Proceedings of the 4th International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/2026.32907

Passive Detection Techniques for Artificial Intelligence Generated Images and Videos

Jingfei Lang

School of Science and Technology, Hong Kong Metropolitan University, Hong Kong, China
s1373373@live.hkmu.edu.hk

Abstract. Artificial Intelligence (AI) can now make very real images and videos. This helps create digital content, but it also brings big security risks. People can make fake news easily. The people need ways to find these fakes. Passive detection is a good method. It does not need watermarks added before making the image. Instead, it looks directly at the media files to find mistakes made by the computer. This paper reviews different passive detection methods. For images, this paper looks at pixel patterns and frequency data. For videos, this paper checks if frames connect smoothly over time and looks for body signals like heartbeats. Right now, detection programs work well on things they have seen before. However, they usually fail on new types of AI fakes. Future work must fix this problem so detectors can find any fake media, no matter how it was made. This paper aims to classify and review existing passive detection methods, reveal the common shortcomings of current algorithms in generalization ability, and point out the necessary path to build a general forgery detector in the future to address the security challenges brought about by the continuous evolution of deepfake technology.

Keywords: Artificial intelligence generated content, passive detection, deepfake, image recognition, information security

1. Introduction

Artificial intelligence has grown rapidly recently. Computer programs can now make pictures and videos that look exactly like real life [1]. These tools are useful for making movies, games, and art. But they also cause serious problems. Bad actors can use them to make fake news or fake videos of real people to spread lies [2]. This makes it hard for people to trust what they see online [3]. People must build tools to check if the media is real [4]. There are two main ways to do this: active defense and passive detection. Active methods put hidden marks, like watermarks, into the image when it is made. This only works if the software maker agrees to add them. Passive methods are different. They do not need any inside information from the software. They just look closely at the picture or video to find signs that a computer made it. Because it is easier to use in the real world, many researchers focus on passive detection to solve this problem.

DOI: 10.54254/2755-2721/2026.32907

2. Passive detection for fake images

To find fake images, people look for small differences between real camera photos and computer-made ones. There are two main ways to do this: looking at pixels in the spatial domain and looking at frequencies.

2.1. Checking image pixels (spatial features)

This method looks at how pixels are placed and how colors change. Deep learning models often leave small mistakes in the pixels when they make images larger (upsampling). People cannot see these mistakes, but computers can. Wang et al. [5] showed that basic computer vision models can learn to spot these mistakes. Figure 1 shows the flowchart of spatial domain detection for images using multi-scale feature fusion CNN.

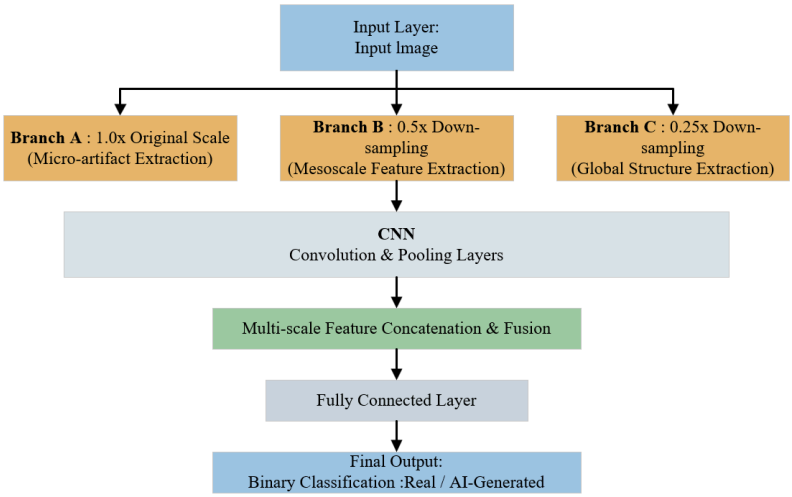

Figure 1. Flowchart of spatial domain detection for images using multi-scale feature fusion CNN [5]

Other researchers made better programs to find these clues. For example, Zhou et al. combined different networks to find changed faces [6]. However, new AI tools like diffusion models make much better images [7]. They make images step by step from noise, so they do not leave the same pixel mistakes. Corvi et al. found that old detection methods do not work well on these new diffusion models [8]. So, researchers need to look for new kinds of structural clues to catch them.

2.2. Checking image frequencies

Computer models often make images too smooth. They have trouble making high-frequency details, like sharp edges or random camera noise. Durall et al. showed that deep learning networks cannot copy the frequency patterns of real photos. People can change the image into a frequency map to see this better [9]. Frank et al. found that some AI models leave grid patterns in the high frequencies [10]. Real photos do not have these grids at all. Dzanic et al. also proved that AI networks make physical mistakes in these frequency areas [11]. Some researchers try to fix this by filtering the frequencies on purpose. Qian et al. used a method to make these hidden mistakes stand out more [12]. He et al. found that fake and real images act differently when people try to compress or change their frequencies [13]. A big benefit of checking frequencies is that it still works even if the picture

DOI: 10.54254/2755-2721/2026.32907

is resized or saved as a smaller JPEG. Figure 2 shows the flowchart of frequency domain detection for images based on the Fourier transform.

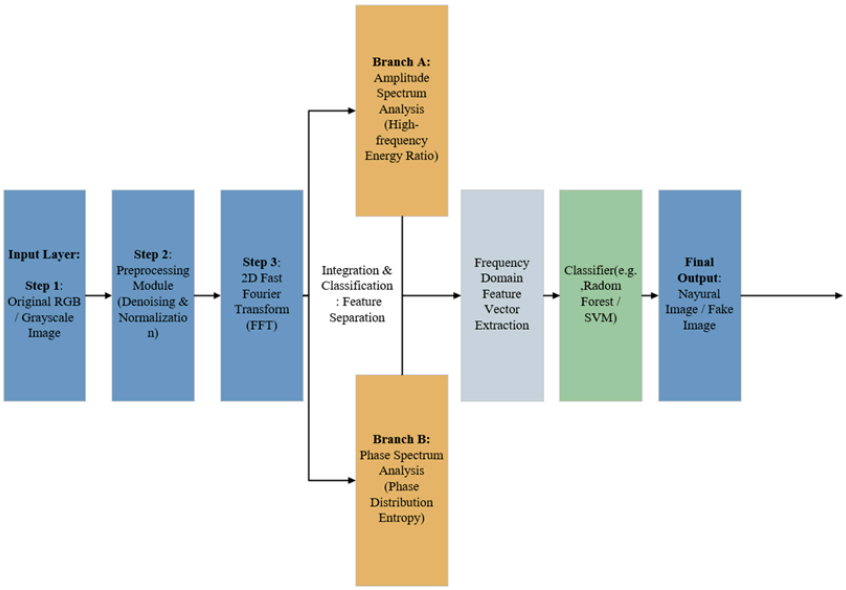

Figure 2. Flowchart of frequency domain detection for images based on the Fourier transform [12]

3. Passive detection for fake videos

It is harder to find fake videos than fake images. A video has many frames. Detectors must check each frame and also check if the frames fit together naturally over time.

3.1. Checking movement and time

Older AI models had trouble making smooth videos. The faces would shake, or the expressions looked stiff. Güera and Delp built a system to find these problems [14]. They used a network to check how frames relate to each other over time. Amerini et al. used a method called optical flow [15]. Figure 3 shows the flowchart of joint spatiotemporal (CNN+LSTM) detection for deepfake videos.

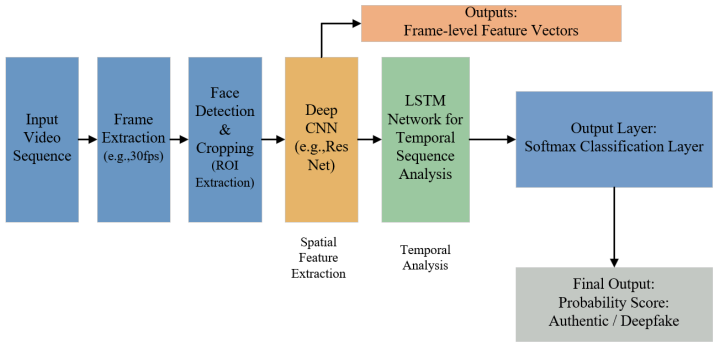

Figure 3. Flowchart of joint spatiotemporal (CNN+LSTM) detection for deepfake videos [14]

DOI: 10.54254/2755-2721/2026.32907

This checks how pixels move from one frame to the next to find unnatural face movements. Sometimes, a video is mostly real, and only a few frames are fake. Li et al. made a system to handle this [16]. It checks the video in parts. If even a small part is fake, it flags the whole video. Masi et al. and others added more features to make detection better [17]. They tell the computer to focus mostly on the face area instead of the background [18].

3.2. Checking body signals

As AI gets better, simple video mistakes disappear. So, researchers started looking at human biology. Computers have a hard time copying how real bodies work. Yang et al. looked at how people move their heads. Fake videos often get the 3D angles wrong [19]. Haliassos et al. looked at mouths [20]. They found that in fake videos, the lip movements do not match the sound perfectly. Another deep method looks at heartbeats. When a real heart beats, blood moves in the face, changing the skin color very slightly. Computers cannot see this, but special algorithms can. Figure 4 shows the flowchart of video verification based on rPPG physiological signals

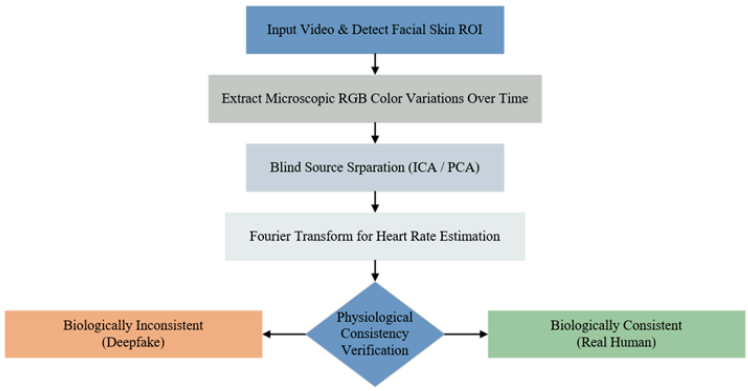

Figure 4. Flowchart of video verification based on rPPG physiological signals [21]

AI models do not know how to add real heartbeat signals to the pixels when they make a video. This makes heartbeat checking a very strong way to find fakes.

4. Discussion

Right now, detectors work very well in the lab. They can be over 90% accurate on the data they have studied. But they often fail in the real world. The biggest problem is that they cannot handle new AI models [22]. If a detector learns on one type of fake, it will probably miss a fake made by a brand-new program. Researchers are trying to fix this. Ojha et al [23]. and Shiohara et al [24]. found ways to train detectors on many different fake patterns. This helps them guess better on new fakes. Also, fake media now includes bad audio and video together. Future detectors must check both at the same time to see if they match.

5. Conclusion

This paper looked at how to find AI-made images and videos without using watermarks. For images, looking at pixels and frequencies works best. For videos, people have to look at how things move

DOI: 10.54254/2755-2721/2026.32907

over time and check biological signs like heartbeats. The technology to find fakes is getting better. However, the main problem is that detectors still fail on new types of AI fakes. Future research must find basic rules that apply to all fakes, no matter what program made them. As AI fakes look more like real life, people need strong tools to check them. This is very important to keep the internet safe and truthful.

References

[1] T. Brown, B. Mann, N. Ryder, et al., Language models are few-shot learners. Adv. Neural Inf. Process. Syst. 33, 1877–1901 (2020)
[2] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, et al., Deepfakes and beyond: A comprehensive survey of face manipulation and fake detection. Inf. Fusion 64, 131–148 (2020)
[3] A. Rössler, D. Cozzolino, L. Verdoliva, et al., Faceforensics++: Learning to detect manipulated facial images, in Proceedings of the IEEE/CVF International Conference on Computer Vision (2019), pp. 1–11
[4] H. Dang, F. Liu, J. Stehouwer, et al., On the detection of digital face manipulation, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2020), pp. 5781–5790
[5] S.Y. Wang, O. Wang, R. Zhang, et al., CNN-generated images are surprisingly easy to spot... for now, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2020), pp. 8695–8704
[6] P. Zhou, X. Han, V.I. Morariu and L.S. Davis, Two-stream neural networks for tampered face detection, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (2017), pp. 1831–1839
[7] R. Rombach, A. Blattmann, D. Lorenz, et al., High-resolution image synthesis with latent diffusion models, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2022), pp. 10684–10695
[8] R. Corvi, D. Cozzolino, G. Zingarini, et al., On the detection of synthetic images generated by diffusion models, in ICASSP 2023 – 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (IEEE, 2023), pp. 1–5
[9] R. Durall, M. Keuper and J. Keuper, 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 (2020), pp. 7890–7899
[10] J. Frank, T. Eisenhofer, L. Schönherr, et al., Leveraging frequency analysis for deep fake image recognition, in International Conference on Machine Learning (PMLR, 2020), pp. 3247–3258
[11] T. Dzanic, K. Shah and F. Witherden, Fourier spectrum discrepancies in deep network generated images. Adv. Neural Inf. Process. Syst. 33, 3022–3032 (2020)
[12] Y. Qian, G. Yin, L. Sheng, et al., Thinking in frequency: Face forgery detection by mining frequency-aware clues, in European Conference on Computer Vision (Springer, Cham, 2020), pp. 86–103
[13] P. He, H. Xin and J. Gao, Beyond the spectrum: Detecting deepfakes via re-synthesis, in IJCAI International Joint Conference on Artificial Intelligence (2021), pp. 3496–3502
[14] D. Güera and E.J. Delp, Deepfake video detection using recurrent neural networks, in 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (IEEE, 2018), pp. 1–6
[15] I. Amerini, L. Galteri, R. Caldelli and A. Del Bimbo, Deepfake video detection through optical flow based CNN, in Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (2019), pp. 0–0
[16] X. Li, Y. Lang, Y. Chen, et al., Sharp multiple instance learning for deepfake video detection, in Proceedings of the 29th ACM International Conference on Multimedia (2021), pp. 1864–1872
[17] I. Masi, Y. Wu, T. Hassner, et al., Two-branch recurrent network for isolating deepfakes in videos, in European Conference on Computer Vision (Springer, Cham, 2020), pp. 667–684
[18] H. Zhao, W. Zhou, D. Chen, et al., Multi-attentional deepfake detection, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2021), pp. 2185–2194
[19] X. Yang, Y. Li and S. Lyu, Exposing deep fakes using inconsistent head poses, in ICASSP 2019 – 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (IEEE, 2019), pp. 8261–8265
[20] A. Haliassos, K. Vougioukas, S. Petridis, et al., 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 (2021), pp. 5039–5049
[21] U.A. Ciftci, I. Demir and L. Yin, FakeCatcher: Detection of synthetic portrait videos using biological signals. IEEE Trans. Pattern Anal. Mach. Intell. 43(12), 4359–4370 (2020)

DOI: 10.54254/2755-2721/2026.32907

[22] S. Tariq, S. Lee, H. Kim, et al., Detecting both machine and human created fake face images in the wild, in Proceedings of the 2nd International Workshop on Multimedia Privacy and Security (2018), pp. 81–87
[23] U. Ojha, Y. Li and Y.J. Lee, Towards universal fake image detectors that generalize across generative models, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2023), pp. 24443–24452
[24] K. Shiohara and T. Yamasaki, Detecting deepfakes with self-blended images, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2022), pp. 13220–13229.