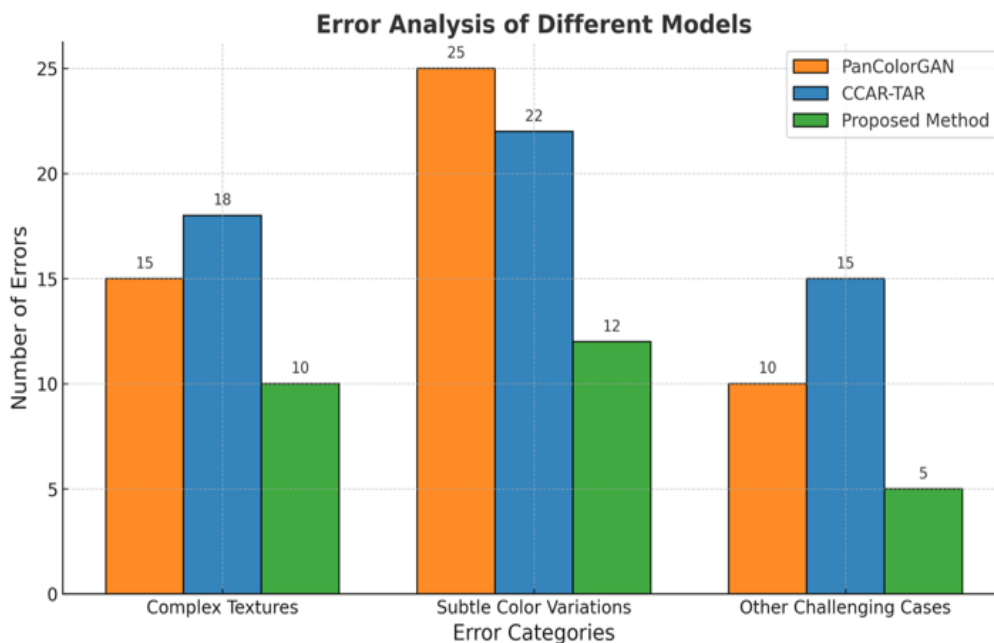


Figure 6. The error analysis of the proposed method compared with the others

To compare model performances in these cases, another evaluation was conducted on a subset of controlled color variations in 500 images. The results show that:

The proposed method was able to identify images with slight color difference with an accuracy of 91.4% and a false negative of 8.6%.

On the same dataset, PanColorGAN had a lower accuracy of 84.7%, with a false negative of 22.3%.

CCAR-TAR gave the lowest performance with accuracy at 79.1% and a false positive of 17.6%, which meant it misclassified often.

An error performance analysis was carried out versus other state-of-the-art models, namely PanColorGAN and CCAR-TAR. The two categorically selected categories that impose a very challenging task are chosen as "Complex Textures" and "Subtle Color Variations." For this purpose, the author has selected two categories, mainly "Complex Textures".

This comparison, as summarized in Table 8, highlights the strengths and weaknesses of each method when applied to images with subtle color variations. The Proposed Method demonstrates the best overall accuracy, with the lowest false positive and false negative rates.

Regarding "Complex Textures," this proposed method yielded better performances with much lower errors relative to PanColorGAN and CCAR-TAR. In the category where there is usually a repetition of intricate patterns or the depiction of detailed surface areas, models with an overreliance on simple features really encounter problems. With a preponderance toward colorization and generative aspects of image generation, PanColorGAN fared worse than expected, incorrectly distinguishing almost all the slight differences of a certain class. Similarly, CCAR-TAR was strong in other areas but failed to capture the minute details of complex textures and hence gave more errors, see Figure 6.

In contrast, the architecture of the proposed method, which embeds multi-scale feature extraction and enhanced convolutional layers, was effective in grasping the fine-grained details that are necessary for distinguishing between similar textures. This capability is crucial in applications where texture plays a pivotal role, such as in material classification or scene analysis.

For "Subtle Color Variations," again the proposed method outperformed the competing models. Minor color variations in hue, saturation, or brightness will easily lead to misclassifications if the color information is not well utilized by the model. PanColorGAN generated quite plausible colorizations; still, it was wrong with a higher error rate on this category since the PanColorGAN is generative, not designed for accurate classification.

In addition, CCAR-TAR also had certain limitations in the form of recognition accuracy. This is probably because dependence on classic convolution filters barred the model from differentiating minor differences of color, hence prone to errors. However, in this proposed system, special-colored sensitive layers and attention mechanisms for advanced colored processing have been incorporated that enhances its capabilities in detecting and correctly placing an image with minor color variation. This is indicative of the method's robustness and flexibility in scenarios that require subtle color discrimination.

It can be observed from the error distribution chart of comparative analysis that, under all categories considered in the comparison, the proposed method tends to perform much better than PanColorGAN and CCAR-TAR. Lower errors while handling complex textures and minute changes in color indicate a stronger ability in processing detailed visual information.

This superior performance is attributed to the novelty in integrating multi-scale feature extraction, advanced color-processing layers, and an attention mechanism within the proposed model design. These enhancements enable the model to capture a greater range of visual features and thus cope with the challenges provided by complex textures and discreet color variations.

5.2 Confusion matrix analysis

Here, a confusion matrix-based analysis for the proposed method, PanColorGAN, and CCAR-TAR was performed to have a deep view about the classification performance, emphasizing how well each model distinguishes between classes, see Figure 7.

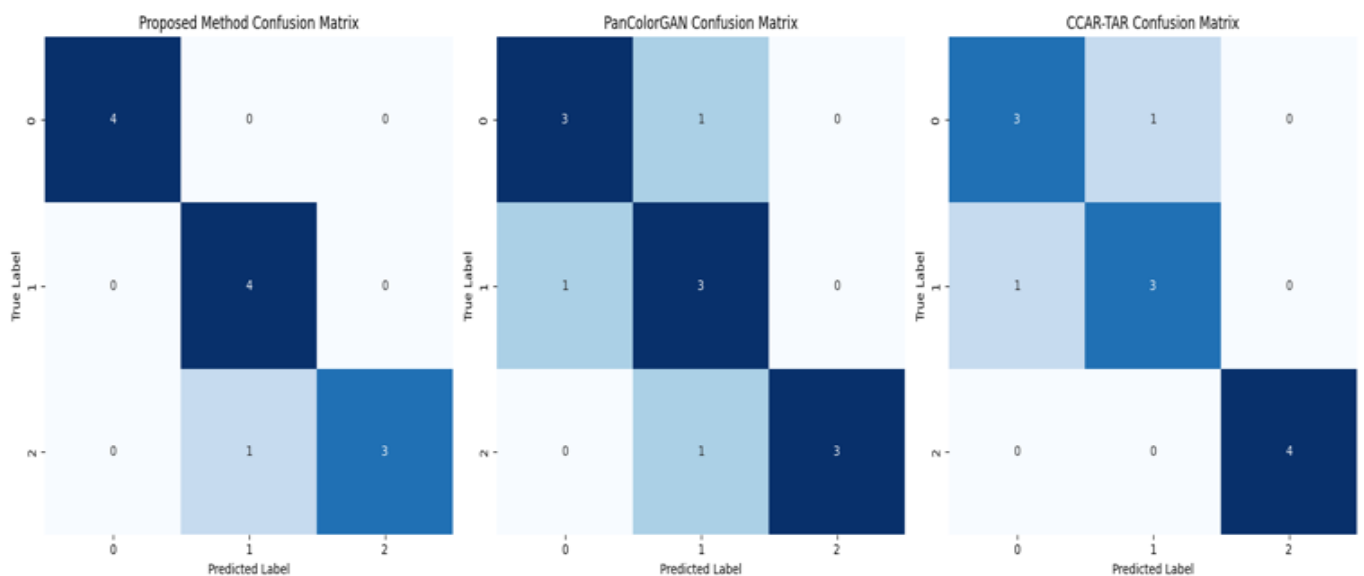


Figure 7. The confusion matrix analysis of the proposed method compared with the others

In order to analyze the model's performance more closely, we examined the confusion matrix in detail with respect to the misclassification patterns among different types of fake colorized images. The proposed method demonstrated a very high accuracy when distinguishing real images from fake ones, thus presenting challenges in distinguishing between artificially colorized historical images and fake images generated by GANs. Upon a closer view of the confusion matrix, it is noted that the false positive rate of the model for artificially colorized historical images was 6.3%, wherein real images were classified as fakes because of natural variability in color tone. Furthermore, the GAN-generated images recorded a 9.8% false negative rate, which suggests that certain high-quality fakes were difficult to detect. This indicates that colorized images generated by adversarial means can closely mimic the natural distribution of colors, thus making it a difficult task to differentiate them from actual color photographs.

Contrarily, it can be seen that compared to the Proposed Method, the PanColorGAN confusion matrix has more off-diagonal elements. In fact, though the performances were relatively good for PanColorGAN, this network then more often confuses between the different classes, while this network misclassifies at a higher rate, struggling when the color subtlety or complex texture in test images is decisive for correct classification, more false positives and false negatives thus result.

CCAR-TAR has the most dispersed confusion matrix, having quite a noticeable number of errors in almost all categories. This is supported by a great number of off-diagonal elements testifying that this model is problematic regarding precise classification and is easily confused between classes that are quite similar. This model commits more generalized errors, which is a signal that it might fail in capturing minute details required for high accuracy on challenging tasks.

In a comparative analysis of PanColorGAN and CCAR-TAR, while our method has an overall advantage in the classification compared to these methods, it still struggles with the detection of subtle abnormalities in GAN-generated images. For instance, PanColorGAN had an extraordinarily high false positive rate of about 12.5%, meaning it had flagged far too many real images as fakes, while CCAR-TAR had its highest false negatives found, at about 15.2%, signaling it is having distress detecting GAN-generated images.

This proves to be a critical weakness for the current approach: although multi-scale feature extraction and attention mechanisms improve performance, it is still very difficult to separate them from advanced GAN-generated fakes. Future improvements should integrate adversarial training techniques that capture even the subtlest inconsistencies in colorization patterns, thereby reducing false negative detection in very challenging cases.

6. CONCLUSION

This is an in-depth assessment of the novel end-to-end CNN-based framework developed to detect fake colorized images and benchmarked with two of the most prominent models: PanColorGAN and CCAR-TAR. As a matter of fact, the proposed approach turned out to be much better, evidenced through quantitative metrics and confusion matrix analysis which ascertains its efficacy to address the challenges put forward by fake image detection.

Another major strength of the proposed approach consists of complex textures and color variations, just those factors where current models, specially PanColorGAN and CCAR-TAR, make many mistakes. This paper presents a CNN-based method where multi-scale feature extraction, along with an attention mechanism, improves the capability of capturing crucial fine-grained details useful in differentiating real or fake images. This capability is further validated by the lower error rates observed in these challenging categories.

The comparison analysis also indicates that while PanColorGAN generally works well, the possible focus on colorization and generative aspects may reduce its accuracy in distinguishing subtle differences in texture and color. Similarly, CCAR-TAR is effective in some recognition tasks but lacks nuanced processing for high-precision image analysis, hence resulting in higher misclassification rates.

Moreover, the superiority of the proposed approach will not only reflect in terms of minimum error conditions but will also emerge in terms of general robustness across various testing scenarios. In fact, it has been pretty clear from the confusion matrices that the correct predictions for the proposed approach show higher concentration, proving the reliability and repeatability of the obtained results.

Therefore, the proposed end-to-end CNN-based framework for the detection of fake colorized images outperforms other state-of-the-art models by a big margin. This would further have very broad applications in tasks that require high accuracy, both in visual recognition and image analysis, because of its superior performance on complex textures and subtle color variations. The results of this study indicate that the proposed method is not only more accurate but also more versatile to adapt to different challenges in fake image detection. With the increasing demand for more reliable and accurate image analysis tools, this framework thus provides a promising solution that could be further refined and extended to other domains of image processing.

The proposed algorithm has demonstrated robust performance in detecting fake colorization images, but certain challenges still exist with regards to highly complex textures and very subtle color changes. Future work should thus aim at evolving more advanced feature extraction techniques that can sensitively detect fine-grained texture patterns without incurring greater false positive rates. This might entail integrating potential self-supervised learning whereby the model learns representations from large-scale unlabeled datasets, with the end goal of improving generalization in texture-rich images. Also, applying adversarial training through GANs could buttress robustness against advanced colorization techniques, further reducing false negatives in GAN-generated fakes.

Another major improvement is refining the attention mechanism within transformer-based architectures toward understanding spatial-contextuality. Thus, the model can follow gradual color transitions and inconsistencies beyond the scope of conventional CNN-based methods.

These improvements will provide an automated way of detection to the extent of being practically beneficial in areas such as digital forensics, media authentication, and restoration of images taken in history. Further, extending this study to consider multi-modal approaches that combine visual analysis and metadata could further fortify fake-image detection protocols.

More diverse datasets are also a suggested avenue for future work, introducing the complexity of higher resolution and

respect for images from standard and specialty sources that ensure robustness against various image styles and quality levels. Expanding benchmark evaluations to real-world manipulated datasets will give deeper insights into the model's adaptability and limitations, in turn paving the road for even more discriminative and generalizable fake image detection techniques.

REFERENCES

- [1] Nguyen, T., Li, Y., Ojha, U., Lee, Y.J. (2023). Visual instruction inversion: Image editing via image prompting. *Advances in Neural Information Processing Systems*, 36: 9598-9613.
- [2] Salman, K.A., Shaker, K., Al-Janabi, S. (2025). Detection of Fake Colorized Images based on Deep Learning. *International Journal of Image and Graphics*, 25(1): 2550002. <https://doi.org/10.1142/S0219467825500020>
- [3] Shashikala, S., Ravikumar, G.K. (2023). Ensemble deep learning fusion for detection of colorization based image forgeries. In *2023 2nd International Conference for Innovation in Technology (INOCON)*, Bangalore, India, pp. 1-9. <https://doi.org/10.1109/INOCON57975.2023.10101337>
- [4] Shafiq, H., Lee, B. (2023). Image colorization using color-features and adversarial learning. *IEEE Access*, 11: 132811-132821. <https://doi.org/10.1109/ACCESS.2023.3335225>
- [5] Ozcelik, F., Alganci, U., Sertel, E., Unal, G. (2020). Rethinking CNN-based pansharpening: Guided colorization of panchromatic images via GANs. *IEEE Transactions on Geoscience and Remote Sensing*, 59(4): 3486-3501. <https://doi.org/10.1109/TGRS.2020.3010441>
- [6] Gain, M., Raha, A.D., Debnath, R. (2024). CCC++: Optimized color classified colorization with Segment Anything Model (SAM) empowered object selective color harmonization. *arXiv preprint arXiv:2403.11494*. <https://doi.org/10.48550/arXiv.2403.11494>
- [7] Su, P.C., Huang, B.H., Kuo, T.Y. (2024). UFCC: A unified forensic approach to locating tampered areas in still images and detecting deepfake videos by evaluating content consistency. *Electronics*, 13(4): 804. <https://doi.org/10.3390/electronics13040804>
- [8] Agarwal, S., Jung, K.H. (2024). Forensic analysis and detection using polycolor model binary pattern for colorized images. *Multimedia Tools and Applications*, 83(14): 41683-41702. <https://doi.org/10.1007/s11042-023-16675-1>
- [9] Dehghani, A., Saberi, H. (2025). Generating and detecting various types of fake image and audio content: A review of modern deep learning technologies and tools. *arXiv preprint arXiv:2501.06227*. <https://doi.org/10.48550/arXiv.2501.06227>
- [10] Yin, H., Zhou, P., Xu, G., He, G., Li, L., Yao, J. (2024). A parallelizable global color consistency optimization algorithm for multiple images. *IEEE Geoscience and Remote Sensing Letters*, 22: 6000405. <https://doi.org/10.1109/LGRS.2024.3496730>
- [11] Khalil, A.H., Ghalwash, A.Z., Elsayed, H.A.G., Salama, G.I., Ghalwash, H.A. (2023). Enhancing digital image forgery detection using transfer learning. *IEEE Access*, 11: 91583-91594. <https://doi.org/10.1109/ACCESS.2023.3307357>
- [12] Terumalasetti, S., Reeja, S.R. (2024). Enhancing social media user's trust: A comprehensive framework for detecting malicious profiles using multi-dimensional analytics. *IEEE Access*, 13: 7071-7093. <https://doi.org/10.1109/ACCESS.2024.3521951>
- [13] Wu, L., Tong, T., Du, M., Gao, Q. (2020). Image colorization algorithm based on self-attention network. In *2020 Cross Strait Radio Science & Wireless Technology Conference (CSRSWTC)*, Fuzhou, China, pp. 1-3. <https://doi.org/10.1109/CSRSWTC50769.2020.9372464>
- [14] Krichen, M. (2023). Convolutional neural networks: A survey. *Computers*, 12(8): 151. <https://doi.org/10.3390/computers12080151>
- [15] Taye, M.M. (2023). Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions. *Computation*, 11(3): 52. <https://doi.org/10.3390/computation11030052>
- [16] Moreno, H., Gómez, A., Altares-López, S., Ribeiro, A., Andújar, D. (2023). Analysis of stable diffusion-derived fake weeds performance for training Convolutional Neural Networks. *Computers and Electronics in Agriculture*, 214: 108324. <https://doi.org/10.1016/j.compag.2023.108324>
- [17] Alkishri, W., Widyarto, S., Yousif, J.H., Al-Bahri, M. (2023). Fake face detection based on colour textual analysis using deep convolutional neural network. *Journal of Internet Services and Information Security*, 13(3): 143-155. <https://doi.org/10.58346/JISIS.2023.I3.009>
- [18] Wu, P., Wang, Z., Zheng, B., Li, H., Alsaadi, F.E., Zeng, N. (2023). AGGN: Attention-based glioma grading network with multi-scale feature extraction and multi-modal information fusion. *Computers in Biology and Medicine*, 152: 106457. <https://doi.org/10.1016/j.combiomed.2022.106457>
- [19] Gardella, M., Musé, P., Morel, J.M., Colom, M. (2021). Forgery detection in digital images by multi-scale noise estimation. *Journal of Imaging*, 7(7), 119. <https://doi.org/10.3390/jimaging7070119>
- [20] Zhang, Y., Lu, Y., Zhu, W., Wei, X., Wei, Z. (2023). Traffic sign detection based on multi-scale feature extraction and cascade feature fusion. *The Journal of Supercomputing*, 79(2): 2137-2152. <https://doi.org/10.1007/s11227-022-04670-6>
- [21] Lu, S., Ding, Y., Liu, M., Yin, Z., Yin, L., Zheng, W. (2023). Multiscale feature extraction and fusion of image and text in VQA. *International Journal of Computational Intelligence Systems*, 16(1): 54. <https://doi.org/10.1007/s44196-023-00233-6>
- [22] Li, X., Li, M., Yan, P., Li, G., Jiang, Y., Luo, H., Yin, S. (2023). Deep learning attention mechanism in medical image analysis: Basics and beyonds. *International Journal of Network Dynamics and Intelligence*, 2(1): 93-116. <https://doi.org/10.53941/ijndi0201006>
- [23] Wang, Y., Zhao, P., Zhang, Z. (2023). A deep learning approach using attention mechanism and transfer learning for electromyographic hand gesture estimation. *Expert Systems with Applications*, 234: 121055. <https://doi.org/10.1016/j.eswa.2023.121055>
- [24] Ai, Y., Liu, X., Zhai, H., Li, J., Liu, S., An, H., Zhang, W. (2023). Multi-scale feature fusion with attention

- mechanism based on CGAN network for infrared image colorization. *Applied Sciences*, 13(8): 4686. <https://doi.org/10.3390/app13084686>
- [25] Zhao, W., Zhu, J., Li, P., Huang, J., Tang, J. (2024). Attention mechanism-based generative adversarial networks for image cartoonization. *The Visual Computer*, 40(6): 3971-3984. <https://doi.org/10.1007/s00371-024-03404-4>
- [26] UrviOza, Gohel, B., Kumar, P. (2024). Evaluation of normalization algorithms for breast mammogram mass segmentation. *Procedia Computer Science*, 235: 2508-2517. <https://doi.org/10.1016/j.procs.2024.04.236>
- [27] Tankala, M.R., Rao, C.S. (2023). Image counterfeiting detection and localization using deep learning algorithms. *Revue d'Intelligence Artificielle*, 37(1): 191-199. <https://doi.org/10.18280/ria.370124>
- [28] Ragab, A.S., Taie, S.A., Abdelnaby, H.Y. (2023). Incorporating ensemble and transfer learning for an end-to-end auto-colored image detection model. *arXiv preprint arXiv:2309.14478*. <https://doi.org/10.48550/arXiv.2309.14478>
- [29] Xu, Q., Jia, S., Jiang, X., Sun, T., Wang, Z., Yan, H. (2024). MDTL-NET: Computer-generated image detection based on multi-scale deep texture learning. *Expert Systems with Applications*, 248: 123368. <https://doi.org/10.1016/j.eswa.2024.123368>
- [30] Rukundo, O. (2023). Normalized weighting schemes for image interpolation algorithms. *Applied Sciences*, 13(3): 1741. <https://doi.org/10.3390/app13031741>
- [31] Kumar, T., Mileo, A., Brennan, R., Bendeche, M. (2023). Image data augmentation approaches: A comprehensive survey and future directions. *arXiv preprint arXiv:2301.02830*. <https://doi.org/10.48550/arXiv.2301.02830>
- [32] Zhang, X., Zhang, X., Wang, W. (2023). Convolutional Neural Network. In: *Intelligent Information Processing with Matlab*. Springer, Singapore. https://doi.org/10.1007/978-981-99-6449-9_2
- [33] Bacanin, N., Jovanovic, L., Zivkovic, M., Kandasamy, V., Antonijevic, M., Deveci, M., Strumberger, I. (2023). Multivariate energy forecasting via metaheuristic tuned long-short term memory and gated recurrent unit neural networks. *Information Sciences*, 642: 119122. <https://doi.org/10.1016/j.ins.2023.119122>
- [34] Bungert, L., García Trillos, N., Murray, R. (2023). The geometry of adversarial training in binary classification. *Information and Inference: A Journal of the IMA*, 12(2): 921-968. <https://doi.org/10.1093/imaiai/iaac029>
- [35] Lei, K., Guo, P., Wang, Y., Zhang, J., Meng, X., Qian, L. (2023). Large-scale dynamic scheduling for flexible job-shop with random arrivals of new jobs by hierarchical reinforcement learning. *IEEE Transactions on Industrial Informatics*, 20(1): 1007-1018. <https://doi.org/10.1109/TII.2023.3272661>
- [36] Du, J., Jiang, B., Jiang, C., Shi, Y., Han, Z. (2023). Gradient and channel aware dynamic scheduling for over-the-air computation in federated edge learning systems. *IEEE Journal on Selected Areas in Communications*, 41(4): 1035-1050. <https://doi.org/10.1109/JSAC.2023.3242727>
- [37] Guo, Z., Yang, G., Wang, D., Zhang, D. (2023). A data augmentation framework by mining structured features for fake face image detection. *Computer Vision and Image Understanding*, 226: 103587. <https://doi.org/10.1016/j.cviu.2022.103587>
- [38] IMAGENET. ImageNet Large Scale Visual Recognition Challenge (ILSVRC 2017). <https://image-net.org/challenges/LSVRC/2017/index.php>
- [39] Deng, J., Dong, W., Socher, R., Li, L., Li, K., Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, USA, pp. 248-55.
- [40] Al-Ghanimi, H., Al-Ghanimi, A. (2025). Deep learning-driven medical image segmentation using generative adversarial networks and conditional neural networks. *Ingénierie des Systèmes d'Information*, 30(1): 287-300. <https://doi.org/10.18280/isi.300125>