# **CHAPTER** **TWO**

# **LITERATURE REVIEW**

# **INTRODUCTION**

Fake image detection using machine learning is a challenging and emerging research area due to image editing tools' increasing availability and ease of use. Detecting and effectively mitigating the spread of fake images is crucial for various applications, including image forensics, social media content moderation, and maintaining the integrity of visual information. Machine learning and deep learning techniques have emerged as effective tools for addressing this challenge.

In recent years, numerous studies have been conducted to develop automated systems that can detect fake images using machine learning techniques. This literature review, which employs the chronological method, summarizes the existing research in this field and highlights the challenges and future directions.

# **OVERVIEW OF EXISTING SYSTEM**

The existing systems in use for detecting fake images have evolved from using simple machine learning classifiers and image metadata, to the use of the convolutional neural network deep learning algorithm and the involvement of generative adversarial networks in discerning between real and falsified images. Various combinations of the above methods have been employed in different studies with some unique characteristics being noticeable in some.

# **REVIEW OF RELATED WORK**

According to Lyu et al. (2004), who conducted one of the earliest studies on fake image detection, who proposed a statistical method for detecting digital image forgeries by analyzing the inconsistencies in image properties. Since then, many machine learning-based techniques have been developed for detecting fake images, including deep learning models such as convolutional neural networks (CNNs), generative adversarial networks (GANs), and recurrent neural networks (RNNs).

In a study by Strigl et al. (2010), the implementation of a framework for accelerating training and classification of arbitrary Convolutional Neural Networks (CNNs) on the Graphic Processing Unit (GPU) was presented and the performance and scalability improvement that can be achieved by shifting the computation-intensive tasks of a Convolutional Neural Network to the GPU was displayed. Depending on the network topology training and classification on the GPU, it performs two to twenty-four times faster than on the CPU. Furthermore, the GPU version scales much better than the CPU implementation with respect to the network size.

In a study of image similarity measure by Roy (2013), images were compared using a content-based image retrieval system. This system required feature extraction which involved mapping the image pixels onto the feature space. Image feature extraction techniques were said to be classified into two groups, the pixel domain and compress domain feature extraction technique. Their study employed the pixel domain extraction as they were more concerned with visual features such as color, texture, and shape. Their system proposed a multi-feature extraction method (color histogram, color coherence vector, and canny edge detection technique) and considered 8-bit color images and this combination of techniques enabled high accuracy in detection of similar images.

In a study by Bunk et al. (2017), two systems were proposed to detect and localize fake images using a mix of resampling properties and deep learning. In the initial system, the Radon conversion of resampling properties is determined on overlapping pictures corrections. Deep learning classifiers and a Gaussian conditional domain pattern are then used to construct a heat map. A Random Walker segmentation method uses total areas. In the next system, for identification and localization, software resampling properties are passed on overlapping object patches over a long-term memory (LSTM)- based network. In addition, the detection/ localization performance of both systems was compared. The results confirmed that both systems are active in detecting and settling digital image fraud.

According to Villan et al. (2017), he proposed a machine learning-based approach for detecting fake images. The authors utilized neural networks to identify manipulated regions within images, making it possible to distinguish between original and tampered content. Additionally, the study introduced the use of Error Level Analysis (ELA) and image metadata as supplementary features for improving detection accuracy. The combination of these techniques demonstrated promise in identifying fake images, particularly in the context of social media sharing.

According to Kim (2017), digital forensics techniques are needed to detect manipulation and fake images used for illegal purposes. Thus, the researchers in this study have been working on an algorithm to detect fake images through deep learning technology, which has achieved remarkable results in modern research. First, a converted neural network is applied to image processing. In addition, a high pass filter is used to get at hidden features in the image instead of semantic information in the image. For experiments, modified images are created using intermediate filter, Gaussian blurring, and added white Gaussian noise.

According to a study conducted by Yang et al. (2018), the identification of fake news and images is very difficult, as fact-finding of news on a pure basis remains an open problem and few existing models Can be used to resolve the problem. It has been proposed to study the problem of "detecting false news." Through a thorough investigation of counterfeit news many useful properties are determined from text words and pictures used in counterfeit news. There are some hidden characteristics in words and images used in fake news, which can be identified through a collection of hidden properties derived from this model through various layers. A pattern called TI-CNN (Text and Image information based Convolutional Neural Network) has been proposed. By displaying clear and embedded features in a unified space, TI-CNN is trained with both text and image information at the same time.

In a study by Do et al. (2018), the focus was the development of anti-forensic techniques with the main purpose being to propose a deep convolutional neural network for detecting real image or fake image from GANs. The results were based on evaluation from an AI challenge contest. A training dataset that could adapt to the test data was first built, then a deep learning network was built based on face recognition networks to extract face features, these features were then fine-tuned to make them suitable for classification, and finally, results were obtained from the validation data. Deep convolutional generative adversarial networks (DC-GANs) were used to generate images of size 64x64, while Progressive Generative Adversarial Networks (PG-GANs) were used for images of size 256x256, and 1024x1024. The VGG16 architecture was used and an accuracy of 80% was obtained.

Recent advancements in generative adversarial networks (GANs) have led to a cat and mouse game in fake image creation and detection, making this an ever-evolving research area. A study carried out by Marra et al. (2018) focused on detecting fake images generated using Generative Adversarial Networks (GANs) on social networks. Their research highlighted the importance of distinguishing between GAN-generated images and authentic ones. The authors introduced novel methods for identifying these synthetic images, which are often used for disinformation campaigns on social media.

In a study by AlShariah et al. (2019), the aim was to propose an approach to extract image content, classify it, and verify the authenticity of digital images on social media platforms. Specifically, the research focuses on detecting threats and forged images on Instagram, as these pose problems to society and national security. They stated the importance of social media in people's daily lives and the prevalence of image sharing on platforms like Instagram and WhatsApp and the ease with which individuals and groups can fabricate and disseminate fake images, posing a threat to the credibility of news and public trust in social communication. The model used in this study is based on deep algorithms learning, specifically Convolutional Neural Network (CNN) and Alexnet network. Transfer learning using Alexnet is also utilized. The results show that the proposed Alexnet network offers more accurate detection of fake images compared to other techniques, with a 97% accuracy rate.

According to a study by Hsu et al. (2020), Traditional image forgery detection methods are ineffective in recognizing fake images generated by GANs. As a result, their study proposes a new method that utilizes pairwise learning to detect fake images. The approach involves generating a fake-real image pair using different state-of-the-art GANs and training a reduced DenseNet to learn common fake features from these pairs. The network is then equipped with a classification layer to distinguish between fake and real images. The study highlights the challenges associated with detecting fake face images and addresses them by employing a two-step learning method. The first step involves pairwise learning to extract common fake features, and the second step involves training a classifier to differentiate between fake and real images generated by various GANs. Experimental results show that the proposed method outperforms other state-of-the-art fake image detectors. The authors collected a dataset from the CelebA dataset and used five different GANs, including DCGAN, WGAP, WGAN-GP, LSGAN, and PGGAN, to generate fake images for training the model.

In a study of deep fakes by Negi et al. (2021), the use of neurons in neural networks to represent the values of pixels was explored. The common methods of creating and detecting deep fakes were stated with the creation of deep fakes using the encoder decoder network as well as Convolutional Neural Networks (CNN), (RNN) or a combination of both as the creation method and the separation of image parameters and the detection of gaussian noise as key detection methods. This study also outlined the metamorphosis of neural networks from the use of GANs which could generate faces but were not as efficient and led to poor image quality, to Complex Convolutional Networks or Deep Convolutional GANs (DCGAN) which showed more processing results as they gave more importance to learnable weights and biases and compared them, and finally Coupled GAN or Co-GAN which used two generators and two detectors.

In a study into Deep learning model for deep fake face detection and recognition by Suganthi et al. (2022). A deep learning model specifically designed for deep fake face recognition and detection was proposed. Previous research in the realm of fake images detection has predominantly relied on traditional computer vision methods, but the adoption of deep learning models has shown remarkable improvements in accuracy and robustness. The study addressed the growing concern of deep fake videos and images that manipulate facial features. Their deep learning approach exhibited high accuracy in detecting manipulated facial content, contributing to the ongoing efforts to combat deep fake technology.

According to Sabah (2022), the study discussed the widespread use of fake images in social media, especially in politics and the entertainment industry, and the harmful effects they can have. It mentions the emergence of powerful mobile applications and the Generative Adversarial Network (GAN) as contributing factors to the ease of generating indistinguishable fake images. To address this problem, the study proposes the use of CNN as the most popular algorithm in deep learning for detecting fake images. The process involves preprocessing the images by converting them to the YCbCr color space, applying gamma correction, and using the Canny filter for edge detection. The study compares two different methods of detection: one with Principal Component Analysis (PCA) and one without. The results of the study show that using CNN with PCA resulted in acceptable accuracy, while using CNN only gave the highest level of accuracy in detecting manipulated images.

According to a study by khudeyer (2023), deep learning has emerged as a potent tool to counter the proliferation of fake images in today's digital landscape. Their study focused on the application of deep learning techniques for fake image detection. Their work highlighted the effectiveness of deep neural networks in identifying various types of manipulated images. By leveraging the power of deep learning, their approach showed promising results in distinguishing fake images from genuine ones.

# **STRENGTH AND WEAKNESS OF THE EXISTING SYSTEMS**

The existing systems have been able to make groundbreaking improvements in the field of fake image detection by making use of deep learning algorithms such as neural networks and other advanced techniques to combat the quickly developing adoption of generative adversarial networks and the widespread use of image manipulation software, however, these systems have not been integrated into easily accessible platforms for use by the massive number of users who are subject to fake images and their effects on social media and similar platforms where the verification of content authenticity is crucial.

# **CONCLUSION**

In summation, this literature review has offered a comprehensive overview of the field of machine learning as it pertains to the detection of counterfeit images, based on a careful analysis of six pivotal articles. Throughout this examination, we have observed the remarkable advancements in machine learning, particularly within the realm of deep learning algorithms, which have played an instrumental role in addressing the issue of fake image detection. Furthermore, a recurring theme across these articles is the pivotal role of high-quality and diverse datasets in training accurate models. This emphasizes the significance of robust data collection and curation in the pursuit of improving detection accuracy. The collaborative efforts of multidisciplinary teams, uniting computer scientists, mathematicians, and experts from various domains, underscore the intricate nature of this challenge and the necessity for diverse perspectives in solving it.

This project differs from the works above in its integration of the machine learning model with a web application which serves to facilitate ease of use and access for users to verify the authenticity of images.

# **REFERENCES**

AlShariah, N. M., Khader, A., & Saudagar, J. (2019). Detecting fake images on social media using machine learning. *International Journal of Advanced Computer Science and Applications*, *10*(12), 170-176.

Bunk, J., Bappy, J. H., Mohammed, T. M., Nataraj, L., Flenner, A., Manjunath, B. S., ... & Peterson, L. (2017, July). Detection and localization of image forgeries using resampling features and deep learning. In *2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW)* (pp. 1881-1889). IEEE.

Do, N. T., Na, I. S., & Kim, S. H. (2018). Forensics face detection from GANs using convolutional neural network. *ISITC*, *2018*, 376-379.

Hsu, C. C., Zhuang, Y. X., & Lee, C. Y. (2020). Deep fake image detection based on pairwise learning. *Applied Sciences*, *10*(1), 370.

Khudeyer, R. S., & Almoosawi, N. M. (2023). Fake Image Detection Using Deep Learning. *Informatica*, *47*(7).

Kim, D. H., & Lee, H. Y. (2017). Image manipulation detection using convolutional neural network. *International Journal of Applied Engineering Research*, *12*(21), 11640-11646.

Lyu S., Rockmore D., and Farid H. 2004. A digital technique for art authentication. *Proceedings of the National Academy of Sciences* 101, 49 (2004), 17006-17010. DOI:<https://doi.org/10.1073/pnas.0406398101>

Marra, F., Gragnaniello, D., Cozzolino, D., & Verdoliva, L. (2018, April). Detection of gan-generated fake images over social networks. In *2018 IEEE conference on multimedia information processing and retrieval (MIPR)* (pp. 384-389). IEEE.

Negi, S., Jayachandran, M., & Upadhyay, S. (2021). Deep fake: an understanding of fake images and videos. *International Journal of Scientific Research in Computer Science Engineering and Information Technology*, *7*(3), 183-189.

Roy, K., & Mukherjee, J. (2013). Image similarity measure using color histogram, color coherence vector, and sobel method. *International Journal of Science and Research (IJSR)*, *2*(1), 538-543.

Sabah, H. (2022). A Detection of Deep Fake in Face Images Using Deep Learning. *Wasit Journal of Computer and Mathematics Science*, *1*(4), 94-111.

Strigl, D., Kofler, K., & Podlipnig, S. (2010, February). Performance and scalability of GPU-based convolutional neural networks. In *2010 18th Euromicro conference on parallel, distributed and network-based processing* (pp. 317-324). IEEE.

Suganthi, S. T., Ayoobkhan, M. U. A., Bacanin, N., Venkatachalam, K., Štěpán, H., & Pavel, T. (2022). Deep learning model for deep fake face recognition and detection. *PeerJ Computer Science*, *8*, e881.

Villan, M. A., Kuruvilla, A., Paul, J., & Elias, E. P. (2017). Fake image detection using machine learning. *IRACST-International Journal of Computer Science and Information Technology & Security (IJCSITS)*.

Yang, Y., Zheng, L., Zhang, J., Cui, Q., Li, Z., & TI-CNN, P. Y. (2018). Convolutional neural networks for fake news detection. *arXiv preprint arXiv:1806.00749*, *2*(6).