Journal of Engineering Sciences

Vol 13, Issue 05, May /2022

ISSN NO:0377-9254

A CONGRESSIONAL INVESTIGATION OF FAKE IMAGES ON SOCIAL MEDIA USING MACHINE LEARNING

[1]M.Ayyappa Chakravarthi, Mallampalli Raghuvamsi[2], Kancharla Sunny Priyal[3], Koppolu Sai Jaideep[4],Makkapati Kartheek[5]

[1]Asst.Professor, Department of CSE, KKR & KSR Institute of Technology and Sciences, Guntur, Andhra Pradesh 522017, India

[2,3,4,5]UG Student, Department of CSE, KKR & KSR Institute of Technology and Sciences, Guntur, Andhra Pradesh 522017, India

Abstract—

Social media plays a significant part in people's everyday lives in this modern age. The majority of individuals often post text, photographs, and videos on social media (e.g. Twitter, Snapchat, Facebook, and Instagram). Images are one of the most widely shared forms of material on social media. As a result, there is a demand for social media image monitoring. Individuals and small organisations may now easily generate these photographs and spread them widely in a short period of time, jeopardising the credibility of the news and public trust in social media. This study aimed to provide a method for extracting picture information, classifying it, verifying its validity, and detecting modifications in digital photographs. Instagram is one of the most popular social networking websites and smartphone picture-sharing apps. Users may use this to snap images, apply digital photographic filters, and then post them. Many undesired items, such as threats and faked photographs, may be found in Instagram postings, posing a danger to society and national security. The goal of this study is to develop a model that can be used to categorise Instagram material (pictures) in order to identify threats and falsified photographs. The model was created using deep learning methods such as Convolutional Neural Networks (CNN), Alexnet networks, and Alexnet transfer learning. According to the findings, the proposed Alexnet network is 97 percent more effective than the other strategies at detecting fraudulent photos. The results of this study will help find strange content and fake photos in photos that people share on social media. They will also help protect social media from electronic attacks and threats.

Keywords—Convolution Neural Network (CNN); Image forgery; Classification; Alexnet; Rectified Linear Unit (ReLU); SoftMax function; Features extraction.

I. INTRODUCTION

It is undeniable that social media has altered the way individuals connect and go about their daily lives. Social networking sites have become a popular media phenomena in recent years, attracting a considerable number of users. The number of users [1] worldwide has already surpassed three billion.

The number of active users in the Gulf area has increased by more than 66 percent [2]. Saudi Arabia is ranked sixth in the world for social media use, with more than 75 percent of its estimated 25 million people [3] using it. The foundations of social media are distinct foundations that connect people together and enable them to express themselves, share their interests and views, and form new friendships with others who share their interests. The most popular social networking sites now are Facebook, Twitter, and Instagram. It is common practice to post photographs on social networking sites like Instagram. Every day, at least 80 million photographs [4] are posted on Instagram.

Users may use Instagram to capture photos, add digital photographic filters, and then publish the photos to the website for social networking with brief descriptions. Every day, billions of photos [5] are uploaded and shared on social media. In our electronic era, a large number of individuals have been victims of picture fabrication. Some criminals misuse software and utilize photographs as evidence to pervert the legal system [17]. To eliminate this, all images shared on social media should be identified as authentic or fraudulent. Social media is an excellent tool for disseminating and exchanging information. People may be duped and even influenced by inadvertent misleading propaganda if there is no vigilance. Though most picture altering using Photoshop is obvious, some of these photographs may seem genuine owing to pixelization and amateurish work [16]. Edited visuals, in particular, may undermine a politician's credibility in the policy arena. The researcher will try to propose a classifier model through a convolutional neural network (CNN) that can take use of information to take a picture from social media and then identify and detect it using machine learning methods [6, 7].

This study provides a strategy that uses an effective system (the CNN model) to classify a picture as input [20]. The findings of this proposed study will aid in the monitoring and tracking of social media material as well as the detection of fraud on social networking sites, particularly in the area of photos.

II. 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 about counterfeit pictures online. Adobe recognizes the way in which Photoshop is misused and has tried to offer a sort of antidote [8]. The following provide 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 According to [15] Kim's and Lee's, 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. 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.

Raturi'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.

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 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 a 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.

In 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 level. 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. It develops and tests reliable fake image detection program by combining the results of metadata analysis (40%) and neural network output (60%).

III. Methodology

The goal of this research is to detect false photographs (Fake images are the images that are digitally altered images). The difficulty with current false picture detection systems is that they can only identify particular types of manipulation, such as splicing and coloring. To identify practically all types of picture manipulation, we used machine learning and neural networks to address the issue.

It is possible to create changes to a picture that are too subtle for the human eye to perceive with the newest image editing tools. Even with a powerful neural network, determining whether a picture is phony or not requires discovering a common element that appears in almost all fraudulent photos. Instead of feeding the neural network raw pixels, we fed it a picture with error levels.

The picture is analyzed on two levels in this research. It examines the picture information at the first level. Because image metadata may be changed using simple tools, it is not very dependable. However, the majority of the photographs we come across will include non-altered information, which will aid in identifying the changes. If a picture is modified using Adobe Photoshop, for example, the metadata will include the Adobe Photoshop version utilized.

The picture is transformed to an error level assessed format and shrunk to 100px x 100px at the second level. Then these 10,000 pixels with RGB values (30,000 inputs) are sent into the Multilayer perceptron network's input layer. Two neurons make up the output layer. One for the fictitious picture and the other for the genuine image. We evaluate whether the picture is false or not based on the value of these neuron outputs and metadata analyser output, as well as the likelihood of the image being tampered with.

Neural network structure

Layer	Neurons
Input Layer	30,000
Hidden Layer 1	5000 - Sigmoid
Hidden Layer 2	1000 - Sigmoid
Hidden Layer 3	100 - Sigmoid
Output Layer	2

Tools Used

Neuroph Studio

Neuroph studio is an open source Java neural network framework that helps to easily build and use neural networks. It also provides a direct interface for loading images

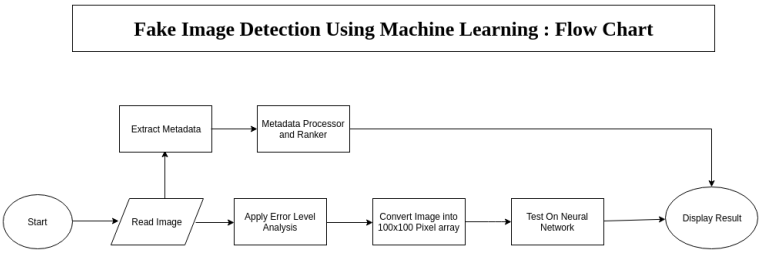
Metadata-extractor

Metadata-extractor is an open source java library used to extract metadata from images.

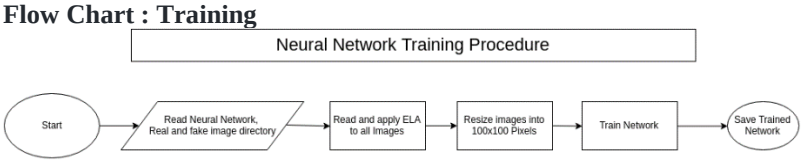
JavaFX

JavaFX is used to implement modern user interface for the application.

Flow Chart : Detection



Flow Chart : Training



2. Error Level Analysis

Error Level Analysis resaves a specific image at a certain error rate, such as 96 percent, and then looks for a virtual change; if one is found, it signifies the cells have hit their local minima for error at that quality level. However, if significant alterations are identified, the pixels are more likely to be original. These studies distinguish between actual and false pixels.

The technology stores an image at 100 percent quality before converting it to a 90 percent quality image. Difference technique is used to determine the difference between these two. The output image is the input

image's needed error level analysis (ELA) image. This image is now stored as a buffered image and delivered to the neural network to be processed further.

Example: -

By displaying changing error levels across the image, Error Level Analysis suggests a kind of digital tampering. Eyes, hair color, nose, and clothing are all mentioned. Instead of their surroundings, all of these traits ultimately reach various levels of error. This suggests that various areas have been brightened and the colors have been changed.

Fig. 3.3 original image Fig. 3.4 ela image

Convolutional Neural Network

A multilayer perceptron neural network with a few hidden layers on both the input and output levels. When an image is selected for review, it is first transformed from the Compression and Error Level Analysis stage to an ELA representation. Because 90% of the photos are utilized to generate the ELA image, the second step is to calculate the ELA. The image is then preprocessed and transformed to a width and height of 100x100px. While representing 10,000 pixels, the image is serialized into an array comprising around 30,000 integer values. Because red, green, and blue components are present in these pixels, 10,000 pixels will have 30,000 values.

The array will be supplied as input to the neural network and output neurons will be established when the data is being trained. The two output neurons represent the false and genuine image, respectively. The neuron is set to one if the image is fraudulent, whereas it is put to zero if the image is true. The image array will be fed into the input neurons during testing, and the output neurons' values will be collected to present the analysis' result.

Learning Transfer

By upgrading the learner, transfer learning transfers knowledge from one domain to the necessary domain. It's a method of creating a model for one activity and then using it as a starting point for another.

Let's say you want to teach two individuals to play the drums. The first individual has never learnt or played any musical instrument, but the second has vast experience with the xylophone. As a result, someone with xylophone expertise will learn drums much more quickly by applying previously learned information to the new job.

When the quantity of target training data is limited, transfer learning is necessary. The main cause for this might be the high cost and scarcity of data. However, we must employ it since it reduces model training time and provides a lower error level.

Model VGG 16

VGG16 is a convolutional neural network design that focuses on having a stride 1 Convolution layer with the same padding and a stride 2 maxpool layer.

The VGG Network employs 3x3 convolutional layers that are stacked one on top of the other. With each layer, the depth grows. It differs from AlexNet in that it substitutes kernel-sized filters with numerous 3x3 kernel-sized filters, resulting in more sophisticated alterations.

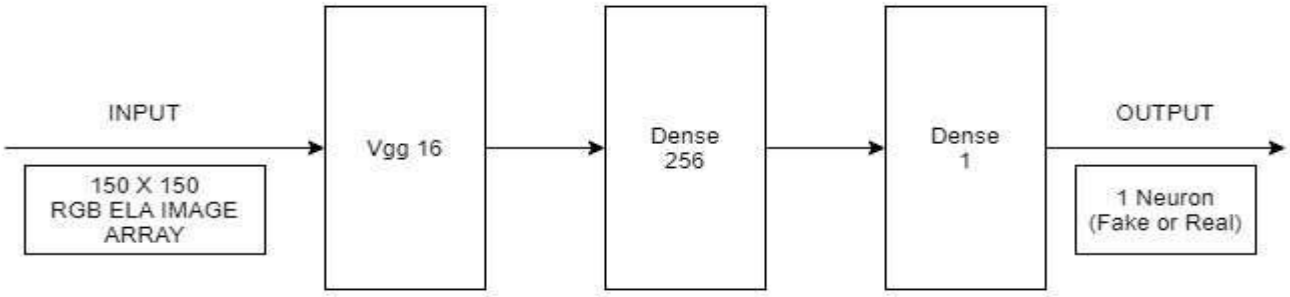


Fig. 3.5 Neural network architecture

IV. RESULTS

We were able to get the following results utilizing transfer learning on the VGG16 model. We can see from the graph below that the approach we used was significantly optimal in both Training and Validation, as our neural network was neither over-fitting nor under-fitting during both training and validation. Despite our limited resources, we were still able to achieve a validation accuracy of 86.12 percent.

```
Test loss: 21.359026432037354
Test accuracy: 90.9704864025116
```

```
Validation loss: 30.717751383781433
Validation accuracy: 86.12499833106995
```

Fig. 4.1 Training and Validation graphs

CONCLUSION

As the internet continues to progress in contemporary society, various social networking platforms such as Facebook, Instagram, and others have been utilized not just for positive purposes, but also for harmful ones by individuals. Crimes against pictures are emerging for illicit objectives in these situations. Such illicit intents must be detected by digital forensics. We suggested image alteration detection approaches utilizing error level analysis in this study. Following a short overview of relevant publications, the suggested model was thoroughly discussed. The suggested model was assessed after extensive testing, and it was shown to have at least 95% accuracy. The suggested model may be used to assess whether or not an image has been modified, and if a better model is developed in future research, it can be used to detect additional manipulation methods. Furthermore, in future study, it will be feasible to apply it to diverse multimedia as well as movies. Crimes against pictures are emerging for illicit objectives in these situations. As a result, digital forensics must discover these illicit uses.

REFERENCES

1. Luo, Weiqi, Jiwu Huang, and Guoping Qiu. "Robust detection of region-duplication forgery in digital image." Pattern Recognition, 2006. ICPR 2006. 18th International Conference on. Vol. 4. IEEE, 2006.
2. S. Gholap and P. K. Bora, Illuminant colour based image forensics, in Proc. IEEE Region 10 Conf. 2008
3. Leida Li, Shushang Li, Hancheng Z -Journal of Information Hiding and Multimedia Signal Processing, Vol. 4, No. 1, pp. 46-56, January 2013.
4. Tiago and Christian et al Exposing Digital Image Forgeries by Illumination Color Classification. IEEE Transactions on Information Forensics and Security (Page: 1182 1194)Year of Publication: 2013.
5. Reshma P.D and Arunvinodh C IMAGE FORGERY DETECTION USING SVM CLASSIFIER Conference on Innovations in Information, Embedde and Communication Systems (ICIIECS), 2015.
6. S.Shaid.”TypesofImageForgery.”Internet:http://csc.fsksm.utm.my/syed/research/image-forensics/11-types-of-mageforgery.html, Feb.08, 2010 12:17 [Dec. 4, 2012].
7. Z. He, W. Sun, W. Lu, and H. Lu. "Digital image splicing detection based on approximate run length," Pattern Recogn .Lett., vol. 32, pp. 1591-1597, 2011.