*A Progress Report*

*on*

FAKE IMAGE DETECTION USING MACHINE LEARNING

*carried out as part of the course IT 1634 Submitted by*

##### RIA BAHETI

##### 189303004

##### RITIK ANAND

##### 189302133

##### VI-IT

*in partial fulfillment for the award of the degree of*

#### BACHELOR OF TECHNOLOGY

In

**Information Technology**



**Department of Information Technology,**

**School of Computing and IT, Manipal University Jaipur, *June, 2021***

# CERTIFICATE

Date : 17 June 2021

This is to certify that the project titled “**Fake Image Detection”** is a record of the bonafide work done by **Ria Baheti** (189303004) and **Ritik Anand** (189302133) submitted in partial fulfilment of the requirements for the award of the Degree of Bachelor of Technology (B.Tech) in **(Discipline)** of Manipal University Jaipur, during the academic year 2020-21.

#### Dept Guide Name : Dr Gaurav Aggrawal

*Project Guide, Dept of Information Technology Manipal University Jaipur*

#### HOD Name : Dr Pankaj Vyas

*HOD, Dept of Information Technology Manipal University Jaipur*

# 

# ABSTRACT

Images are often manipulated with the intent and purpose to benefit one of the parties. In fact, images are often considered as evidence of a fact or reality, therefore, fake news or any form of publication that uses images that have been manipulated in such a way as to have the capability and greater potential for misleading. To detect falsification of the image, image data is needed in large quantities many, and models to process each *pixel* in the picture. In addition, efficiency and flexibility in data training is also needed to support its use in everyday life. Big data and deep learning concepts is the perfect solution for this problem. Therefore, architecture *Convolutional Neural Network* (CNN) which utilize *Error Level Analysis* (ELA), forgery detection image can reach 91.83% and convergence only with 9 epochs.

# 

# 

# 

# TABLE OF CONTENTS

1. Cover Page 1
2. [Certificate… 2](#_TOC_250004)
3. [Abstract… 3](#_TOC_250003)
4. [Contents…](#_TOC_250000) 4
5. [List of Figures…](#_TOC_250001) 5
6. Introduction
   * 1. Background………………………………………………………………6
     2. Purpose……………………………………………………………………7
     3. Problem Statement………………………………………………….7
7. Methodology Overview
   * 1. Method………………………………………………………………8
     2. Limitations………………………………………………………………9
8. Literature review
9. First research paper name…………………………………………….10
10. Design and Implementation………………………………………………………..11
11. Analysis…………………………………………………………………… 14
12. Conclusion………………………………………………………………………17
13. References Appendix…………………………………………………….. 18

# LIST OF FIGURES

|  |  |  |
| --- | --- | --- |
| **Figure No.** | **Figure Title** | **Page No.** |
| Figure 1 | Original image | 11 |
| Figure 2 | Tampered image | 11 |
| Figure 3 | ELA of original image | 13 |
| Figure 4 | ELA of tampered image | 14 |
| Figure 5 | Accuracy Plot | 15 |
| Figure 6 | Confusion Matrix | 15 |

# INTRODUCTION

### BACKGROUND

According to the *EU High Level Expert Group*(2018), fake news is defined as disinformation, which is any form of inaccurate, false or misleading information that is presented, promoted or designed. Behind the fake news, there are several reasons for the publication. One of them is to gain economic benefits, whether it is through increasing the number of news clicks or making news that is not supposed to benefit either party .

In addition, fake news can also affect stock prices, which can benefit those who release the news. Another reason is to gain support or bring down other parties socially or politically .

Based on statistics belonging to the Indonesian Telematics Society (MASTEL) in 2017, the types of fake news that are most often received are socio-political, SARA (ethnicity, religion, and race), health, food and beverage, financial fraud, and science and technology. As many as 84.5% of all respondents stated that they feel disturbed by the existence of fake news, and more than 70% agreed that fake news disturbs community harmony and hinders development.

Apart from writing, around 40% of respondents stated that the spread of fake news is also often accompanied by pictures. Images are used by humans to reproduce reality, and are often used as evidence of news, publications, or facts. Fake news that has a supporting image tends to be accepted and trusted by the public.

In general, humans are easier to remember the form of pictures than writing. According to the *Social Science Research Network*, as many as 65% of people are people who like to learn through visuals. In *marketing*and *visual*science , it is stated that images have a very big influence on an article. People are more likely to respond when there are pictures than just text. According to an infographic with the theme of the influence of images in the world of marketing, images can increase the number of respondents for an article by up to 94% [8]. Therefore, an image is a strong element in disseminating information.

To determine an image is genuine or fake, it is very difficult to see with the naked eye, special techniques and certain accuracy are needed in order to know for sure an image is an original image or has been modified For ordinary people, this may be difficult to do. For this reason, image forgery detection technology needs to be developed, so that it can be used as a means to assist people in determining the authenticity of an image.

This technology requires a lot of image data, and each image has many constituent *pixels*. With ordinary machine learning, this technology would be difficult to develop. Thus, *big data*and *deep learning* is the right solution to solve this image forgery detection problem.

### PURPOSE

Data mining in the form of image forgery detection has two main objectives as follows.

1. Propose a new method using *deep learning*to classify images as original images and modified images with a simpler architecture, thereby reducing computational costs
2. Involving the use of ELA in machine learning as an effort to increase efficiency

There are several impetus behind these two main goals. As is well known, there have been several past studies which also aimed to detect image falsification . However, most of these studies require large computational costs (as can be seen from the number of *epochs*and *layers*required), so that the flexibility of the proposed method is reduced and difficult to apply in everyday life due to computational costs. In fact, there is a need for image forgery detection methods to be able to adapt to the addition of original image data and modifications over time.

### 1.3 PROBLEM STATEMENT

The models which can help us to detect image forgery can be built using machine learning algorithms and ELA for feature extraction.

1. METHODOLOGY

## *METHOD*

There are two main methods used in this data mining, namely Error Level Analysis (ELA) and machine learning with deep learning techniques in the form of Convolutional Neural Network (CNN).

***1. Error Level Analysis (ELA)***

*Error Level Analysis*is one of the techniques used to detect image manipulation by re-saving the image at a certain quality level and calculating the ratio between the compression levels. In general, this technique is performed on images that have a *lossy*format ( *lossy compression*). The image type used in this data mining is JPEG. In JPEG images, compression is performed independently for every 8x8 pixels in the image. If an image is not manipulated, every 8x8 pixel in the image must have the same *error*rate .

***2. Convolutional Neural Network (CNN)***

CNN is a type of *network*based on *feedforward*, where the flow of information is only one way, namely from input to output. Although there are several types of CNN architectures, in general, CNNs have several *convolutional layers*and a *pooling layer*. Then, followed by one or more *fully connected layers*. In image classification, the input to CNN is in the form of images, so that each *pixel*can be processed.

In short, the *convolutional layer is*used as a feature extractor to study the representation of these features from the image that is input to the CNN. Meanwhile, the pooling layer is tasked with reducing the spatial resolution of feature maps. Generally, before the *fully connected layer*, there is a stack of several *convolutional*and *pooling layers*that serve to extract more abstract feature representations. After that, the *fully connected layer*will interpret these features and perform functions that require *high-level reasoning*. The classification at the end of CNN will use the *softmax*function .

## *LIMITATIONS*

There are some limitations that apply to mining this image forgery detection data, namely that the raw data must be an image with *lossy compression*(eg.jpg), nor is it a  *computer generated image*

1. LITERATURE REVIEW

*3.1 Image forgery detection using error level analysis and deep learning, April 2019*

*By Ida Bagus Kresna Sudiatmika, Fathur Rahman, Trisno, Suyoto*

* In this paper, the dataset taken consists of 7491 original images and 5123 tampered images. The size of the dataset is changed to 224x224 pixels and divided into training and test sets.
* We propose a new system from combination Error Level Analysis and Convolutional Neural Network in machine learning and computer vision.
* First error level analysis is used on both the sets which consists of calculating the difference from compression levels.
* Due to insufficient dataset, VGG 16 is used after that to try and get an accurate result. It consists of pre-processing the images, feature extraction using CNN and functional block classification.
* Training accuracy of the model achieved up to 92.2% and for validation 88.46% using 100 epoch.
* Thus, by using the deep learning architecture of VGG 16 in analyzing error level image analysis for image forgery can be applied and get good results on recognition.

1. **DESIGN AND IMPLEMENTATION**

In general, architectural design is divided into two major parts, namely *data preparation*and *model building*. At the initial stage, input data consisting of images with the “.jpg” format, with the following details: 1771 images with *tampered*labels and 2940 images with *real*labels [3], were entered into the *data preparation*stage . The *data preparation*stage is the stage where each image which is the input data is first converted into an *Error Level Analysis*result image . Then, an image of ELA it will be *resize*be an image with a size of 128 x 128.



**Figure 2. a) An example of an original image**

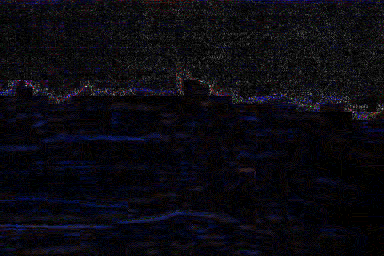


**b) an example of a modified image**

Converting raw data to ELA results is a method used to increase the training efficiency of the CNN model. This efficiency can be achieved because the resulting ELA image contains information that is not as redundant as the original image. The features generated by the ELA image have focused on the portion of the image that has an *error*level above the limit. In addition, *the pixels*in the ELA image tend to have a color that is similar to or very contrasting with *the*adjacent *pixels*, so that training the CNN model becomes more efficient.



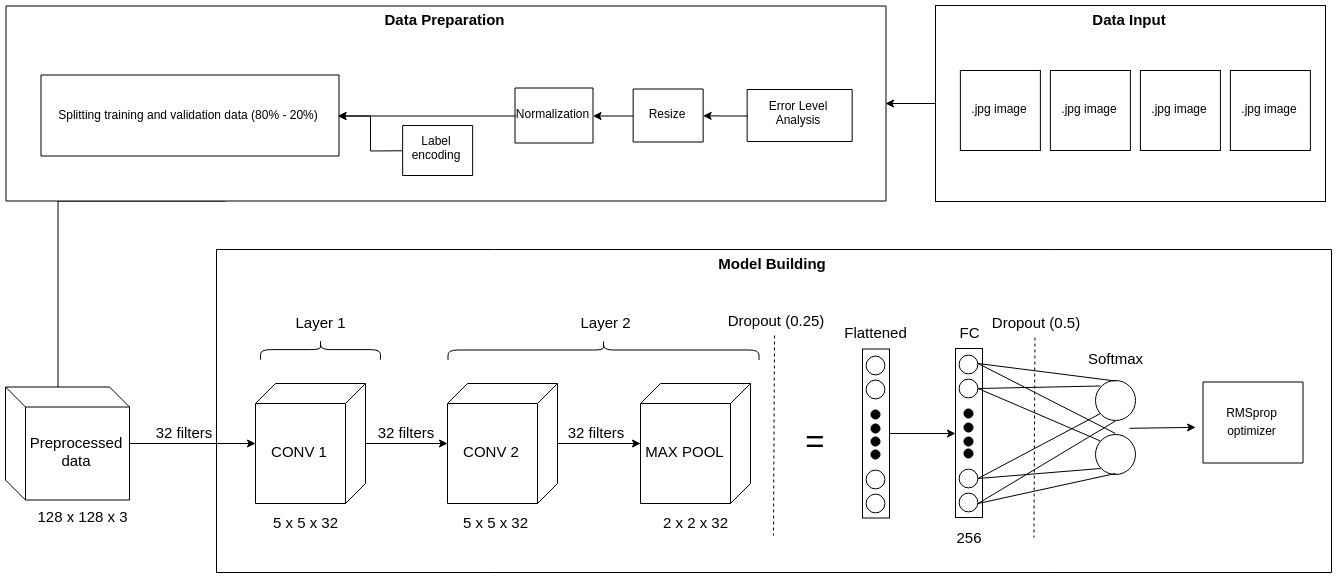
**Figure 3. a) ELA image results from Figure 2a)**



**b) ELA image results from Figure 2b)**

After that, the image size is changed. The next step is to normalize by dividing each RGB value by 255.0 to perform normalization, so that CNN converges faster (reaching the global minimum of the *loss*value belonging to the validation data) because the value of each RGB value only ranges between 0 and 1. The next step is to change the label on a data, where 1 represents *tampered*and 0 represents *real*to *categorical value*. After that, the distribution of training data and validation data was carried out using the distribution of 80% for training data and 20% for validation data.

The next step is to use training data and validation data to conduct *deep learning*model training using CNN. The optimization applied during training is the *RMSProp optimizer*, which is one of the *adaptive learning rate*methods . The complete architecture used in the *model building*section can be seen in the image below or by using the link[] which is a complete architectural drawing.



**Figure 4. CNN model development architecture**

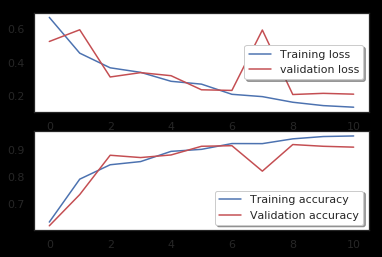
In the *deep learning*model used, the first layer of CNN consists of a *convolutional layer*with a kernel size of 5x5 and a number of filters 32. The second layer of CNN consists of a *convolutional layer*with a kernel size of 5x5 and a number of filters 32, and a *Max Pooling layer*with a size of 2x2. The two *convolutional layers*used use *the glorot uniform kernel initializer*, and the ReLU activation function to make the neurons in the *convolutional layer*select so that they can receive useful signals from the input data [9].

After that, the *MaxPooling*layer added a *dropout*of 0.25 to prevent    *overfitting*.Next Layer is a *fully connected layer*with 256 *neurons*and a ReLU activation function. After *fully connected layer*, a *dropout*of 0.5 will be added to prevent *overfitting*. The *output*layer used has a *softmax*activation function .

In the architecture used, only two *convolutional layers*are needed, because the results resulting from the conversion process into an ELA image can highlight important features to determine whether an image is original or has been modified properly.

**5. ANALYSIS**

The results obtained from the proposed method have a maximum accuracy of 91.83%. Images of accuracy curves and *loss*curves can be seen in the image below.



**Figure 5. A Accuracy curves and loss curves for**

**training data and validation data**

It can be seen in the picture above that the best accuracy is obtained in the 9th *epoch*. The value of *validation loss*after the 9th *epoch*started to flatten and finally increased, which is a sign of *overfitting*. A good method of identifying the number of *epochs*to be used during training is *early stopping*. With this method, the training will be stopped when the validation accuracy value starts to decrease or the *validation loss*value starts to increase.

The number of training *epochs*required is small to achieve convergence, because the use of the ELA converted image feature makes model training much more efficient, and the normalization performed on the RGB values ​​for each *pixel*also accelerates the convergence of the CNN model .

The accuracy results obtained by the model in classifying can be said to be high. This is an indication that the feature is an ELA gam bar image successfully used to classify whether the image is an original image or has been modified.

A screenshot of a computer

Description automatically generated with medium confidence

**Figure 6 :Confusion Matrix**

### Individual Contribution of Project Members

##### Ritik Anand:

*Implemented:* Convolutional Neural Network Classifier, Data Preprocessing

##### Ria Baheti:

*Implemented:*  Convolutional Neural Network Classifier,Data Preprocessing

* 1. CONCLUSION

In this study, there are several things that can be concluded from the results of machine learning using ELA and CNN.

1. CNN uses two *convolutional layers*, one *MaxPooling layer*, one *fully connected layer*, and one *output layer*with *softmax*can reach 91.83% accuracy .

2. The use of ELA can increase efficiency and reduce computational costs of the training process. This can be seen from the reduction in the number of layers from the previous method [11] and the number of *epochs*required. In the proposed model, the number of *epochs*needed to achieve convergence is only 9.

* 1. REFERENCES

[1] zgöbek, zlem, JA Gulla, *“Towards an Understanding of Fake News”, Norwegian Big Data Symposium*(2017).

[2] Kshetri, Nir, J. Voas, *“The Economics of 'Fake News'”, IT Pro (November/December 2017), IEEE Computer Society*.

[3] *Chinese Academy of Sciences. “CASIA Image Tempering Detection Evaluation Database (CAISA TIDE) V2.0*. Taken from [http://forensics.idealtest.org](https://translate.google.com/translate?hl=en&prev=_t&sl=auto&tl=en&u=http://forensics.idealtest.org/)

[4] N. Krawetz, *"A pictures worth digital image analysis and forensics," Black Hat Briefings*, p. 1-31, 2007.

[5] Rawat, Waseem, Z. Wang, *“Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review”*, *Neural Computation 29*(2017), p. 2352-2449.

[6] Gunawan, Teddy Surya, Hanafiah, SAM, Kartiwi, M., Ismail, N., Za'bah, NF, Nordin, AN, *“Development of Photo Forensics Algorithm by Detecting Photoshop Manipulation Using Error Level Analysis”, Indonesian Journal of Electrical Engineering and Computer Science, Vol. 7, No. 1,*July 2017, p. 131-137.

[6] *Photo Forensics: Detect Photoshop Manipulation with Error Level Analysis*, September 2018. Taken from https://resources.infosecinstitute.com/error-level-analysis-detect-i mage-manipulation/#gref

[7] Edelman, *“2018 Edelman Trust Barometer Global Report”*, taken from

https://cms.edelman.com/sites/default/files/2018-01/2018%20Edel man%20Trust%20Barometer%20Global%20Report.pdf

[8] Bulls, Jeff. *“6 Powerful Reasons Why you Should include Images in your Marketing”*, taken from [https://www.jeffbullas.com/6-powerful-reasons-why-you-should-in](https://translate.google.com/translate?hl=en&prev=_t&sl=auto&tl=en&u=http://www.jeffbullas.com/6-powerful-reasons-why-you-should-in)*include*-images-in-your-marketing-infographic /

[9] V. Nair and GE Hinton. *“Rectified linear units improve restricted boltzmann machines,” Proceedings of the 27th International Conference on Machine Learning*, June 21-24 2010, p. 807-814.

[10] Villan, M. Afsal, Kuruvilla, K., Paul, J., Elias, EP, *“Fake Image Detection Using Machine Learning”, International Journal of Computer Science and Information Technology & Security (IJCSITS), Vol. 7, No. 2,*2017.

[11] Rao, Yuan, Ni, J., *“A Deep Learning Approach to Detection of Splicing and Copy-Move Forgeries in Images”, 2016 IEEE International Workshop on Information Forensics and Security (WIFS)*.