# 딥페이크 데이터 셋을 이용한 딥페이크 영상 검출 방법의 성능 비교

Rafiul Hasan Khan*, 이석환**, 권기룡*

*부경대학교 IT융합응용공학과, **동아대학교 컴퓨터공학과

# *Performance Comparison of Deepfake Image Detection Methods using Deepfake Dataset*

Rafiul Hasan Khan*, Suk-Hwan Lee**, Ki-Ryong Kwon*

* Dept. of IT Convergence Engineering, Pukyong National University,

** Dept. of Computer Engineering, DongA University,

**Abstract**

The inception of artificial image generation was groundbreaking in the field of image processing. However, with mass popularity and difficulty in detection, this innovation is posing threats and concerns towards society. In this paper, we have comprised a study on the recent techniques that have been working on fake image detection. These techniques are based on deep learning methods and they are being regarded as the most efficient methods till now. For this research, we have used deep learning methods i.e. Mesonet, EfficientNet, GoogleNet, VGG-16, and VGG-19 to detect deep fake images. For this research, we have used the Deepfake Dataset of Mesonet article. This Deepfake Dataset is made up of fake images collected from 175 forged videos and real face images from various internet sources. We have implemented all those deep learning methods onto the Deepfake Dataset and compared them based on their total accuracy rate. Our study shows that VGG-19 outperformed the other methods with an overall accuracy rate of 97.2%. Although VGG-19 is computationally more expensive, the accuracy puts it at the top.

## I. Introduction

With the advent of social mediums and available smart gadgets, images and videos have become far more accessible than at any time in history. According to the statistics of Facebook, more than 300 million photos get uploaded every day. All of these photos are not original rather they are a mix-up of tempered or altered images. Many freeware software and enthusiastic research have paved the way towards the rise of fake images. The field of digital image forensics research is dedicated to the detection of image forgeries to regulate the circulation of such falsified contents [1]. Recently, deep learning methods have been established as successful methods for digital image forensics. Thus, in this research, we have compiled a comparative study on deep fake image detection based on the recent deep learning methods. There are thousands of deep fake videos to analyze but we have chosen MesoNet's [1] Deepfake dataset so that we can have a common platform to

compare. We believe, this study will give a brief insight into digital image forensics and will guide future research.

## II. Methods

### 2.1 MesoNet

MesoNet [1] automatically and efficiently detects face tampering in videos. Traditional image forensics techniques struggle to detect fake image because the fake videos get degraded due to the high data compression. Mesonet follows a deep learning approach and presents two networks, both with a low number of layers to focus on the mesoscopic properties of images. The proposed two architectures namely "Meso-4" and "MesoInception-4" to solve these problems while producing 27,977 and 28,615 trainable parameters respectively.

### 2.2 EfficientNet-b0

EfficientNet-b0 [2] is the base network of the EfficientNets group. Although EfficientNet-b7 is the strongest network among them, however, they were built on EfficientNet-b0. Moreover, computationally EfficientNet-b0 is the least expensive network among them as it produces 5.3 million parameters whereas, EfficientNet-b7 produces 66 million parameters. EfficientNets [2] are the most promising networks in the field of digital image forensics.

### 2.3 GoogleNet

GoogleNet [3] is one of the most efficient networks for classification tasks. With features such as 1x1 convolution or modified inception module, it successfully applies dimensionality reduction as well as does not compromise with the performance. GoogleNet [3] produces 7 million parameters making it one of the least expensive networks in the least.

### 2.4 VGG

VGG or Visual Geometric Group is a series of the convolution network model starting from VGG11 to VGG19. The main intention behind it was to understand how the depth of convolutional networks affects the accuracy of the models of large-scale image classification and recognition. The VGG-16 has 13 convolutional layers and 3 fully connected layers while VGG-19 has 16 convolutional layers and 3 fully connected layers. The overall structure includes 5 sets of convolutional layers, followed by a MaxPool. The difference between all the VGGs is the increase in the depth as we move from VGG11 to VGG19 more and more cascaded convolutional layers are added in the five sets of convolutional layers. Both the VGG-16 and VGG-19 produce 138 million parameters making them the most computationally expensive networks on our list

## III. Results and Discussion

### 3.1 Dataset

Table 1: Deepfake Dataset.

| Set | Forged Class | Real class |
|---|---|---|
| Deepfake training | 5103 | 7250 |
| Deepfake testing | 2845 | 4259 |

The Deepfake dataset was created using Deepfake technique. Deepfake is a technique that aims to replace the face of a targeted person with the face of someone else in a video [1]. After some modifications, Deepfake technique was created into a user-friendly application called FakeApp. Deepfake images were created from 175 rushes of forged videos. All videos are compressed with different compression levels. All the faces

have been extracted using the Viola-Jones detector [5] and aligned using a trained neural network for facial landmark detection [6]. Then, the dataset has then been doubled with real face images, also extracted from various internet sources and with the same resolutions. Precise numbers of the image count in each class as long as the separation into a set used for training and model evaluation can be found in Table 1.

### 3.2 Experimental Results

We did transfer learning with all the above-mentioned deep learning methods using the Deepfake dataset. The dataset was divided into eighty percent to twenty percent for training and validation/testing. After the division, we applied image augmentation to the training dataset. For all the simulations, we used the Stochastic Gradient Descent with Momentum (SGDM) and set an initial learning rate of 0.0003. We set the max epoch to 30 and the mini-batch size to 16. Also, we applied shuffle after every epoch. Finally, we trained them on a single NVIDIA GeForce RTX 2070 16GB GPU.

Table 2: Classification Score.

| Network | Deepfake Classification Score | |
|---|---|---|
| Class | Forged | Real |
| Meso-4 | 0.882 | 0.910 |
| MesoInception-4 | 0.934 | 0.900 |
| EfficientNet-b0 | 0.913 | 0.934 |
| GoogleNet | 0.950 | 0.965 |
| VGG-16 | 0.955 | 0.9685 |
| VGG-19 | 0.964 | 0.9765 |

Our simulation was done on individual frame / image. Table 2 displays the deepfake classification score of all the networks. Apparently, Meso-4 is the least performing network whereas, VGG-19 outperformed others.

Table 3 displays the overall accuracy of all the networks. The VGG-19 topped the chart with an accuracy rate of 97.2% whereas. Meso-4 achieved the least accuracy rate of 89.1%.

Table 3: Accuracy Rates.

| Network | Accuracy |
|---|---|
| Meso-4 | 0.891 |
| MesoInception-4 | 0.917 |
| EfficientNet-b0 | 0.922 |
| GoogleNet | 0.958 |
| VGG-16 | 0.963 |
| VGG-19 | 0.972 |

## IV. Conclusion

According to our study, VGG-19 proved as the most accurate network. However, VGG-19 is also the most computationally expensive network. Since our concern is about deepfake detection, VGG-19 fits to our choice. In the future, we will extend our study by experimenting on more networks i.e. EfficientNet-b2 to EfficientNet-b7 to find a least computationally expensive method for deepfake detection.

## Acknowledgment

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2020R1F1A1069124 and 2020R1I1A3066594) and by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2020-0-01797) supervised by the IITP(Institute of Information & Communications Technology Planning & Evaluation).

## [Reference]

[1] D. Afchar, V. Nozick, J. Yamagishi, and

I. Echizen, “MesoNet: a Compact Facial Video Forgery Detection Network,” *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 1 – 7, Dec. 2018, doi: 10.1109/WIFS.2018.8630761.

[2] M. Tan and Q. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” *International Conference on Machine Learning*, pp. 6105 – 6114, May 2019, Accessed: Feb. 04, 2021.

[3] C. Szegedy et al., “Going Deeper With Convolutions,” *arXiv:1409.4842*, pp. 1 – 9, 2015, https://arxiv.org/abs/1409.4842,

[4] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition.,” *The 3rd International Conference on Learning Representations, ICLR 2015*, San Diego, CA, USA, May 2015,

[5] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, p. I – I, Dec. 2001, doi: 10.1109/CVPR.2001.990517.

[6] D. E. King, “Dlib-ml: A Machine Learning Toolkit,” *Journal of Machine Learning Research*, vol. 10, no. 60, pp. 1755 – 1758, 2009.