An Effective Technique for Deepfake Video Detection

***Baneen Musa Mahdi, *Prof. Dr. Ali Mohammad Sahan**

**Technical College of Management - Baghdad,*
Middle Technical University,
Baghdad, Iraq

DOI:10.37648/ijps.v20i01.013

[1]*Received: 18/07/2025; Accepted: 01/09/2025; Published: 07/09/2025*

Abstract

As fake videos cause numerous problems affecting people's lives in various fields, they have received increasing attention. In this paper, we present a successful method for detecting fake videos based on the Scattered Wavelet Transform (SWT) and a pre-trained deep learning model, EfficientNet-B0. Several experiments were conducted on the fake video detection dataset to evaluate the effectiveness of the proposed method. The Deepfake Detection database (DFD) database was used. The model was tested by adding noise to images to evaluate its robustness and accuracy under various noise conditions. It was tested on salt-and-pepper noise and white Gaussian noise, and horizontal misalignment noise, a type of noise commonly used in fake video detection, was applied. 98% accuracy was achieved using the noise.

Keywords— *EfficientNet B0; Scatter Wavelet transform; deepfake video detection; Deepfake Detection database.*

1. Introduction

Deepfake technologies have becoming widely used to create incredibly lifelike fake videos, posing a direct danger to cybersecurity, media credibility, and privacy [1].These videos provide new difficulties for digital verification systems since they are frequently utilized for identity theft, defamation, and the dissemination of misleading information [2].Even with the advancement of deep learning-based detection algorithms, many still have trouble making generalizations when examining compressed or realistic films, especially since the underlying manipulation methods keep changing [3]. According to recent research, temporal and spatial irregularities in frame coherence can be learned from real data without the use of phony videos for training, and they are promising markers for identifying manipulation [4].Improved detection performance has also been a result of convolutional neural networks' (CNNs) integration with transformers and attention processes [5]. Additionally, on large datasets like FF++ and DFDC, ensemble-based models that incorporate many trained networks, including EfficientNet and XceptionNet, have demonstrated efficacy in attaining high accuracy, particularly when employing dual training procedures like Siamese training [6].

The following is a summary of the relevant work:

> Researchers Shouhong Ding, Jilin Li et al. [7] developed a method based on learning local relationships between facial parts by analyzing the similarity between different facial regions in the face to detect inconsistencies using the Local Relation Learing algorithm and tested it on the FaceForensics++ database and achieved an accuracy of 91.47%.

[1] *How to cite the article:* Mahdi B.M., Sahan A.M. (September 2025); An Effective Technique for Deepfake Video Detection; *International Journal of Professional Studies*; Jul-Dec 2025, Vol 20, 192-201; DOI: http://doi.org/10.37648/ijps.v20i01.014

researchers Alexandros Haliassos, Konstantinos et al. [8] proposed a pre-trained neural network for lipreading. The model consists of ResNet-18 for feature extraction and lip analysis using Multi-Scale TCN, achieving an accuracy of 97.1%.

Researchers Junyi Cao, Chao Ma et.al.[9] in their research on face image forgery proposed a Reconstruction-Classification Learning (RECCE) model using reconstruction and classification to understand the features of faces. The model consists of an Encoder-Decoder Network and a graphical model for relationship analysis, which enhances the model's ability to detect inconsistencies and achieved an accuracy of 83.25% on the WildDeepfake database.

Researchers Haixu Song and Shiyu Huang. [10] analyzed the performance of deepfake detection tools on images generated from diffusion models. They created a database known as DeepFakeFace (DFF) containing 90,000 fake images and 30,000 real images and found that the best performance was with InsightFace.

Peter Peer et.al.[11] detected deep face fakes using Deep Information Decomposition (DID) in addition to traditional CNN networks ResNet-50 and EfficientNet-V2-L on the FaceForensics++ (FF++), Celeb-DF V2 and DFFD (DeepFake Detection Dataset) databases and achieved an accuracy of 0.970, outperforming CFFs (0.742) and NoiseDF (0.759).

The DuB3D technique was created by Lichuan Ji et al. [12] and is based on a bi-branch that combines appearance and motion to create a reliable model for identifying phony videos. The model was trained using the GenVidDet dataset, which included 2.66 million videos from authentic sources like InternVid and fakes from potent generators like Opnen-Sora/ModelScope. The Video Swin Transformer served as the foundation for the algorithm, which cleverly combined motion and visual data. It achieved 79.19 percent accuracy on unseen evaluation data and 96.77% accuracy in a closed setting. The significance of motion cues in fosteringdetection generalization to today's misleading videos is emphasized by this work.

2. Video Forgery

Over the past few years, facial manipulation technologies have advanced significantly and become widely available. They are even capable of accurately and easily altering faces in films without the need for specialized equipment. A person's identity in a video can be completely altered, or their facial expressions can be mimicked using tools such as deepfakes and facial expression changers. While these technologies are useful in areas such as special effects and filmmaking, they pose a significant risk in terms of private viewing, the spread of misinformation, and the falsification of videos. These numerous videos demonstrate the veracity of media information and can be used in chants and cyberattacks, highlighting the urgent need to create intelligent systems capable of accurately and successfully identifying manipulation. [13]

The following are the main risks of video falsification technologies: [14]

- Spreading misinformation and negative media influence: The public is misled, and their trust in official sources of information is undermined when fake videos are used to create false political claims or distort media content.
- Financial fraud and torture: Advanced fraud schemes, such as fake video calls impersonating government officials, have used deepfake technologies.
- Privacy invasion and personal abuse: When real people's faces are used to create pornographic material without their permission, it directly violates their privacy and negatively impacts victims psychologically and socially, undermining the democratic process and influencing elections. Deepfake technology poses a threat to political stability and the integrity of elections, as it can be used to influence politicians' statements or spread misinformation during campaigns.
- Difficulty distinguishing between real and fake: With the rapid advancement of forgery technologies, fake videos are becoming more realistic, making it difficult even for experts to verify their legitimacy. This makes detection even more challenging.

- Beyond traditional detection systems: Although generation tools are evolving more rapidly than traditional detection tools, they are less effective and require more complex algorithms to address today's risks.

3. Wavelet Scattering Transform

The SWT is a non-trainable hierarchical analytic technique that aims to extract stable and invariant representations of images and signals by applying a series of complex wavelet transforms followed by absolute value-based nonlinear operations and then local smoothing. This technique was developed by researcher Stefan Malla and is widely used in fields such as image classification, pattern recognition, and medical analysis because it preserves high-frequency structures and ensures representation stability under the influence of local distortions [15]. The algorithm is based on implementing a multilevel wavelet transform, then applying a nonlinear modulus operator, followed by a low-pass filter, which produces coefficients known as scattering coefficients. These coefficients are similar in structure to the features extracted by convolutional neural networks, but without the need for training [16, 17].

4. Deep Learning

Multilayer neural networks are used in deep learning, a state-of-the-art artificial intelligence technology. In this sector, Convolutional Neural Networks (CNNs) are essential tools [18]. CNNs are excellent in visual analysis-based object recognition and classification [19]. The input layer, which receives data, including images of particular dimensions, is the first of many interconnected layers that make up these networks.

After that, data is handled via hidden layers: Convolutional kernels are used by convolutional layers to extract various picture information and carry out mathematical calculations. After that, the feature maps are sent to deeper layers for further examination [20].

Feature maps can be made simpler without losing important information by using pooling layers like max pooling or average pooling. The lowering of dimensionality improves computing performance. [21].

The data is then sent to the output layer, where the model produces final predictions, some of which may be classifications based on the study of the network [22]. The basic parts of CNNs are shown in Figure 1.

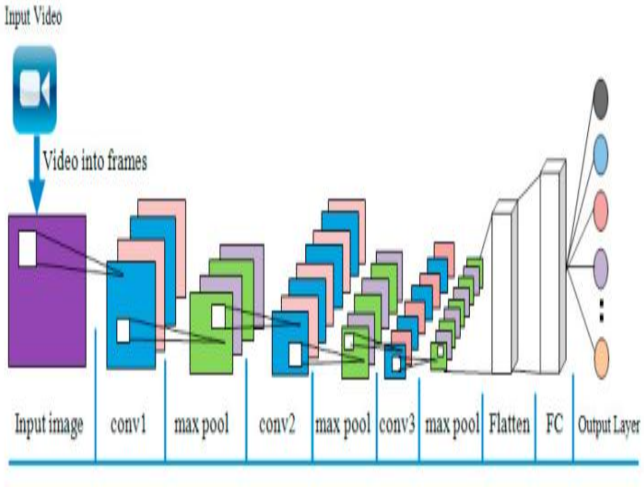

Fig. 1. Basic components of Convolutional Neural Networks [23].

By efficiently managing and attaining optimal performance, activation functions such as ReLU, sigmoid, and tanh are essential to network optimization [24]. By reducing overfitting, enhancing network stability, and averting the negative consequences of overfitting, regularization techniques like dropout improve model performance [25].

5. Experimental Part

The proposed technique was evaluated through extensive experiments on a DFD database. The experiments were conducted using Python code, a 2.60 GHz laptop processor, and 16.0 GB of RAM.

6. Used Database

This paper uses the Deepfake Detection Dataset (DFD), a sophisticated database designed to support research into face detection in videos, particularly those resulting from deepfake technologies. Developed byGoogle in collaboration with Jigsaw, it is one of the first and most important databases designed to enable researchers to train and test algorithms capable of accurately distinguishing between real and deepfake videos. The DFD database was built by recording high-quality videos of a group of actors and participants of various genders, ethnicities, and ages, who provided explicit consent for the use of their images in research. After collecting the original videos, deepfake versions were created using open-source tools based on face-swapping techniques, while maintaining a balance between real and deepfake videos to ensure fair representation. The database contains approximately 363 real videos and 3,068 deepfake videos, all produced with professional-quality filming under a variety of lighting conditions and shooting angles. These videos were recorded under the supervision of specialists and using precise digital photography techniques to ensure the quality of the content.[26] Figure 2 represents some samples from a DFD database.

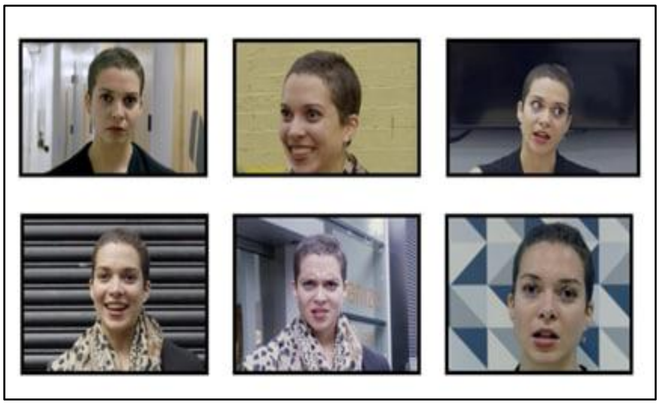

Fig.2. Some samples from DFD database [27].

7. Testing the Technology on Noise

The proposed technique was tested for noise suppression. Several types of noise were added to images to test the proposed technique's ability to handle various image distortions. These included salt and pepper noise, white Gaussian, and horizontal misalignment. To address these distortions, a set of specialized noise removal filters was used: a Gaussian filter, a Telea Inpainting Method, and a Median filter, as follows:

- **salt and pepper noise**

Salt and pepper noise is one of the most prominent types of random noise affecting digital images [28]. This noise is characterized by the appearance of individual pixels within the image with extreme values, where the values of some pixels are suddenly replaced by zero (representing pepper) or 255 (representing salt), while other values in the image remain unchanged. This random distribution of noise is a serious problem that clearly affects image quality [29, 30]. It can be represented by the following equation:

$$p(x) = \begin{cases} p & if\ x = 0 \\ q & if\ x = 225 \\ 1 - p - q & otherwise \end{cases} \quad (1)$$

Where p represents the probability of "pepper" appearing (value = 0), q represents the probability of "salt" appearing (value = 255), and 1-p-q represents the probability of the pixel remaining unchanged.

- **White Gaussian Noise**

White Gaussian noise is one of the most common types of noise in digital signal and image processing. It is used as an approximate model for many random physical phenomena affecting digital communication and image systems. This noise is defined as a random signal with a Gaussian probability distribution with a mean = 0 and a specific variance (σ^2), where noisy pixels follow this statistical distribution. It is also described as white because it has a flat frequency spectrum, meaning that all frequencies are represented by the same power, making them equally influential on all parts of the image or signal [31, 32]. It is represented by the following equation.

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}}.exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right) \qquad (2)$$

Where x represents the pixel intensity, θ represents the mean, and σ^2 represents the noise variance.

- **Horizontal misalignment noise**

Misalignment noise is a complex type of noise that appears particularly in applications that rely on analyzing time-sequences of images or video, such as motion tracking systems, video processing, and image fusion techniques. This noise arises from a failure to accurately align consecutive frames or images, leading to incorrect overlay of visual data and, consequently, to visible distortions in the final representation [33]. It is represented by the following equation.

$$\mathrm{P(I)}\frac{\sum_{\mathrm{q}\in\cap}\mathrm{I(q).w(p,q)}}{\sum_{\mathrm{q}\in\cap}\mathrm{w(p,q)}} \qquad (3)$$

$$\mathrm{w(p,q)} = \frac{1}{\|\mathrm{p}-\mathrm{q}\|} \cdot \frac{(\mathrm{p}-\mathrm{q}).\mathrm{N(p)}}{\|\mathrm{p}-\mathrm{q}\|} \qquad (4)$$

Where:

(p): The value we want to calculate for the corrupted pixel.

I(q): The value of a known neighboring pixel.

w(p, q): The weight that determines the contribution of pixel q to the calculation of p.

||p − q||: The Euclidean distance between the two pixels.

N(p): The normal vector (edge direction) at point p.

8. Results

Here, we present the results of applying the proposed technique and its resistance to noise using a DFD database containing 3,068 videos, divided into real and fake videos. The data was tested at a training rate of 80% and a testing rate of 20%. It demonstrated good noise resistance.

Table (1) shows the results of SWT in resisting salt-and-pepper noise, white Gaussian noise, and horizontal misalignment noise.

Table (1): The effect of noise on SWT

Noise Name		**Ratio**	**Accuracy**
salt and pepper noise	Noise	0.1	92.90
	With filter	0.1	94.40
	With SWT	0.1	96.00
	Noise	0.05	95.80
	With filter	0.05	96.30
	With SWT	0.05	98.00
White Gaussian noise	Noise	25	94.70
	With filter	25	97.10
	With SWT	25	98.22
	Noise	35	90.04
	With filter	35	95.25
	With SWT	35	96.55
Horizontal misalignment noise	Noise	10 pixel	94.64
	With filter	10 pixel	96.00
	With SWT	10 pixel	97.36
	Noise	5 pixel	95.23
	With filter	5 pixel	96.11
	With SWT	5 pixel	98.23

From the table above, we note that in the case of salt and pepper noise, at a noise level of (0.05), accuracy increased from 95.80% before processing to 96.30% using the Median filter, while it reached 98% using the proposed SWT technique, an increase of 1.77%. At a noise level of (0.1), results increased from 92.90% to 94.40% using the Median filter, reaching 96% using SWT, an increase of 1.70%.

For white Gaussian noise, experiments showed that a noise level of (25) increased accuracy from 94.70% to 97.10% using the Gaussian filter, while it reached 98.22% using SWT, an increase of 1.15%. At a higher noise level (35), the results increased from 90.04% to 95.25% using the Gaussian filter, and reached 96.55% using the SWT transform, a difference of 1.37%.At a noise level of 5 pixels, the accuracy increased from 95.23% before processing to 96.11% after using the filter, and then to 98.23% using the SWT transform, a difference of 1.42%. At a higher noise level, the results increased from 94.64% to 96% using the filter, and then to 97.36% using the SWT transform, a difference of 2.21%. These results clearly demonstrate that the use of the SWT technique in the proposed technique improves the system's resistance to many types of noise. Figure (3) Effect of the proposed technique on noise.

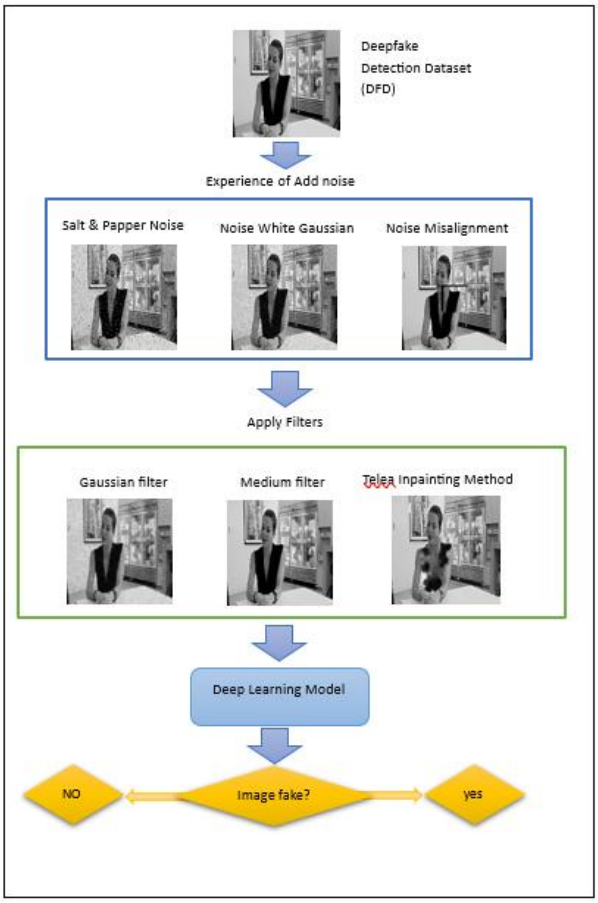

Fig 3. Effect of the proposed technique on noise.

The rotation resistance of the suggested method was also examined. Three rotational angles were used for testing: 60°, 90°, and 180°. Since these angles do not significantly alter the spatial distribution of face features, it was discovered that the recognition accuracy rises for angles equal to multiples of 90°. The application of the suggested SWT approach with rotation is demonstrated in Table (2).

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

Rotation degree	**accuracy**	
	Before the SWT	After the SWT
60°	95.00	96.84
90°	95.04	97.00
180°	97.20	98.69

9. Conclusion and Suggestions

The noise removal process in the preprocessing stage contributed to improving the accuracy of forgery detection, which had a positive impact on the overall performance of the proposed technique. The combination of noise removal techniques with image analysis methods demonstrated great effectiveness in preserving the quality of visual features even in the presence of complex noise such as horizontal misalignment. The proposed technique also demonstrated its high resistance to the effects of rotation on video images from different angles.

References

Agarwal, S., & Farid, H. (2021). Detecting deep-fake videos from aural and oral dynamics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 981–989). https://doi.org/10.1109/TCSVT.2023.3281474

Al-Tamimi, M. S. H. (2019). Combining convolutional neural networks and slantlet transform for an effective image retrieval scheme. *International Journal of Electrical and Computer Engineering*, *9*(5), 4382–4395. https://doi.org/10.11591/ijece.v9i5.pp4382-4395

Aneja, S., & Nießner, M. (2020). *Generalized zero and few-shot transfer for facial forgery detection*. arXiv preprint. https://arxiv.org/abs/2006.11863

Bar, L., Sochen, N., & Kiryati, N. (2005). Image deblurring in the presence of salt-and-pepper noise. In *Proceedings of the International Conference on Scale-Space Theories in Computer Vision* (pp. 107–118). Springer. https://doi.org/10.1007/11408031_10

Brodarič, M., Štruc, V., & Peer, P. (2024). Cross-dataset deepfake detection: Evaluating the generalization capabilities of modern deepfake detectors. In *Proceedings of the 27th Computer Vision Winter Workshop (CVWW)* (pp. 47–56).

Cao, J., Ma, C., Yao, T., Chen, S., Ding, S., & Yang, X. (2022). End-to-end reconstruction-classification learning for face forgery detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 4113–4122).

Chen, S., Yao, T., Chen, Y., Ding, J., Li, J., & Ji, R. (2021). Local relation learning for face forgery detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(2), 1081–1088.

Dey, S., Singh, P., & Saha, G. (2023). *Wavelet scattering transform for improving generalization in low-resourced spoken language identification*. arXiv preprint. https://arxiv.org/abs/2310.00602

Eickenberg, M., Exarchakis, G., Hirn, M., & Mallat, S. (n.d.). *Solid harmonic wavelet scattering for molecular energy regression*. Unpublished manuscript.

Gerstner, C. R., & Farid, H. (2022). Detecting real-time deep-fake videos using active illumination. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 53–60). https://doi.org/10.1007/978-3-031-73661-2_22

Ghosh, A., Sufian, A., Sultana, F., Chakrabarti, A., & De, D. (2019). Fundamental concepts of convolutional neural network. In *Intelligent Systems Reference Library* (Vol. 172, pp. 519–567). Springer. https://doi.org/10.1007/978-3-030-32644-9_36

Hadi, T. H. (2024). Deep learning-based DDoS detection in network traffic data. *International Journal of Electrical and Computer Engineering Systems*, *15*(5), 407–414. https://doi.org/10.32985/ijeces.15.5

Haliassos, A., Vougioukas, K., Petridis, S., & Pantic, M. (2021). Lips don't lie: A generalisable and robust approach to face forgery detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 5039–5049).

Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., Yang, Z., Zhang, Y., & Tao, D. (2023). A survey on vision transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(1), 87–110. https://doi.org/10.1109/TPAMI.2022.3152247

Hu, J., Liao, X., Gao, D., Tsutsui, S., Wang, Q., Qin, Z., & Shou, M. Z. (2023). *Mover: Mask and recovery based facial part consistency aware method for deepfake video detection*. arXiv preprint. https://arxiv.org/abs/2303.01740

Hwang, H., & Haddad, R. A. (1995). Adaptive median filters: New algorithms and results. *IEEE Transactions on Image Processing*, *4*(4), 499–502. https://doi.org/10.1109/83.370679

Ji, L., Wang, Y., Chen, K., Wu, Y., & Huang, D. (2024). *Distinguish any fake videos: Unleashing the power of large-scale data and motion features*. arXiv preprint. https://arxiv.org/abs/2405.15343

Kaur, A., Hoshyar, A. N., Saikrishna, V., Firmin, S., & Xia, F. (2024). Deepfake video detection: Challenges and opportunities. *Artificial Intelligence Review*, *57*(6), Article 159. https://doi.org/10.1007/s10462-024-10810-6

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

Qi, P., Cao, J., Li, Y., Liu, X., Meng, Y., & Shen, W. (2023). Fakesv: A multimodal benchmark with rich social context for fake news detection on short video platforms. In *Proceedings of the AAAI Conference on Artificial Intelligence*, *37*(12), 14444–14452. https://doi.org/10.1609/aaai.v37i12.26689

Rehman, M., Ahmed, F., Khan, M., Tariq, U., Alfouzan, F., Alzahrani, N. M., & Ahmad, J. (2022). Dynamic hand gesture recognition using 3D-CNN and LSTM networks. *Computers, Materials & Continua*, *70*(3), 4675–4690. https://doi.org/10.32604/cmc.2022.019586

Santhoshkumar, R., & Geetha, M. K. (2019). Deep learning approach for emotion recognition from human body movements with feedforward deep convolution neural networks. *Procedia Computer Science*, *152*, 158–165. https://doi.org/10.1016/j.procs.2019.05.038

Sarma, D., Kavyasree, V., & Bhuyan, M. K. (2022). Two-stream fusion model using 3D-CNN and 2D-CNN via video-frames and optical flow motion templates for hand gesture recognition. *Innovations in Systems and Software Engineering*. Advance online publication. https://doi.org/10.1007/s11334-022-00477-z

Sekar, V., & Jawaharlalnehru, A. (2022). Semantic-based visual emotion recognition in videos: A transfer learning approach. *International Journal of Electrical and Computer Engineering*, *12*(4), 3674–3683. https://doi.org/10.11591/ijece.v12i4.pp3674-3683

Song, H., Huang, S., Dong, Y., & Tu, W.-W. (2023). *Robustness and generalizability of deepfake detection: A study with diffusion models*. arXiv preprint. https://arxiv.org/abs/2309.02218

Telea, A. (2004). An image inpainting technique based on the fast marching method. *Journal of Graphics Tools*, *9*(1), 23–34.

Xu, G., & Aminu, M. J. (2022). An efficient procedure for removing salt and pepper noise in images. *Informatica*, *46*(2). https://doi.org/10.31449/inf.v46i2.3530

Zhang, D., Xiao, Z., Li, S., Lin, F., Li, J., & Ge, S. (2024). Learning natural consistency representation for face forgery video detection. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 407–424). Springer. https://doi.org/10.1109/TIFS.2025.3567110

Zhang, M., & Gunturk, B. K. (2008). Multiresolution bilateral filtering for image denoising. *IEEE Transactions on Image Processing*, *17*(12), 2324–2333. https://doi.org/10.1109/TIP.2008.2006658

Zhang, S., & Karim, M. A. (2002). A new impulse detector for switching median filters. *IEEE Signal Processing Letters*, *9*(11), 360–363. https://doi.org/10.1109/LSP.2002.805310