Fake Face Detection using Metadata Analyzer and Deep Learning

Mr.C.Murale
Assistant Professor, Dept of Information Technology Coimbatore Institute of Technology
Coimbatore, India
murale@cit.edu.in

Mr.R.Srinivas
Student, Dept of Information Technology Coimbatore Institute of Technology
Coimbatore, India
1907049it@cit.edu.in

Mr.K.Kalkidas
Student, Dept of Information Technology Coimbatore Institute of Technology
Coimbatore, India
1907018it@cit.edu.in

Mr.V.Gokul
Student, Dept of Information Technology Coimbatore Institute of Technology
Coimbatore, India
1907010it@cit.edu.in

Abstract **-** With advancements in technology, it is now possible to create representations of human faces in a seamless manner for fake media, leveraging the large-scale availability of videos. These fake faces can be used to conduct personation attacks on the targeted subjects. This research aims to evaluate the working of different deep learning techniques in the novel "Real and Fake Face detection" dataset by Computational Intelligence Photography Lab, Yonsei University. For the detection of forged faces, the first step of the proposed method is image normalization for real and fake image recognition. Normalized images are then preprocessed and train to different pre-trained deep learning models. We finetune these models for categorization of 2 classes that are forged and real to evaluate these model's performance. From all tested models, VGG models give the best training accuracy of 91.97% and 92.09% on VGG-16 and VGG-19, whereas VGG-16 shows the good test set Accuracy using a smaller number of epochs, which is competitively better than all other techniques. Another feature used along with deep learning models is image metadata.Here it is used as a supporting parameter for deep learning model. Results of these models were evaluated using confusion matrix evaluation measures and compared with state of the art techniques.

Keywords— Deeplearning,VGG 16,Metadata analyzer, VGG 19,CNN.

I. 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 photoshoped 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.

II. THEORY

A. Metadata Analysis

Most image files do not just contain a picture. They also contain information (metadata) about the picture.Metadata provides information about a picture's pedigree, including the type of camera used, color space information, and application notes. Different picture formats include different types of metadata. Some formats, like BMP, PPM, and PBM contain very little information beyond the image dimensions and color space. In contrast, a JPEG from a camera usually contains a wide variety of information, including the camera's make and model, focal and aperture information, and timestamps.

PNG files typically contain very little information, unless the image was converted from a JPEG or edited with Photoshop. Converted PNG files may include metadata from the source file format.

Metadata provides information related to how the file was generated and handled. This information can be used to identify if the metadata appears to be from a digital camera,processed by a graphical program, or altered to convey misleading information. Common things to look for include:

1) Make, Model, and Software

These identify the device or application that created the picture. Most digital cameras include a Make and Model in the EXIF metadata block. (However, the original iPhone does not!) The Software may describe the camera's firmware version or application information.

2) Image size

The metadata often records the picture's dimensions. Does the rendered image size (listed at the bottom of the metadata) match txhe other sizes in the metadata? Many applications resize or crop pictures without updating other metadata fields.

3) Timestamps

Look for fields that detail timestamps. These typically identify when a picture was taken or altered. Do the timestamps match the expected timeframe?

4) Types of metadata

There are many different metadata types. Some are only generated by cameras, while others are only generated by applications.

5) Descriptions

Many pictures include embedded annotations that describe the photo, identify the photographer, or itemize alteration steps.

6) Missing metadata

Are any metadata fields missing? If the picture came from a digital camera, then it should have camera-specific information. Some applications and online services strip out metadata. A lack of specific metadata usually indicates a resaved picture and not an original photo.

7) Altered Metadata

Metadata is analogous to the chain of custody for evidence handing. It can identify how a picture was generated, processed, and last saved. However, some people intentionally alter metadata. They may edit timestamps or photo information in an attempt to deceive.

B. Machine Learning

The process of machine learning is similar to that of data mining. Both systems search through data to look for patterns.However, instead of extracting data for human comprehension as is the case in data mining applications machine learning uses that data to detect patterns in data and adjust program actions accordingly. Machine learning algorithms are often categorized as being supervised or unsupervised. Supervised algorithms can apply what has been learned in the past to new data. Unsupervised algorithms can draw inferences from datasets.

III. LITERATURE SURVEY

Sudiatmika [1] proposed an image forgery classification technique using Error Level Analysis to determine the compression ratio between the original image and the fake image. When images were compressed, the original image and the fake images were different. Results of performed experiment showed training accuracy but no test accuracy.

Meera [2] used CASIA TIDE v.1 Dataset for forgery detection using Gabor Wavelets and Local Phase Quantization.

Hakimi [3] proposed a method in which they used chromatic components to enhance image forgery detection. The performance of Cb is almost like that of Cr on CASIA v2.0 dataset. The results are comparable to each other, which shows the consistency of their proposed research.

Jing [4] used self-evaluation and the purpose of creating CASIA Dataset for image tampering recognition.

Li [5] used Forgery detection using the block artifact grid technique on images downloaded directly from NASA.

Muhammad [6] proposed method Passive copy move image forgery detection in which they analyzed image forgery using dyadic wavelet transform.

Akhtar [8] suggested an effective method for handling the Digital Image Forgery Detection problem. LBP and HOG was used for feature extraction.

Muhammad [9] proposed a new method for detecting forgery based on ELA; the method was evaluated on CASIAv2.0 publicly available datasets. These experiments showed different accuracies using different learning rates. Gypsy

Nandi [10] discussed some of the most effective forgery detection techniques that helped forgeries identify images , either single or composite.

Mahale, V. H [12] presented a unique approach based on Local Binary Pattern (LBP) for evaluating image inconsistency and detection, the system was tested on COMOFOD dataset.

Mohamad [13] used an efficient method of combined un decimated wavelet transform from which scale-invariant feature transform were judged from evaluation measures.

Wu-Chi [14] summarized some forgery methods based on watermarking and alpha mattes to analyze images.

Many state-of-art techniques worked on image detection and recognition using different feature extraction techniques and CNN's. This research presents forgery recognition on a novel "Real and Fake Face detection" dataset using deep learning models and convolutional neural networks. We also compared our results of each deep learning model with other and with state-of-art techniques.

.

IV. PROPOSED METHOD

Our dataset consists of RGB images with an extension of (.jpg) which were classified into real and fake classes. After preprocessing the images of the whole dataset, it was converted into ELA images. These images were split up into training and test sets, which were then forwarded to the Deep CNN model to recognize real and fake images. The proposed method is shown in Fig.1.

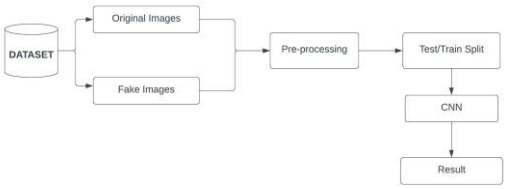


Fig. 1. Proposed method

A. Pre-processing

The first step of the proposed approach was preprocessing in which the whole dataset was resized into (128*128) pixels, which make the whole dataset normalized.

B. Metadata analysis

The entire system is developed using python programming language. For extracting metadata of images, pillow library is used. Metadata-extractor is able to extract metadata information of large no of different image types. Once an image is selected for processing, it is tunneled into 2 separate stages. First stage is metadata analysis. After

extracting metadata, the metadata text is fed into metadata analysis module.

Metadata analyzer is basically a tag searching algorithm. If keywords like Photoshop, Gimp, Adobe or DateTimeOriginal,DateTimeDigitized,etc. is found in the text and then the possibility of being tampered is increased. Two separate variables are maintained which are called fakeness and realness. Each variable represents the weight of being real or fake image. Once a tag is taken, it is analyzed and corresponding variable is incremented by a certain predefined weight. After processing the entire tags, final values of fakeness and realness variable is fed into the output stage as shown in Fig 2.

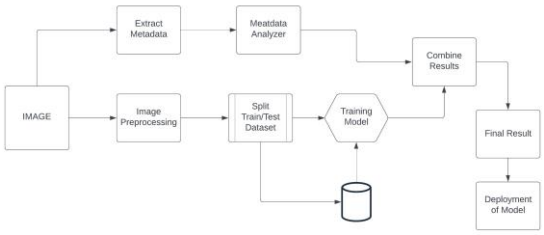


Fig. 2. System Architecture

C. Image Recognition using Deep Learning

Once pre processed images were converted and all images were saved in NumPy array and split into train set and test set and then the train set was passed to the finetuned pretrained deep learning models. Pre-trained model were used to evaluate our dataset results, which are VGG-16. This models was already pretrained.We changed these models' input layers as per the size of pre-trained image size (128*128) and the last layers were changed according to the number of dataset classes. We modified this model by adding sequential model layers as output layers.

1)Training parameters for deep learning models: We used the object recognition models, which were already pre trained on the many image datasets. Using deep learning, we finetune these models for "Fake and Real face detection" dataset for image recognition. RMSprop was used as the optimization function with 0.0001 learning rate and batch size 5 per epoch.

D. Convolutional neural network

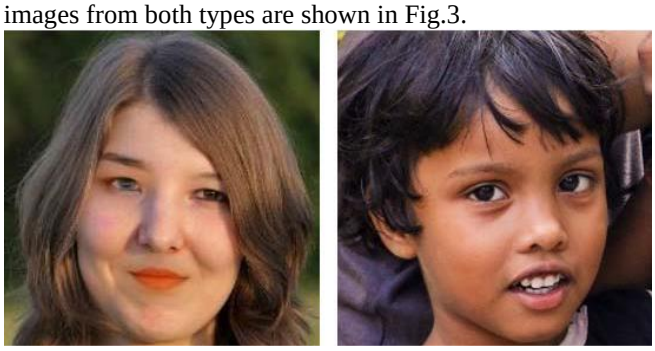
Using keras libraries, we used the flatten layer and dense layer as a sequential CNN model. After images were trained from a finetuned preprocessed model, they were passed to the flatten layer, which transformed the feature map by flattening them into the feature vector and forwarding them to a fully connected layer.

The fully connected layer was used for pattern recognition in the last dense layer of models, SoftMax activation function was used to convert the feature vector in a probabilistic manner. Based on SoftMax activation, the training set was compared with the test set and return a probability distribution on real and forged images. These layers were used to finetune all the deep learning models used in this research.

V. IMPLEMENTATION AND RESULTS

A. Dataset

"Real and Fake Face detection" by Computational Intelligence Photography Lab, Yonsei University. This novel dataset consists of 2041 facial images that were further divided into a fake (960) and real (1081). Sample images from both types are shown in Fig.3.



(a) Real image *(b) Fake image*

Fig. 3.Example of some images from dataset

B. Data preprocessing

Machine learning models cannot use real-time data which contains missing values, noisy data and it also may be in an unsuitable format . To rectify the data, data preprocessing method is used. The Python libraries that we have used to perform data preprocessing are pandas, numpy, label encoder and matplotlib.

Fig. 4.Preprocessed Images

C. Experiments

This section presents the experiments performed using different deep learning pre-trained techniques on "Real and Fake Face detection" dataset.

1)VGG-16: In the first experiment, ELA converted training set images were passed to finetuned VGG-16, the simplest model consisted of 22 layers, we changed its input layer according to our dataset images size of 128 and replaced its last two layers with sequential CNN layers, then used early stopping to prevent the model from over-fitting. The confusion matrix of VGG-16 model is shown in Fig.5.

Training parameters were used for the evaluation of results produced by the model. This technique gave Accuracy of 91.97% training and 64.49% test using 10 epochs. Model truly predict 267 images from 414 test set images, from which 116 were fake and 151 real images.

D. Performance evaluation and metrics

a. Confusion matrix

The performance of our ensemble model was measured using a confusion matrix. Two or more types of classes were obtained as an outcome of the confusion matrix. ACTUAL and PREDICTED were the two dimensions of the confusion matrix. If the value obtained for actual and predicted class is 1, then it falls under true positives (TP). If the value obtained for actual and predicted class is 0, then it falls under the category of true negatives (TN). If the value obtained for actual and predicted class is 0 and 1 respectively, then it falls under false positives (FP). If the value obtained for actual and predicted class is 1 and 0 respectively, then it comes under false negatives (FN).

b. Accuracy

It is the most often used statistic for assessing the performance of classification algorithms. It is expressed as the number of right predictions [4]. The following formula [5] is used to compute.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)} \quad (1)$$

For determining, accuracy of our classification model13], accuracy score function has been used.

b. Precision

The number of correct documents put back by our proposed model can be termed as precision. The formula which is used to determine precision based on a confusion matrix [5][13].

$$\text{Precision} = TP / (TP + FP) \quad (2)$$

c. Recall

Recall is nothing but the number of positives put back by our proposed model. For the determination of recall using the confusion matrix [5][13], the formula is

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive + False Negative}} \quad (3)$$

d. F1 score

The F1 score is calculated as average of precision and recall. F1 will have a highest value of 1 and a worst value of 0. The formula used to compute F1 score is [1] [5]

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision + Recall}} \quad (4)$$

e. Support

Number of samples of the genuine response that fall into each class of target values is termed as support [13].

E. Comparison

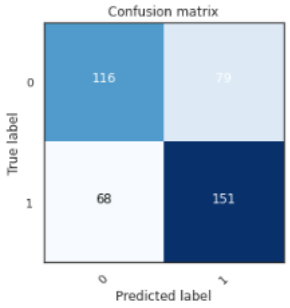


Fig. 5. Confusion matrix of VGG 16

Measures / Models	Precision	Recall	Accuracy	F1-score	Error
VGG-16	0.65	0.68	0.64	0.66	0.35

Fig. 6.Evaluation Measure of model

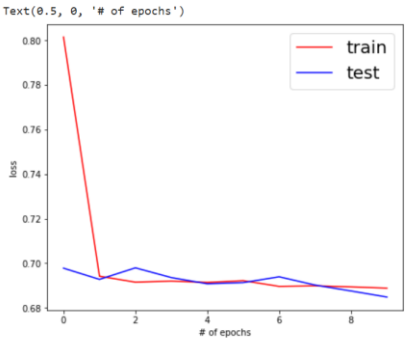


Fig. 7. Value loss of Model

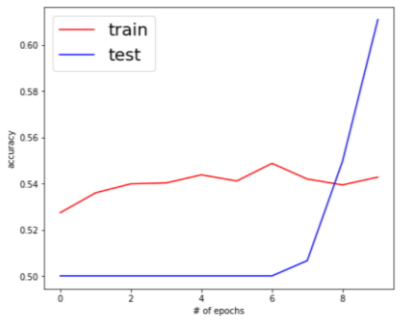


Fig. 8. Value Accuracy of Model

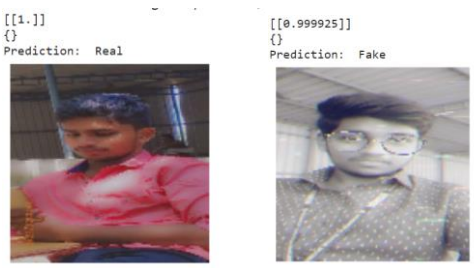


Fig.9. Prediction of Real and Fake Faces

VI. CONCLUSION

A user friendly, self-intuitive dataset of customized faces is created to enable ease of use for both society and government. The otherwise cumbersome and heavyweight models is made simpler with the use of VGG 16 framework. We have achieved the maximum accuracy of 91.97% using the VGG 16 algorithm and metadata analyser, which is much better than the results provided by other existing systems.

VII. SCOPE FOR FUTURE

In future the size of the customizable dataset can be increased by adding more sample images. The dataset could also be made available on a cloud to make it universally accessible. We could detect faces to estimate whether the face is real or get morphed. We could export this model and estimate Originality of face in that picture..An interface can be made for the ease of accessing the system.

VIII. REFERENCES

[1]Grampurohit Sudiatmika, I. B. K., & Rahman, F. (2019). Image forgery detection using error level analysis and deep learning.Telkomnika,17(2), 653-659.

[2] Isaac, M. M., & Wilscy, M. (2015). Image forgery detection based on Gabor wavelets and local phase quantization. Procedia Computer Science, 58, 76-83.,

[3] Khan Hakimi, F., Zanjan, I., & Hariri, I. (2015). Image-splicing forgery detection based on improved lbp and k-nearest neighbors algorithm. ElectronInf Plan,3.

[4] Dong, J., Wang, W., & Tan, T. (2013, July). Casia image tampering detection
eva`luation database. In 2013 IEEE China Summit and International Conference
on Signal and Information Processing (pp.422-426). IEEE.

[5] Li, W., Yuan, Y., & Yu, N. (2009). Passive detection of doctored JPEG image via block artifact grid extraction. Signal Processing, 89(9), 1821-1829.

[6] Villan, Muhammed Afsal, et al. "Fake Image Detection Using Machine Learning." IRACST-International Journal of Computer Science and Information Technology & Security (IJCSITS) (2017).

[7] Hsu, Chih-Chung, Yi-Xiu Zhuang, and Chia-Yen Lee. "Deep fake image detection based on pairwise learning." Applied Sciences 10.1 (2020): 370.

[8] Akhtar, F., & Qayyum, H. (2018). Two Fold Image Forgery Detection System Using Combined Key point based method and Block based Method. Journal of Information Communication Technologies and Robotic Applications, 62-70.

[9] 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).

[10] Sarma, B., & Nandi, G. (2014). A Study on Digital Image Forgery Detection. International Journal of Advanced Research in Computer Science and Software Engineering,

[11] Jeon, Hyeonseong, Youngoh Bang, and Simon S. Woo. "Facing off fake images using fake detection fine-tuning network." IFIP International Conference on ICT Systems Security and Privacy Protection. Springer, Cham,2020.

[12] Ramkumar,] Mahale, V. H., Ali, M. M., Yannawar, P. L., & Gaikwad, A. T. (2017).Image inconsistency detection using local binary pattern(LBP). Procedia computerscience, 115, 501-508.

[13] Hashmi, M. F., Anand, V., & Keskar, A. G. (2014). Copy-move image forgery detection using an efficient and robust method combining undecimated wavelet transform and scale invariant feature transform Aasri Procedia, 9, 8491.

[14] Hu, W. C., Chen, W. H., Huang, D. Y., & Yang, C. Y. (2016). Effective image forgery detection of tampered foreground or background imagebased on image watermarking and alpha mattes.Multimedia Tools and Applications, 75(6), 3495-3516.

[15] Tariq, Shahroz, et al. "Detecting both machine and human createdfake Face images in the wild." Proceedings of the 2nd international workshop on multimedia privacy and security. 2018.