IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

A Review on Fake Image Detection Using Machine Learning

Vaishali Gedam[1], Sahil Meshram[2], Shrinivas Chinchanikar[3], Ganesh Pandey[4], Yash Lad[5], Punam Bhandarkar [6]

[1]Associate Professor, Department of Computer Science and Engineering, NIT, Nagpur
[2]B.E Graduate (iv year), Department of Computer Science and Engineering, NIT, Nagpur
[3]B.E Graduate (iv year), Department of Computer Science and Engineering, NIT, Nagpur
[4]B.E Graduate (iv year), Department of Computer Science and Engineering, NIT, Nagpur
[5]B.E Graduate (iv year), Department of Computer Science and Engineering, NIT, Nagpur
[6]B.E Graduate (iv year), Department of Computer Science and Engineering, NIT, Nagpur

Abstract : Nowadays biometric systems are useful in recognizing a person's identity, but criminals change their appearance in behaviour and psychological to deceive recognition system. To overcome this problem we are using a new technique called Deep Texture Features extraction from images and then building train machine learning model using CNN (Convolution Neural Networks) algorithm. This technique refers as LBPNet or NLBPNet as this technique is heavily dependent on features extraction using LBP (Local Binary Pattern) algorithm. In this project, we are designing LBP Based machine learning Convolution Neural Network called LBPNET to detect fake face images. Here first we will extract LBP from images and then train LBP descriptor images with Convolution Neural Network to generate a training model. Whenever we upload a new test image then that test image will be applied to the training model to detect whether the test image contains a fake image or a non-fake image. Below we can see some details on LBP.

Index Terms **-** Biometry, Identity, Recognition, Detection, Fake face.

INTRODUCTION

In this technological era a huge number of people have become victims of image forgery. A lot of people use technology to manipulate images and use it as evidences to mislead the court. So to put an end to this, all the images that are shared through social media should be categorized as real or fake accurately. Social media is a great platform to socialize, share and spread knowledge but if caution is not exercised, it can mislead people and even cause havoc due to unintentional false propaganda. While manipulation of most of the photos hoped images is clearly evident due to pixelization & shoddy jobs by novices, some of them indeed appear genuine. Especially in the political arena, manipulated images can make or break a politician's credibility.

Current forensic techniques require an expert to analyze the credibility of an image. We implemented a system that can determine whether an image is fake or not with the help of machine learning and thereby making it available for the common public. This project will unfold into three sections whereby first will focus on the second will focus on the Implementation details while the last part showcase the experimental result.

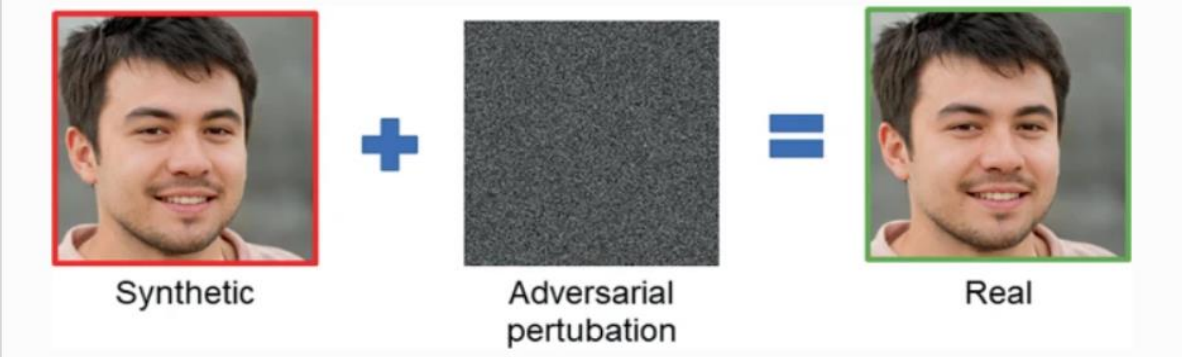
Objective and Scope of the Project

The advent of deep learning has given extraordinary impulse to both face manipulation methods and forensic detection tools. We have seen that successful detectors rely on inconsistencies at different levels, looking for both hidden and visible artifacts. One first important observation is that visual imperfections on faces will likely disappear soon. Newer GAN architectures [28] already improved upon this aspect by producing faces with even more details and highly realistic. Thus, relying exclusively on these traces could be a losing strategy in the long term. Turning to generic deep learning based-solutions, the main technical issue is probably the inability to adapt to situations not seen in the training phase. Misalignment between training and test, compression, and resizing are all sources of serious impairments and, at the same time, highly realistic scenarios for real-world applications. Also, to deal with the rapid advances in manipulation technology, deep networks should be able to adapt readily to new manipulations, without a full re-training, which may be simply impossible for lack of training data or entail catastrophic forgetting phenomena.

A more fundamental problem is the two-player nature of this research which is common to many security-related fields. In fact, detection algorithms must confront with the capacity of an adversary to fool them. This means that new solutions are needed in order to cope with unforeseen attacks. This applies to any type of classifier and is also very well known in forensics, where many counter-forensics methods have been proposed in the literature in order to better understand weaknesses of current approaches and help to improve them over time.
In the following, we analyze some works that have shown the vulnerabilities of GAN detectors to different types of threats.

It is well known, from the object recognition field, that suitable slight perturbations can induce misclassification . Following this path, it has been investigated the robustness of GAN detectors to imperceptible noise both in a white-box and in a black-box scenario. The authors show that it is possible to generate appropriate adversarial perturbations so as to misclassify fake images as real (see Fig.1), but also the opposite. In addition, they show that the attack can survive JPEG compression. Interestingly, it is also possible to design an effective strategy in a black-box threat model when the adversary does not have perfect knowledge of the classifier but is aware about the type of classifier. A similar analysis is conducted , where adversarial attacks are designed to fool co-occurrence-based GAN detectors.
Removing GAN fingerprints. Instead of adding noise, one can take a different perspective and remove the specific fingerprints that are used to discriminate GAN images from real ones. This approach is pursued , where an autoencoder-based strategy is proposed, that is trained using only real faces and is able to remove the high-frequency components that correspond to the fingerprints of the models used to generate synthetic images. At test time the autoencoder takes as input synthetic face images and modifies them so as to spoof GAN detection systems.

Inserting camera fingerprints. Another possible direction to attack GAN detectors is to insert the specific camera traces that characterize real images. In fact, real images are characterized by their own device and model fingerprints, as explained before. Such differences are important to carry out camera model identification from image content but can also be used to better highlight anomalies caused by image manipulations . It is proposed a targeted black-box attack that is based on a GAN architecture, able to insert specific real camera traces in a synthetic images. In this way it is possible not only to fool a GAN detector without any prior information on its architecture, but also to fool a camera model identification algorithm, that will attribute the GAN image to the targeted camera under attack.And automatic tools capable of distinguishing synthetic faces from real ones. The scientific community is making a huge research effort in this field, proposing several interesting approaches. However, a universal detector is yet to come. Fundamentally, the research in this field is like a cat and mouse game, with new detectors that are designed to deal with powerful synthetic face generators, while the latter keep improving to produce more and more realistic images. In this chapter we will present the most effective techniques proposed in the literature for the detection of synthetic faces. We will analyze their rationale, present real-world application scenarios , and compare different approaches in terms of accuracy and generalization ability.



A small and imperceptible adversarial perturbation can be added to the synthetic face image in order to fool the detector

What contribution would the project make

Educational deepfake examples

Deepfake technology holds positive potential for education. It could revolutionise our history lessons with interactivity. It could preserve stories and help capture attention. How? With deepfake examples of historical figures.
For instance, in 2018 the Illinois Holocaust Museum and Education Centre created hologrammatic interviews. So, visitors could talk to and interact with Holocaust survivors. They could ask questions and hear their stories. As deepfake technology advances, this kind of virtual history could boco

me achievable on a much wider scale.

Students interacting with artificially rendered Holocaust survivors. Source: ***YouTube***

Another example comes from **CereProc**, a company that ‘resurrected’ JFK in voice. This deepfake made it possible to hear the late president deliver the speech he would have delivered, if not for his assassination.

In this way, deepfake technology could help us preserve not just the facts in history books, but the impact **historical events** had on real people.

LITERATURE REVIEW

Very little work has been finalized around detecting forge audio, images, and videos. Yet, several studies and tasks are underway to identify what can be done around the incredible proliferation of counterfeit pictures online. Adobe recognizes the way in which Photoshop is misused and has tried to offer a sort of antidote [8]. The following provides a summary of a few of these studies: According to a study [9] conducted by Zheng 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 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. Ratri’s 2018 architecture [10] was proposed to identify counterfeit accounts in social networks, especially on Facebook. In this research, a machine learning feature was used to better predict fake accounts, based on their posts and the placement on their social networking walls. Support Vector Machine (SVM) and Complement Naïve Bayes (CNB) were used in this process, to validate content based on text classification and data analysis. The analysis of the data focused on the collection of offensive words, and the number of times they were repeated. For Facebook, SVM shows a 97% resolution where CNB shows 95% accuracy in recognizing Bag of Words (BOW) -based counterfeit accounts. The results of the study confirmed that the main problem related to the safety of social networks is that data is not properly validated before publishing. In a 2017 study by Bunk et al [11], 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. Aphiwongsophon and Chongstitvatana [12], aimed to use automated learning techniques to detect counterfeit news. Three common techniques were used in the experiments: Naïve Bayes, Neural Network, and Support Vector Machine (SVM). The normalization method is a major step to disinfect data before using the automatic learning method to sort information. The results show Naïve Bayes to have a 96.08% accuracy in detecting counterfeit news. There are two other advanced methods, the Neural Network Machine and the Support Network (SVM), which achieve 99.90% accuracy. In [13] by Kuruvilla et al., a neural network was successfully trained by analyzing the 4000 fake and 4000 real images error levels. The trained neural network has succeeded in identifying the image as fake or real, with a high success rate of 83%. The results showed that using this application on mobile platforms significantly reduces the spread of fake images across social networks. In addition, this can be used as a false image verification method in digital authentication, court evidence assessment, etc. This research develops an approach that takes an image as input and classifies it, using the CNN model. For a completely new task/problem, CNNs are very good feature extractors. It extracts useful attributes from an already trained CNN with its trained weights by feeding your data at each level and tuning the CNN a bit for the specific task. This means that a CNN can be retrained for new recognition tasks, enabling it to build on pre-existing networks. This is called pre-training, where one can avoid training a CNN from the beginning and save time. CNN can carry out automatic feature extraction for the given task. It eliminates the need for manual feature extraction since the features are learned directly by the CNN. In terms of performance, CNNs outperform many methods for image recognition tasks and many other tasks where it gives high accuracy and accurate result. Another key feature of CNNs is weight sharing, which basically means that the same weight is used for two layers in the model. Due to the above features and advantages, CNN is used in this research in comparison to other deep learning algorithms.

[1] G.Mohamed Sikandar, "100 Social Media Statistics You must know," [online] Available at: https://blog.statusbrew.com/social- mediastatistics-2018-for-business/ [Accessed 02 Mar 2019]

In this paper we have to learn that The results showed that the proposed net network offers more accurate detection of fake images compared to the other techniques with 97%. The results of this research will be helpful in monitoring and tracking in the shared images in social media for unusual content and forged images detection and to protect social media from electronic attacks and threats. Keywords—Convolution Neural A huge number of people have become victims of photo forgery in this technological age. Some criminals use software to exploit and use pictures as evidence to confuse the courts of justice [17]. To put an end to this, all photographs exchanged via social media should be labeled as true or fake.

[2] L. Zheng, Y. Yang, J. Zhang, Q. Cui, X. Zhang, Z. Li, et al. (2018). TICNN: Convolutional Neural Networks for Fake News Detection. United States

In this paper we have to learn that n this technological era, social media has a major role in people's daily life. Most people share text, images, and videos on social media frequently (e.g. Twitter, Snapchat, Facebook, and Instagram). Images are one of the most common types of media share among users on social media. So, there is a need for monitoring of images contained in social media. It has become easy for individuals and small groups to fabricate these images and disseminate them widely in a very short time, which threatens the credibility of the news and public confidence in the means of social communication.

[3].GMI_BLOGGER,"Saudi Arabia Social Media Statistics," GMI_ blogger. [online] Available at :https: //www.globalmediainsight.com/ blog/saudi-arabia-social-media-statistics/ [Accessed 04 May 2019].

In this paper we have to learn that In this project, we are designing LBP Based machine learning Convolution Neural Network called LBPNET to detect fake face images. Here first we will extract LBP from images and then train LBP descriptor images with Convolution Neural Network to generate a training model.
Whenever we upload a new test image then that test image will be applied to the training model to detect whether the test image contains a fake image or a non-fake image. Below we can see some details on LBP

[4]. A. Krizhevsky, I. Sutskever, & G. E. Hinton, (2012). Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems, 1097–1105.

This research develops an approach that takes an image as input and classifies it, using the CNN model. For a completely new task/problem, CNNs are very good feature extractors. It extracts useful attributes from an already trained CNN with its trained weights by feeding your data at each level and tuning the CNN a bit for the specific task. This means that a CNN can be retrained for new recognition tasks, enabling it to build on pre-existing networks. This is called pre-training, where one can avoid training a CNN from the beginning and save time. CNN can carry out automatic feature extraction for the given task. It eliminates the need for manual feature extraction since the features are learned directly by the CNN. In terms of performance, CNNs outperform many methods for image recognition tasks and many other tasks where it gives high accuracy and accurate result. Another key feature of CNNs is weight sharing, which basically means that the same weight is used for two layers in the model. Due to the above features and advantages, CNN is used in this research in comparison to other deep learning algorithms.

[5]. G.Mohamed Sikandar, "100 Social Media Statistics You must know," [online] Available at:https://blog.statusbrew.com/social-mediastatistics-2018-for-business/ [Accessed 02 Mar 2019].

In this paper we have to learn Very little work has been finalized around detecting forge audio, images, and videos. Yet, several studies and tasks are underway to identify what can be done around the incredible proliferation of counterfeit pictures online. Adobe recognizes the way in which Photoshop is misused and has tried to offer a sort of antidote [8]. The following provides a summary of a few of these studies: According to a study [9] conducted by Zheng 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

METHODOLOGY

The proposed method to detect the fake image can achieve by processing image in many steps as follow:

1.First the original image is transformed using SVD Where and are the orthogonal matrices, denotes the transpose of, and is a diagonal matrix whose diagonal elements can form a column vector

2.Two secret column vectors, are constructed, which satisfyide and Where denotes the inner product.

3. The main goal is to protect the image before publishing it to the public; this will be achieved by changing the diagonal of □□ matrix result from relation (2) with new elements counted by the following equation:□□□(□□ □ □□□□) □□ □□□□□□□□ □□□□□□□□□□□□□□□□□□□□□ □ □ □ □ Where □ is a scalar factor, which is set to 0.0001 for the purpose of this research. The vector □□ from relation (3) is restored as new diagonal elements into zero matrix □□correspondingly, for that new image will be constructed (A') as a protected image from the following relation:

□□□ □ □□□□□ …………….. 4 □' is the preprocessed image and publish to public's is robust to slight alteration of images, i.e., the vector v is stable under slight alteration of the image. In proposed image preprocessing procedure, the alteration of the vector □ in relation (2) is very small, keep □□ □ □ by Proposition 2.2 ,so the image preprocessing does not change the quality of origin image significantly (which will be demonstrated inthe later examples).

4. Another suggestion in current research is using auto threshold by Theorem 2.3. instead of constant threshold (0.01) for all images as in previous researches, which mean for each imagee .

Hardware and software to be use

Figure 1: Steps in Creating a Deep fake

- **Gathering of source and destination video (CPU)**—A minimum of several minutes of 4K source and destination footage are required. The videos should demonstrate similar ranges of facial expressions, eye movements, and head turns. One final important point is that the identities of source and destination should already look similar. They should have similar head and face shape and size, similar head and facial hair, skin tone, and the same gender. If not, the swapping process will show these differences as visual artifacts, and even significant post-processing may not be able to remove these artifacts.
- **Extraction (CPU/GPU)**—In this step, each video is broken down into frames. Within each frame, the face is identified (usually using a DNN model), and approximately 30 facial landmarks are identified to serve as anchor points for the model to learn the location of facial features. An example image from the Face Swap framework is shown in Figure 2 below.

CONCLUSION

In this study, we have proposed a novel common fake feature network-based pairwise learning, to detect the fake face/general images generated by state-of-the-art GANs successfully. The proposed CFFN can be used to learn the middle- and high-level and discriminative fake features by aggregating the cross-layer feature representations into the last fully connected layers. The proposed pairwise learning can be used to improve the performance of fake image detection further. With the proposed pairwise learning, the proposed fake image detector should be able to have the ability to identify the fake image generated by a new GAN. Our experimental results demonstrated that the proposed method outperforms other state-of-the-art schemes in terms of precision and recall rate. For future, work square measure for instance employing an additional complicated and deeper model for unpredictable issues. Integration of deep neural networks with the idea of increased learning, wherever the model is simpler. Neural network solutions seldom take under consideration non-linear feature interactions and non-monotonous short-run serial patterns, that square measure necessary to model user behavior in thin sequence information. A model is also integrated with neural networks to unravel this downside. The dataset can be inflated and another variety of images can be used for coaching, for instance, gray-scale pictures.

REFERENCES

[1]. G.Mohamed Sikandar, "100 Social Media Statistics You must know," [online] Available at:https://blog.statusbrew.com/social-mediastatistics-2018-for-business/ [Accessed 02 Mar 2019].

[2]. Damian Radcliffe, Amanda Lam, "Social Media in the Middle East,"[online]Available:https://www.researchgate.net/publication/32318 5146_Social_Media_in_the_Middle_East_The_Story_of_2017 [Accessed 06 Feb 2019].

[3]. GMI_BLOGGER,"Saudi Arabia Social Media Statistics," GMI_ blogger. [online] Available at:https://www.globalmediainsight.com/ blog/saudi-arabia-social-media-statistics/ [Accessed 04 May 2019].

[4]. Kit Smith,"49 Incredible Instagram Statistics,". Brandwatch. [online] Available at: https://www.brandwatch.com/blog/instagram-stats/ [Accessed 10 May 2019].

[5]. Selling Stock. (2014). Selling Stock. [online] Available at: https://www. selling-stock.com/Article/18-billionimages-uploaded-to-the-web-everyd [Accessed 12 Feb 2019].

[6]. Li, W., Prasad, S., Fowler, J. E., & Bruce, L. M. (2012). Localitypreserving dimensionality reduction and classification for hyperspectral image analysis. IEEE Transactions on Geoscience and Remote Sensing, 50(4), 1185–1198.

[7]. A. Krizhevsky, I. Sutskever, & G. E. Hinton, (2012). Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems, 1097–1105.

[8]. K. Ravi, (2018). Detecting fake images with Machine Learning. Harkuch Journal

[9]. L. Zheng, Y. Yang, J. Zhang, Q. Cui, X. Zhang, Z. Li, et al. (2018). TICNN: Convolutional Neural Networks for Fake News Detection. United States

[10]. R. Raturi, (2018). Machine Learning Implementation for Identifying Fake Accounts in Social Network. International Journal of Pure and Applied Mathematics, 118(20), 4785-4797.

[11] .J. Bunk, J. Bappy, H. Mohammed, T. M. Nataraj, L., Flenner, A., Manjunath, B., et al. (2017). Detection and Localization of Image Forgeries using Resampling Features and Deep Learning. The University of California, Department of Electrical and Computer Engineering, USA.

[12]. S. Aphiwongsophon, & P. Chongstitvatana, (2017). Detecting Fake News with Machine Learning Method. Chulalongkorn University, Department of Computer Engineering, Bangkok, Thailand.

[13]. M. Villan, A. Kuruvilla, K. J. Paul, & E. P. Elias, (2017). Fake Image Detection Using Machine Learning. IRACST—International Journal of Computer Science and Information Technology & Security (IJCSITS).

[14]. S. Shalev-Shwartz, & S. Ben-David, (2014). Understanding Machine Learning: From Theory to Algorithms. New York: Cambridge University Press.