Detecting Forgery Face Images Based on Local Feature Descriptors (SIFT)

***Marym Taha Alsaed, **Prof. Dr. Ali Mohammad Sahan**

**Technical College of Administration*

***Department of Information Technology Technical -College of Administration*

Middle Technical University

Baghdad, Iraq

DOI:10.37648/ijps.v20i01.013

[1]*Received: 18/07/2025; Accepted: 27/08/2025; Published: 31/08/2025*

Abstract

Digital facial images represent people's identity and play a pivotal role in documenting events and their visual communication. Fake faces are detected using image processing and feature extraction techniques. The proposed methodology relies on a series of pre-steps that enhance the quality of visual representation and reduce computational complexity. The image size is reduced by half at each stage, the image is converted to grayscale, and the features are extracted using the SFT algorithm. A pre-trained deep learning model, EfficientNet-B0, was also applied to classify face images into real and fake ones. The FaceForensics++ database was used. The model was tested using noise added to images to evaluate its robustness and accuracy under different noise conditions. It was tested on salt-and-pepper noise and Gaussian noise, and horizontal geometric displacement, a type of image noise, was applied. Satisfactory results were achieved, with the accuracy of the proposed technique reaching 99.85%.

Keywords—*SIFT; CLAHE; FaceForensics++; CNN; EfficientNet-B0; deep neural network; DoG; Gaussian noise; Horizontal misalignment noise; salt and pepper noise.*

1. Introduction

It should come as no surprise that images are essential to human perception given that vision is the most developed of the human senses. Images have evolved since the 1950s, There are many negative consequences of fake images, including security risks (faces are stolen and used for biometric authentication), reputational damage and personal injury (real people are superimposed on their faces), and a loss of trust in the media due to the widespread use of fake images[1, 2]. These negative impacts also affect justice and the law, and violate individual rights. It also has several advantages, as it is used as a tool to test the strength of artificial intelligence systems and their importance in medicine and education[3]. Fake images have several disadvantages that can be exploited as a loophole to detect images, such as inconsistent lighting, distorted edges and features, and differences in quality and accuracy where elements of different quality are combined[4]. There are many forgery tools that researchers use to forge faces and benefit from them, such as modifying facial expressions[5] such as anger, smiling, sadness, etc., and generating a new facial identity[6, 7], which refers to creating a completely fake human face that does not belong to any other person while maintaining the visual identity of the target, and re-representing facial expressions[8], example, moving facial expressions from one person to another, as well as attempting to record the source face's expressions and move them from one person's face to another .Previous studies on the subject can be summarized as follows:

Fatima Zahra Nourddine and a group of researchers have been working on combating the phenomenon of face forgery by using fake or printed images to deceive biometric verification systems. The aim of this study is to develop a model

[1]*How to cite the article:* Alsaed M.T., Sahan A.M (August 2025); Detecting Forgery Face Images Based on Local Feature Descriptors (SIFT); *International Journal of Professional Studies*; Jul-Dec 2025, Vol 20, 178-191; DOI: http://doi.org/10.37648/ijps.v20i01.013

that can distinguish between a real face and a fake one. The FaceNet algorithm was used to extract distinctive identity features, and the DenseNet201 network was used to extract visual features from images and then combine the features into a unified layer. The database used was the NUAA Imposter Database and they achieved an accuracy of 99%[9].

Darius Afchar and colleagues presented two lightweight networks that focus on the mesoscopic features of an image, with good computational efficiency, and demonstrated strong facial tampering detection capabilities across the Deepfake and Face2Face datasets in 2018, achieving 98% on Deepfake and 95% on Face2Face[10].

Vijay V. Chakole et al. published a research paper aiming to design a framework based on traditional machine learning techniques to detect fake images. The Difference of Gaussian algorithm was used with the Beltrami filter and these images were tested on Replay Attack images. The MLbFA framework achieved a classification accuracy higher than 0.11% compared to the best competing model (Edge-Net Autoencoder)[11].

The rapid advancement of AI-based face-spoofing technologies has made it easy to produce high-quality fake videos and images, raising widespread concerns about their use in malicious activities such as fraud, defamation, and spreading lies[12]. This undermines trust in media content, necessitating research and development into innovative fake detection technologies that offer more effective and accurate solutions to counter and mitigate the risks[13]. The use of fake images can influence public opinion and tarnish the reputation of political leaders and individuals[14]. AI- and deep learning-based classifiers designed to detect fake images can be deceived using advertising attacks[15].

2. Research Problem

High-quality fake movies and photographs may now be easily created because to the quick development of AI-based face-spoofing technologies, raising widespread concerns about their use in malicious activities such as fraud, defamation, and spreading lies. This undermines trust in media content, necessitating research and development into innovative fake detection technologies that offer more effective and accurate solutions to counter and mitigate the risks. The use of fake images can influence public opinion and tarnish the reputation of political leaders and individuals. Advertising attacks can fool classifiers that use AI and deep learning to identify phoney images.

3. Aim of Research

The goal is to propose a technology to detect digital face forgery to protect individuals from identity theft and reputational damage, thereby enhancing trust in society and reducing digital misinformation

4. Research Significance

The importance of this research is highlighted by its use of modern technologies to detect forgery, reduce the risk of identity theft, and protect security. This is because forgery poses social and legal risks that harm people's lives and threaten their privacy. Therefore, forgery detection technologies are vital for protecting social, cultural, and political stability, particularly in the Arab world, especially in light of rapid technological developments.

5. Image Forgery

As artificial intelligence and machine learning algorithms have improved significantly, sophisticated programs that can produce realistically accurate false photos have surfaced. There are four primary categories of forgery algorithms:

a) Face replacement: This relies on artificial intelligence to replace a specific person's face with another's, while preserving the original facial expressions, head orientation, lighting, and skin tone[16].

b) Facial expression modification: This refers to altering facial features, such as anger, smiling, and sadness, without replacing or altering the face. This means that the apparent emotions change, but the features remain the same[17].

c) Facial expression reenactment: This is used to transfer expressions from one person's face to another's while preserving the target's visual identity[17, 18].

d) New facial identity generation: This refers to the creation of a completely fake human face that is not related to any real person. This falls under the category of unconditional generation, meaning that it does not use an original image as a starting source. The face is generated from random noise[19, 20].

There are image forgery techniques used by AI models, which researchers can summarize below:

- Adversarial Generative Networks: These consist of two complementary neural networks, one of which is the generator, which generates new data, and the discriminator, which attempts to detect whether the data is real or fake. There are many types of generative networks, such as the Deep Convolutional GAN (DCGAN), the Wasserstein distance GAN (WGAN), and others[21, 22].

- Diffusion Model: This model consists of two main stages: the destruction stage, which gradually adds noise to the data, and the generation stage, which is the reverse of the first stage, gradually removing this noise, thus generating new data[23-25].

- Auto-variational Encoder: This deep learning model performs two steps: the first is encoding, where the original data is entered and converted into a compressed and abstract latent representation. The second is decoding, which takes a sample from this representation and generates a new image similar to the original. This is crucial for generating samples of acceptable quality[26, 27].

6. The Proposed Method

A. Image Enhancement

Enhancement is an important step in the field of face forgery detection to improve image quality and increase the clarity of fine details, which plays a pivotal role in enhancing the effectiveness of deep learning models. CLAHE[28, 29], which is short for Contrast Limited Adaptive Histogram Equalization, is the technique used in this research for contrast enhancement of images. It is an improvement over the traditional HE (Histogram Equalization) technique, which affects each image. CLAHE is applied to small squares within the image, and then these squares are merged. One of the characteristics of this technique is that it provides contrast on pixels instead of general enhancement to highlight fine edges and distorted textures[30]. It can be represented by the following equation[28]:

$$H'(i) = min(H(i), C) \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

where H(i) is the local cumulative distribution and C is the clip limit value. After cutting, the excess is redistributed equally over the levels and then H(i) is used to calculate the transformation[31]:

$$k^s = H'(k)\frac{L-1}{w \times \mathrm{h}} \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

B. Scale-Invariant Feature Transform(SIFT)

It is an algorithm for extracting points of interest and local features from images. It has several characteristics, the most important of which is that it is stable across scale and rotation changes and captures unique local features such as corners, edges, and craters[32]. It operates in several stages. First, it detects extreme points using the octave, then passes them through various Gaussian filters, and calculates the difference of Gaussian (DoG) layers[33]. The points representing the local maximum and minimum values are then identified. In the next stage, weak points are filtered out by deleting unstable points and retaining only those with a strong response. The rotation direction is then calculated, and a stable direction is then calculated based on the local gradient in the neighborhood. Finally, a description is created by taking a window around each key point, dividing it into cells, and calculating the gradient direction within each cell[34].

Feature points are detected by applying a Gaussian filter with different changes of the standard deviation represented by the following equation[35]:

$$G(x,y,\sigma)=\frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}} \ldots\ldots\ldots\ldots\ldots..(3)$$

Where (x,y): pixel coordinates, sigma: standard deviation, controls the level of blur, G(x,y,σ) is the blur output of the image.

Then the Difference of Gaussians is calculated using the following equation[36]:

$$D(x,y,\sigma) = (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y)........(4)$$

Where ($D(x,y,\sigma)$) represents the difference between the two Gaussians and ($I(x,y)$) represents the original image.

Then we can determine the exact locations of the points using the following equation[37]:

$$D(X)=D+\frac{D\partial}{X\partial}X^T+\frac{1}{2}x\frac{D^2\partial}{X^2\partial}X^T \ldots\ldots\ldots\ldots.(5)$$

where D is the value of the DoG function at the potential point, X= $(x,y,\sigma)^T$ It is a displacement from the original site.

Finally, the gradient is calculated using the following equation[33]:

$$\theta(x,y)=\arctan\left(\frac{L(x,y+1)-L(x,y-1)}{L(x+1,y)-L(x-1,y)}\right)..........................(6)$$

Where $L(x,y)$ is the value of the image after applying Gaussian blur and θ(x,y) is the orientation.

C. Deep Neural Network

Based on the architecture of the human brain, it is composed of several layers of neural networks. Complex patterns are extracted from data using these networks [38]. A subfield of machine learning called "deep learning" uses artificial neural networks. Multiple processing layers are characteristics of neural networks.[39].

The components of a deep neural network are:

1- Input layer: Receives raw data.

2- Hidden layers: These layers learn complex representations or features, and each layer consists of a number of nodes.

3- Output layer: Produces the final result (such as classifying real from fake).

Consisting of multiple layers of neural units, they are inspired by the structure of the human brain. These networks are used to extract complex patterns from data. Activation functions (ReLU) (Rectified Linear Unit), sigmoid, and tanh (Hyperbolic Tangent) are a few of the primary parts of neural networks, and others. Their benefits include preserving positive values and being fast in calculations[40]. Neural networks rely on learning by experience. The first stage involves the forward propagation process, where data is passed from one layer to another The loss function is also used to calculate the error in the model's prediction[41]. The lower the loss value, the more accurate the model is. Backpropagation is the stage following the loss calculation and is used to adjust the network weights, thus reducing the error rate. The weights are then optimized using optimization algorithms such as SGD (Stochastic Gradient Descent), Adam, and RMSprop to determine the appropriate adjustment amount. There are also several additional optimization techniques such as Dropout, which disables some neurons during training to prevent overlearning; Batch Normalization, which speeds up training and makes the network more stable; and Regularization, which prevents overlearning. L1 forces some weights to become zero, while L2 prevents some weights from becoming too large[42, 43].

AdamW optimizer was used with a learning rate of 0.001 and batch size was set at 32 with 25 epochs. Data was divided into 80% training and 20% testing. The model was evaluated on the testing set to classify the input images into real and fake images.

7. Testing the Technology on Noise

The proposed technology has been tested to resist noise that negatively affects image quality. There are many impurities that can occur in images in nature. Some noise may occur due to the camera sensor in low light, problems in the sensor circuits, and also digital transmission errors. Data may be lost as a result of sending data over wireless networks, camera shake, sudden movement, or video compression. The most common types of image noise that can be tested are:

- **Gaussian noise**

It is one of the most common types of noise in digital images and results from thermal disturbances in sensors or electronic interference. It is widely used in experiments and modeling. Its statistical properties include being unbiased, being centered around zero, and being controllable through the standard deviation to determine the intensity of light[44].

It affects the image by smoothing out fine details and hiding weak edges, making the image appear blurry, making it difficult to detect fake images using traditional methods[45].

It can be represented by the following equation[35]:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} . e^{\frac{(x-\mu)^2}{2\sigma^2}} \text{..................}(7)$$

Where μ represents the mean, while σ represents the standard deviation, which determines the spread of values around the mean.

This type of noise is removed by a Gaussian filter, where the noise is removed without significant loss of data. This filter can be represented by the following equation[35]:

$$(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \cdots\cdots\cdots\cdots\cdots\cdots (8)$$

while σ represents the standard deviation.

- **Salt and Pepper Noise**

It is known as one of the most popular types of noise, where noise appears in the form of black dots (value 0) and white dots (value 255) scattered throughout the image. It was given this name because it resembles grains of salt and pepper that are distributed randomly[46], and it is expressed by the following equation[44]:

$$p(x) = \begin{cases} p_a & for\ x = a \\ p_b & for\ x = b \\ 0 & otherwise \end{cases} \ldots\ldots\ldots\ldots\ldots\ldots (9)$$

This type of noise can be removed using a median filter, which replaces the central pixel value with the most average value in the vicinity, making it effective at removing outliers. It can be represented by the following equation[47]:

$$M(r,s) = med_{(a,b)\in R_{uv}}\{I(u,v)\} \ldots\ldots\ldots\ldots\ldots\ldots (10)$$

Where $I(u,v)$ represents the value of the original image at point (u,v), and R_{uv} represents a local frame around point (r,s).

- **Horizontal Misalignment Noise**

This type of noise refers to the displacement of a horizontal part of the image or one of its components from its normal position as a result of errors in temporal and spatial synchronization. This usually leads to a loss of consistency between parts of the image[48]. The distortion appears in the form of a gap or interruption in the consistency between the right and left halves of the image, as the image is divided into two halves $I_{right}\ and\ I_{left}$As in the following equation[49]:

$$I_{right\_shifted}(x,y) = (\delta_x, y - I_{right}) \ldots\ldots\ldots (11)$$

The final image is collected using the following equation[49]:

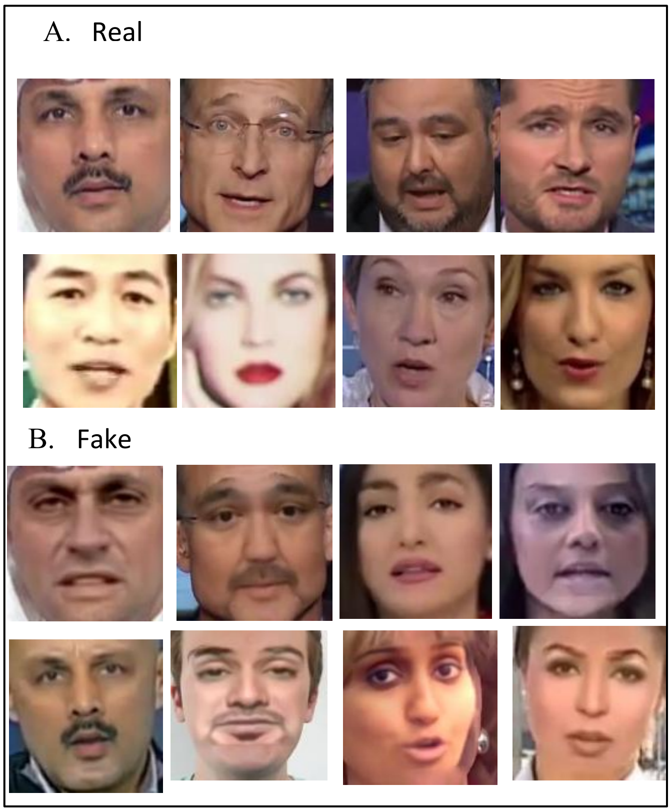
$$I_{misaligned}(x,y) = \begin{cases} I(x,y), & x < \frac{w}{2} \\ I(x+\delta,y), & x \geq \frac{w}{2} \quad (edges\ offset\ with\ ) \end{cases} \ldots (12)$$

Where $I(x,y)$ is the original grayscale image, δ_x is the shift value, and w is the width of the image, This type of noise is removed using internal paint.

8. Dataset

The technology was tested on the FF++ standard database, one of the most popular databases used in fake detection research. It was developed by a research team at the Technical University of Munich. This database contains fake versions of clips created using four main algorithms (DeepFakes, Face2Face, FaceSwap, NeuralTextures)[50]. It has evolved to include several versions. Among these versions, a modified and pre-configured version with a fixed size of (128x128) pixels, called “FF++ 128x128”, emerged. Each video was converted into a set of frames, and then the face inside the frame was detected using a specific algorithm only, without the background. The image was then resized to (128x128) pixels, and then organized into folders classified by type, real and fake. “FF++ 128x128” database consists of 16,000 images divided into real and fake. This data was separated into training and test data at ratios of 80% and 20%, respectively. Figure (1) represents samples from the database used[51, 52].

Figure (1): Sample from FF++ 128x128 database



9. Results

The proposed technique was implemented using the Python programming language, and tests were conducted in a Windows 10 operating system. A computer with a 10th-generation Intel Core i7-10610U processor, an NVIDIA Quadro P520 graphics processor, 16 GB of RAM, and a 1.80 GHz processor was used.

Here we will review the results of applying the proposed technology in all its stages on the database FF++ 128x128, which contained 16,000 images separated into real and fake and tested with a ratio of 80% training and 20% testing.

Table 1 displays the outcomes of applying the CLAHE algorithm to improve image quality. We notice an increase in the discrimination accuracy by (1.83%), which occurred as a result of enhancing the edges of the image and highlighting the fine details.

Table 1: results of using the CLAHE

before CLAHE	After CLAHE	*Time (second)*
97.21	99.04	*31,680*

The SIFT technique was applied at four levels to the database improved by using CLAHE. The results demonstrated its strong effectiveness in accelerating the processing and classification process and achieving a significant increase in accuracy.

The accuracy reached 99.15% in the first level and increased to 99.85% in the second level, but it collapsed in the third and fourth levels due to the small size of the image, which causes weakness in the data transferred and entered into the model.

It should also be mentioned that the image size at the first level was (64×64) and at the second level it became (32×32) and the third and fourth levels (16×16,8×8) respectively. Table (2) shows the results of using this technology and its effect on processing time and accuracy at four levels.

Table 2: The effect of using the SIFT technique on the proposed model

The Level	**accuracy**	**Time**
First level	99.05	15480
Second level	99.85	8520
Third level	79.64	4740
Fourth level	55.61	1980

And when testing the technology on noise, the results were impressive. The technology showed great superiority in resisting noise. Table (3) shows its results in resisting Gaussian noise, salt and pepper noise, and Horizontal misalignment noise

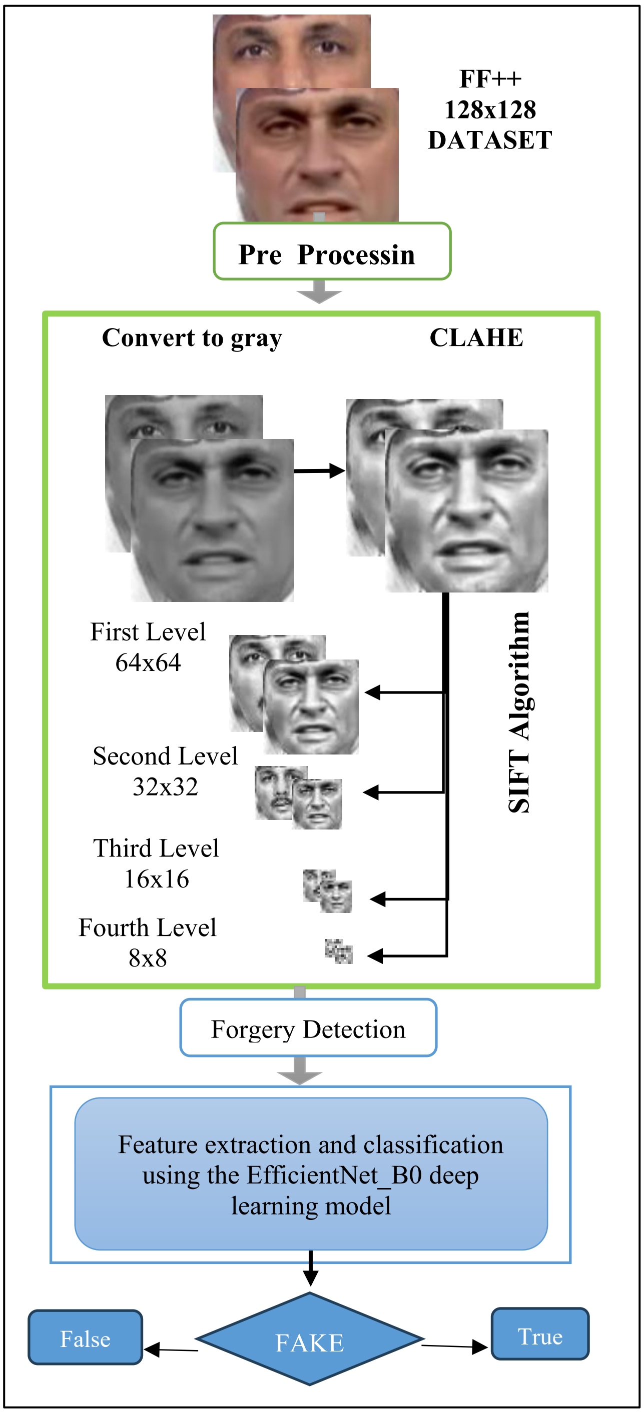
Table (3): Impact of noise on the proposed technique

Noise Name		**Ratio**	**Accuracy**
Gaussian noise	Noise	0.30	91.58
	With filter	0.30	92.89
	With SIFT	0.30	94.92
	Noise	0.15	94.64
	With filter	0.15	95.88
	With SIFT	0.15	97.15
	Noise	0.05	95.05
	With filter	0.05	96.66
	With SIFT	0.05	97.93
salt and pepper noise	Noise	0.05	94.57
	With filter	0.05	95.43
	With SIFT	0.05	96.69
	Noise	0.02	95.59
	With filter	0.02	96.33
	With SIFT	0.02	98.36
Horizontal misalignment noise	Noise	10 pixel	95.45
	With filter	10 pixel	96.15
	With SIFT	10 pixel	97.82
	Noise	5 pixel	96.44
	With filter	5 pixel	97.18
	With SIFT	5 pixel	98.82

It is noted from the above results that the SIFT technology outperforms the image filters used in resisting noise.

It achieved an increase in accuracy of (2.03%) at a noise factor of 0.3, while at a noise factor of 0.15 the increase reached (1.27%) and also the increase reached (1.07%) at a noise factor of 0.05. Figure (2) shows the Stages of the proposed technique for identifying fake images.

Figure (2) Stages of the proposed technique for identifying fake images



The proposed technique was also tested in image rotation resistance. It was tested at 4 degrees of rotation (30°, 60°, 90°, 180°). It appears that the accuracy of discrimination increases at angles that are multiples of 90. This is due to the fact that these angles do not cause radical changes in the spatial distribution of facial features. Table (4) shows the performance of the proposed technology in rotation.

Table (4) the performance of the proposed technology in rotation

Rotation degree	**accuracy**	
	Before the proposed technique	After the proposed technique
30°	97.16	98.71
60°	94.15	95.58
90°	95.93	98.47
180°	96.07	98.39

10. Conclusion and Suggestions

In this study, a new technique for detecting tampered or fake images is proposed, focusing on striking a balance between reasonable processing economy and excellent detection accuracy. According to experimental findings, the suggested approach significantly enhances detection performance while maintaining short processing time. This indicates that the method not only effectively distinguishes real images from fakes but is also suitable for real-time or large-scale applications where response time is critical. Overall, the results confirm that the proposed technique represents a promising approach for reliable and efficient detection of fake images and could form the basis for further improvements in the field of multimedia forensics.

The use of CLHAE technology improved image quality and highlighted edges, followed by the use of SIFT at four levels, which contributed to reducing the image size by half, extracting important features from the image, and neglecting unnecessary features. This contributed significantly to increasing accuracy and reducing processing time by almost half at each stage. It was proven that the second level was the highest degree of accuracy, level 4 takes the least amount of processing time, but accuracy collapsed at this level due to the small size of the image and the weakness of the data it contained.

Since the technology reduces processing time by up to half, it is recommended to use the proposed technology in real-time systems to benefit from security applications and smart surveillance. It can also be used to detect forged identical twins and also to detect fake faces from videos.

11. References

Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018). Mesonet: a compact facial video forgery detection network. In *2018 IEEE International Workshop on Information Forensics and Security (WIFS)* (pp. 1–7). IEEE.

Ali, K., & Hughes, C. E. (2019). *All-in-one: Facial expression transfer, editing and recognition using a single network* [Unpublished manuscript].

Anderson, J., & Akram, N. J. (2024). Denoising Diffusion Probabilistic Models (DDPM) Dynamics: Unraveling Change Detection in Evolving Environments. *International Conference on Systems, 10*(1), 1–10.

Ba, K., et al. (2024a). [Title omitted for brevity].

Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning*. Springer.

Boediman, E. P. J. (2025). Exploring the impact of deepfake technology on public trust and media manipulation: A scoping review. *Journal of Fake Media Analysis, 19*(2), 131–152.

Boutros, F., Klemt, M., Fang, M., Kuijper, A., & Damer, N. (2023). Exfacegan: Exploring identity directions in GAN's learned latent space for synthetic identity generation. In *2023 IEEE International Joint Conference on Biometrics (IJCB)* (pp. 1–10). IEEE.

Buades, A., Coll, B., & Morel, J.-M. (2005). A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (Vol. 2, pp. 60–65). IEEE.

Chakole, V. V., & Dixit, S. R. (2025). MLbFA: A Machine Learning-Based Face Anti-Spoofing Detection Framework under Replay Attack. *International Journal of Electronic and Computer Engineering Systems, 16*(4), 321–332.

Chan, R. H., Ho, C.-W., & Nikolova, M. (2005). Salt-and-pepper noise removal by median-type noise detectors and detail-preserving regularization. *IEEE Transactions on Image Processing, 14*(10), 1479–1485.

Chen, X., et al. (2016). *Variational lossy autoencoder*. arXiv preprint.

Choi, Y., Choi, M., Kim, M., Ha, J.-W., Kim, S., & Choo, J. (2018). StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 8789–8797).

Ciftci, U. A., Demir, I., & Yin, L. (2020). FakeCatcher: Detection of synthetic portrait videos using biological signals. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 44*(12), 8719–8732.

Doersch, C. (2016). *Tutorial on variational autoencoders*. arXiv preprint.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Hinton, G. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Durall, R., Keuper, M., & Keuper, J. (2019). Watch your up-convolution: CNN based generative deep neural networks are failing to reproduce spectral distributions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 7890–7899).

Fang, S., et al. (2025). [Title omitted for brevity].

Fernandes, S., Raj, S., Ortiz, E., Vintila, I., Salter, M., Urosevic, G., & Jha, S. (2019). Predicting heart rate variations of deepfake videos using neural ODE. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops* (pp. 1–9).

Frank, J., Eisenhofer, T., Schönherr, L., Fischer, A., Kolossa, D., & Holz, T. (2020). Leveraging frequency analysis for deep fake image recognition. In *International Conference on Machine Learning* (pp. 3247–3258). PMLR.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems, 27*.

Gui, J., Sun, Z., Wen, Y., Tao, D., & Ye, J. (2021). A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE Transactions on Knowledge and Data Engineering, 35*(4), 3313–3332.

Gu, Z., et al. (2022). [Title omitted for brevity].

He, K., Zhang, X., Ren, & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770–778).

Hernandez-Ortega, J., Tolosana, R., Fierrez, J., & Morales, A. (2020). DeepFakes detection based on heart rate estimation: A generalizable approach. *arXiv preprint arXiv:2010.00400*.

Huang, T., Yang, G., & Tang, G. (1979). A fast two-dimensional median filtering algorithm. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 27*(1), 13–18.

Jähne, B. (2005). *Digital image processing*. Springer.

Jain, A. K. (1989). *Fundamentals of digital image processing*. Prentice Hall.

Jiang, L., Li, R., Zhang, Z., Fang, S., & Ma, C. (2024). EmojiDiff: Advanced Facial Expression Control with High Identity Preservation in Portrait Generation. *arXiv preprint arXiv:2403.12345*.

Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2017). Progressive growing of GANs for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*.

Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4401–4410).

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems, 25*.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444.

Li, L., Bao, J., Zhang, T., Yang, H., Chen, D., Wen, F., & Guo, B. (2020). Face X-ray for more general face forgery detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5001–5010).

Li, Y., Yang, X., Sun, P., Qi, H., & Lyu, S. (2020b). Celeb-DF: A large-scale challenging dataset for deepfake forensics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3207–3216).

Lindeberg, T. (1998). Feature detection with automatic scale selection. *International Journal of Computer Vision, 30*(2), 79–116.

Lindeberg, T. (2012). Scale invariant feature transform. *Scholarpedia, 7*(5), 10491.

Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Proceedings of the Seventh IEEE International Conference on Computer Vision* (Vol. 2, pp. 1150–1157). IEEE.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision, 60*(2), 91–110.

Loy, C. C. (2018). *ReenactGAN: Learning to Reenact Faces via Boundary Transfer*. arXiv preprint arXiv:1807.11079.

Lundberg, E., & Mozelius, P. (2025). The potential effects of deepfakes on news media and entertainment. *AI and Society, 40*(4), 2159–2170.

Luo, Y., Zhang, Y., Yan, J., & Liu, W. (2021). Generalizing face forgery detection with high-frequency features. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 16317–16326).

Man, Q., Cho, Y.-I., Gee, S.-J., Kim, W.-J., & Jang, K.-A. (2024). GAN-Based High-Quality Face-Swapping Composite Network. *Electronics, 13*(15), 3092.

Masi, I., Killekar, A., Mascarenhas, R. M., Gurudatt, S. P., & AbdAlmageed, W. (2020). Two-branch recurrent network for isolating deepfakes in videos. In *European Conference on Computer Vision* (pp. 667–684). Springer.

Miao, C., et al. (2022). HFI-Net: Hierarchical feature integration network for face forgery detection. *arXiv preprint arXiv:2203.12145*.

Mikolajczyk, K., & Schmid, C. (2005). A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*(10), 1615–1630.

Musa, P., Al Rafi, F., & Lamsani, M. (2018). A Review: Contrast-Limited Adaptive Histogram Equalization (CLAHE) methods to help the application of face recognition. In *2018 Third International Conference on Informatics and Computing (ICIC)* (pp. 1–6). IEEE.

Nguyen, H. H., Yamagishi, J., & Echizen, I. (2024). [Title omitted for brevity].

Nielsen, M. A. (2015). *Neural networks and deep learning*. Determination Press.

Nightingale, S. J., Wade, K. A., & Mind, and Media. (2022). Identifying and minimising the impact of fake visual media: Current and future directions. *Journal of Visual Media Integrity, 1*, e15.

Nirkin, Y., Wolf, L., Keller, Y., & Hassner, T. (2021). DeepFake detection based on discrepancies between faces and their context. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 44*(11), 6111–6121.

Ognyanova, K., Lazer, D., Robertson, R. E., & Wilson, C. (2020). Misinformation in action: Fake news exposure is linked to lower trust in media, higher trust in government when your side is in power. *Harvard Kennedy School Misinformation Review, 1*(4).

Olivier, N., Baert, K., Danieau, F., Multon, F., & Avril, Q. (2023). Facetunegan: Face autoencoder for convolutional expression transfer using neural generative adversarial networks. *Computers & Graphics, 110*, 69–85.

Qi, H., Guo, Q., Juefei-Xu, F., Xie, X., Ma, L., & Feng, W. (2020). DeepRhythm: Exposing deepfakes with attentional visual heartbeat rhythms. In *Proceedings of the 28th ACM International Conference on Multimedia* (pp. 4318–4327).

Qian, Y., Yin, G., Sheng, L., Chen, Z., & Shao, J. (2020). Thinking in frequency: Face forgery detection by mining frequency-aware clues. In *Proceedings of the European Conference on Computer Vision* (pp. 86–103).

Reza, A. M. (2004). Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. *Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology, 38*(1), 35–44.

Rochow, A., Schwarz, M., & Behnke, S. (2024). Fsrt: Facial scene representation transformer for face reenactment from factorized appearance head-pose and facial expression features. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 7716–7726).

Rossner, M., & Yamada, K. M. (2004). What's in a picture? The temptation of image manipulation. *The Journal of Cell Biology, 166*(1), 11–15. https://doi.org/10.1083/jcb.200406019

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*(6088), 533–536.

Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning* (pp. 6105–6114). PMLR.

Thies, J., Zollhofer, M., Stamminger, M., Theobalt, C., & Nießner, M. (2016). Face2Face: Real-time face capture and reenactment of RGB videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2387–2395).

Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society, 6*(1), 2056305120903408.

Verdoliva, L. (2020). Media forensics and deepfakes: An overview. *IEEE Journal of Selected Topics in Signal Processing, 14*(5), 910–932.

Wang, J., & Chow, K. P. (2023). [Title omitted for brevity].

Wang, S., et al. (2022). M2TR: Multi-modal multi-scale transformers for deepfake detection. In *Proceedings of the IEEE International Conference on Image Processing* (pp. 1–5).

Wang, Z., et al. (2023). [Title omitted for brevity].

Wu, H., Zhou, J., & Tian, J. (2020). Robust image forgery detection against transmission over online social networks. *IEEE Transactions on Information Forensics and Security, 15*, 3811–3825.

Xiang, W., Gao, P., & Peng, Q. (2015). Robust multiview three-dimensional video communications based on distributed video coding. *IEEE Systems Journal, 11*(4), 2456–2466.

Yang, X., Li, Y., & Lyu, S. (2019). Exposing deep fakes using inconsistent head poses. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 8261–8265).

Yin, B., et al. (2023). [Title omitted for brevity].

Zhang, J., et al. (2020). Freenet: Multi-identity face reenactment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5326–5335).

Zhang, K., et al. (2022). [Title omitted for brevity].

Zhang, Y., et al. (2024). [Title omitted for brevity].

Zhao, H., Zhou, W., Chen, D., Wei, T., Zhang, W., & Yu, N. (2021). Multi-attentional deepfake detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2185–2194).

Zhou, P., Han, X., Morariu, V. I., & Davis, L. S. (2017). Two-stream neural networks for tampered face detection. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 1831–1839).

Zuama, L. R., Susanto, A., Santosa, S., Gan, H.-S., & Ojugo, A. A. (2025). High-Performance Face Spoofing Detection using Feature Fusion of FaceNet and Tuned DenseNet201. *Journal of Facial Analysis and Identification Technologies, 1*(4), 385–400.