Fake Image Detection Using Deep Learning

Raidah S. Khudeyer, Noor M. Al-Moosawi
College of Computer Science & Information Technology, University of Basrah, Iraq
E-mail: raidah.khudayer@uobasrah.edu.iq, almoosawinoor2@gmail.com

Keywords: fake detection, CNN, efficientNetB0, synthetic media, deep learning

Received: March 13, 2023

With the emergence of numerous electronic communication programs and image processing programs, as well as an increase in the number of people who use them with a zeal for publishing everything related to their lives and their special pictures and their fear of those who might use these pictures for malicious or humorous purposes, it has become necessary to have specialized and precise systems to determine whether a picture is real or fake. Our work aims to detect real and fake faces by using and modifying one of the most efficient CNN architectures, EfficientNetB0, after improving the architecture with additional fully connected layers and efficiently training the model by using the Adam optimizer and a scheduler learning rate technique. Our findings on the well-known 140k-real-and-fake-faces Kaggle dataset showed state-of-the-art accuracy with the lowest error rate. We achieved 99.06% accuracy, and 0.0569 error rate respectively.

Povzetek: Predstavljena je metoda globokih mrež, ki z uporabo EfficientNetB0 in optimizatorja Adam na Kagglovih 140.000 obrrazih ločuje prave in lažne obraze z 99,06% točnostjo.

1 Introduction

A fast-expanding field, synthetic media is media created by technology. Because of this, artificial media may also be referred to as "AI-generated media". Some examples of synthetic media include music composed by AI, text generation, imagery and video, voice synthesis, and fake images.

In recent years more research has excelled and increased attention on CNN in several areas, such as the classification of images, object detection, facial recognition, fingerprint analysis, computer-aided diagnosis, and facial expressions [1], [2].

The appearance of and the rapid advancement in deep learning makes the detection of authentic and manipulated facial images and video clips more difficult, which is called Deep Fake. Subsequently, the need for techniques that can discover the integrity of digital visual content as images, videos, text so on, is important.

Machine-based algorithms are crucial in detecting fake images, which can take many different forms. Modeling these methods as binary classification issues. It acquires hand-crafted features, investigates hidden information, and separates fake images from editing procedures like an improvement (histogram equalization, color alteration, etc.), geometry modifications (rotation, cropping, shearing), and content changes (copy-move, cut-paste, etc.) [3].

End-to-end learning solutions based on Deep Learning (DL), mainly using Deep Convolution Neural Networks (DCNNs), are created to benefit from automated learning and feature extracting [4-8].

The related research in fake detection started using deep learning and the CNN model since 2018, as summarized in Table 1, in [9] the researchers mention methods that are used to fake detection. A standard CNN consists of multiple components such as convolutional layers, pooling layers, and fully-connected layers. It is created to automatically pick up on spatial feature hierarchies by learning them using a backpropagation algorithm. CNN architecture receives an input image after going through several building blocks and can tell the difference between real and fake faces. The hyperparameters that need to be optimized during CNN training include learning rate, batch size, activation layer, regularization method, and optimizer. in 2019 proposed an efficient network called Efficient Net [10], although the rapid development of convolutional neural network gradually its deficiencies was reason to replace it with pertained models such as Resent, mobile Net, ..., which are required to get more accuracy accordingly the network depth, network width, and input image resolution that need to be manually adjusted [11].

Training a large deep-learning model is a difficult optimization task due to many difficulties including overfitting, underfitting, finishing derivatives, and choosing suitable hyperparameters [12]. The learning rate is one of the most difficult hyperparameters to set and it has a big impact on how well models perform. The traditional neural network training technique can improve performance and speed up training on some issues by utilizing a learning rate that modifies throughout training [13].

In 2018 [3], the researchers used GANs to create fake faces with multiple resolutions and sizes, then used different DCNN models to detect fake images. They apply a deep-face recognition system to transfer weight, and the network is fine-tuned suitable by using real or fake images in the AI Challenge.

By using the empirical knowledge that the textures of fake faces are very different from those of real faces, the researcher in [14] developed a new architecture called LBP (Local Binary Pattern)-Net in 2021 to detect fake faces, then used ensemble learning which outperformed single model. The ensemble model was selected as robust for detecting fake images.

In another work [15], they present a unique technique for identifying fake faces utilizing features based on Image Quality Assessment (IQA). The majority of the discriminative information will be present in the frequency domain of those images, despite the visual appearance of the original and fake ones being identical. Based on such intuition, they depend on derived frequency- and spatial-domain-based parameters for image quality.

In [16] 2022 produced work based on CNN and DL to distinguish between real and fake images. The best model with the highest accuracy and lowest error rate was determined to be the ResNet50 model after training on 9,000 images over 150 epochs.

Table 1: A summary of the literatures on face detection using different techniques.

research	year	method	accuracy
N. Do [3]	2018	DCNN	80%
Y. Wang [14]	2021	LBP-Net	87.60%
S. [15]	2022	IQA	99%
F. M. Salman [16]	2022	ResNet50	99%

In this research, you will discover how improving one of well none CNN models, EfficientNet-B0 by adding more denes layers and how using the learning rate scheduling technique can enhance performance and reduce error rates in the modified EfficientNetB0 model.

2 EfficientNet-V2 network

EfficientNet employs an efficient and sample compound factor to scale the network in three dimensions, including network depth, network width, and the resolution of the input image, to obtain the optimal set of parameters. Depending on the degree of scaling, the EfficientNet series networks can be divided into eight subnetworks, B0–B7. The EfficientNet is much faster with fewer parameters and has more capability of feature extraction than other classic CNN models [17].

The original architecture is depicted in Figure 1. It is possible to gain from the ideal weights obtained from the training using auto-augmentation because the original networks were trained on ImageNet data, which necessitates a significant amount of resources and techniques that are not part of the model structure.

It should be noted that each model's input takes a different form, with the possibility of removing the top layers to prepare the model for transfer learning.

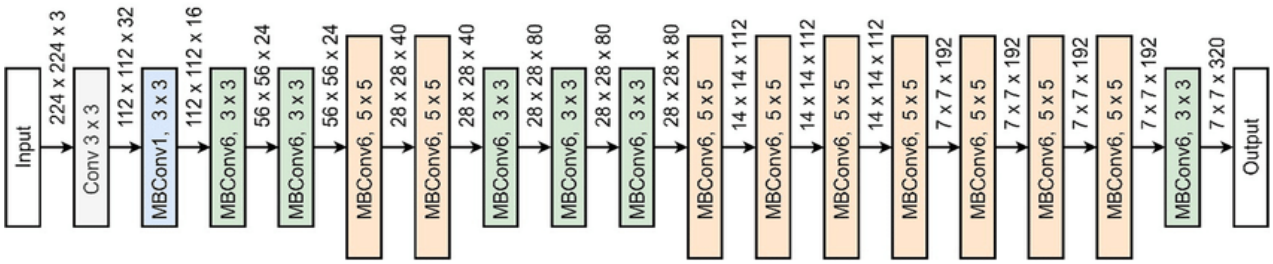


Figure 1: EfficientNetB0 baseline architecture.

3 Methodology

Deepfake detection is typically viewed as a binary classification problem in which classifiers are used to separate real from fake images. This type of approach requires a big dataset of real and fake images in order to train such a classification model and evaluate the final trained model; we utilize 140k actual and fake faces. Kaggle dataset [9] contains a million artificial faces produced using a style GAN with a resolution of 256X256 and 70k real faces obtained from Nvidia's Flickr dataset. In our research, we'll use EfficentNetB0 to transfer learning, fine-tune, and modify the EfficentNetB0 model using a variety of techniques, as follows:

3.1 Model1: Transfer learning method

In this model, we used one of the concepts in machine learning transfer learning that is means reusing a pre-trained model as the starting point for a model on a new task, where transfer knowledge learned in a new task [18], as shown in Figure 2.

It is an optimization task that allows rapid and improved progress when modeling the second task. We used the efficientnetb0 model weights that had already been trained on the ImageNet dataset and saved to initialize the kernels weights on the convolutional part after replacing the top fully connected layer of the original architecture with a custom two-node layer to fit the final binary classification task (real or fake). Therefore, our model has a higher starting point, considerably reducing training time and improving performance. Then retrain the model to fine-tune the kernel's weights based on new data, and the model will be more specific to our task.

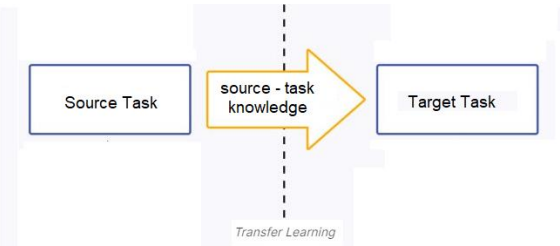


Figure 2: Transfer learning in machine learning with the source of information transfer knowledge from one task to the target task.

3.2 Model2: Modifying EfficentNetB0 model

The dense Layer or fully-connected is a layer of neurons that deploy connected from all the neurons of the previous layer. It is used for classifying tasks depending on the output of convolutional layers [19].

In model 2, the architecture of model 1 was improved by replacing the single layer in the fully connected parts with two additional dense layers with the help of drop-out and batch normalization techniques, which helped combat overfitting and achieve faster convergence. Figure 3 explains model 2 architecture.

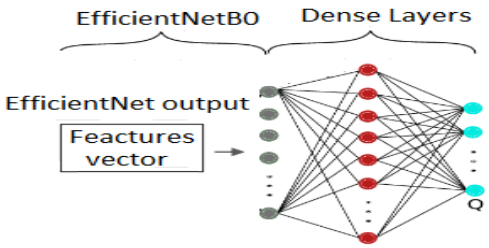


Figure 3: model 2, add two dense layers to EfficientNet architecture.

3.3 Model3: Modifying EfficentNetB0 model with learning rate schedules technique

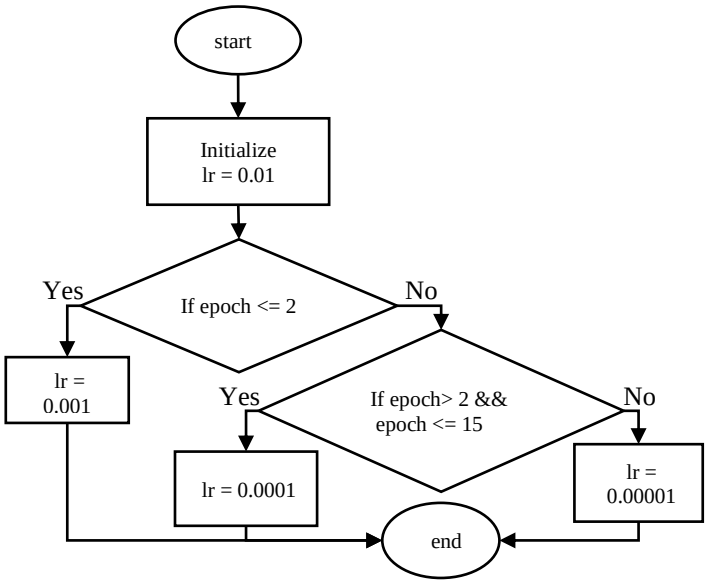


Figure 4: Flowchart of training model3 with learning rate schedule.

The model3 was created in an effort to increase performance and shorten training time after monitoring the performance of model 2 on a dataset and demonstrating the effectiveness of the modifying efficientb0 model and its fit for our task.

While the standard optimization strategy updates network weights at a constant rate for each training epoch, we apply techniques that gradually lower the learning rate by adopting a learning rate that varies according to the training iteration.

Figure 4 shows a flowchart of the learning rate schedule used in our task. This technique has the advantage of making significant weight adjustments early in the training process when higher learning rate values are being used, then lowering the learning rate so that smaller training updates are given to weights later in the training process when the learning rate is smaller. This results in learning good weights rapidly and early.

For training all models, we are leveraging the Kaggle-provided GPU to compute this enormous dataset because it significantly reduces the preprocessing, computation, and training times. For a fair comparison between models, all models were optimized using binary cross-entropy loss and the Adam optimizer. Regarding hyperparameters, we used a batch size of 64 and trained for 30 epochs. Table 2 illustrates all model layers and techniques. The models have been trained a total of 10 separate times, and the models with the best performance on the validation set were saved using the checkpoint technique to be used for final testing on the test set. The early stopping technique was used to avoid wasting time and using conservative weights if the model's performance did not improve after six training epochs.

Table 2: Our 'models architectural details. EB refers to EfficinetNetbB0 Base, GAP to global average pooling, and BN to batch normalization layer.

Models Layers	Model1	Model2	Model3
Efficinetnetb0 Base	EB	EB	EB
Global Average Pooling	GAP	GAP	GAP
Dense Layer	-	256 nodes	256 nodes
Batch Normalization	-	BN	BN
Dropout	-	0.5	0.5
output Layer	2 nodes	2 nodes	2 nodes
Training technique and details			
batch size	64	64	64
epochs	30	30	30
optimizer	Adam	Adam	Adam
learning rate	0.01	0.01	schedule learning rate

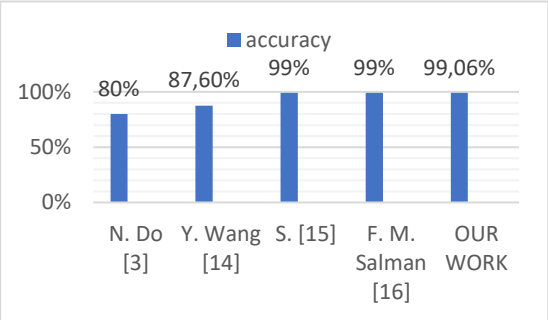
4 Results and discussion

After training models on the 140k real and fake faces dataset by using subsets on data that include: 100000 images as a training set, and 20000 images as a validation set, and then examining the three models using spirited 20000 images as a testing set, various accuracies are illustrated in Table 3.

The varying accuracy that was attained utilizing various model modifications and methodologies was depicted. We observed that the proposed method with improved EfficientNetB0 by adding a dense layer with 256 nodes and dropout techniques in model1 outperformed the base Effitiantnetb0 model.

By using the schedule learning rate technique, the Effitiantnetb0 model with the schedule learning technique in model 3 worked well with our dataset and gave us the highest accuracy score, 99.06% in the test set and 100% in the training set with a lost error rate, while the base EfficieneNetB0 in model 1 achieved only 51% in the same testing set and 76.88 in the training. The compression of our work against published research that used different techniques [3], [14], [15], [16] previously summarized in Table 1 can be shown in Figure 5.

Figure 5: Compression of our outcome with previous works.



5 Conclusion

Scaling the ConvNet is important to balance the network's depth, width, and resolution to increase the model's efficiency and accuracy. EfficientNetB0 provides this feature with fewer parameters. The proposed method aims to enhance the quality of fake image detection using EfficitNetB0 by modifying the model with an additional dense layer and using a learning rate schedule technique.

After several experiments, the results of model 3 concluded that the EfficientNetB0 model with additional dense layer and dropout technique in addition to using schedule learning technique work will in the detection task and our approach aids in the faster and more accurate detection of fake images and achieves good results.

Future studies will focus on refining a different CNN algorithm's performance for improved detection.

Future research could also apply this model to another dataset and check its performance.

Table 3: Models performance on the dataset.

Models	Training Set		Validation Set		Testing Set	
	ACC%	LOSS	ACC%	LOSS	ACC%	LOSS
Model1	76.88	0.76	66.3	0.77	51.88	1.1244
Model2	85.40	0.4195	80.66	0.89	65.88	0.44
Model3	100	1.4650e-05	99.37	0.0378	99.06	0.0569

References

[1] R. S. Khudeyer and N. M. Almoosawi, "Combination of machine learning algorithms and Resnet50 for Arabic Handwritten Classification," *Informatica*, vol. 46, no. 9, pp. 39–44, 2023, doi: 10.31449/inf.v46i9.4375.

[2] J. A. Alhijaj and R. S. Khudeyer, "Integration of EfficientNetB0 and Machine Learning for Fingerprint Classification," vol. 47, pp. 49–56, 2023.

[3] N. Do, I. Na, and S. Kim, "Forensics Face Detection From GANs Using Convolutional Neural Network," no. August 2018.

[4] V. Rachapudi and G. Lavanya Devi, "Improved convolutional neural network based histopathological image classification," *Evol. Intell.*, no. 0123456789, 2020, doi: 10.1007/s12065-020-00367-y.

[5] T. Carneiro, R. V. M. Da Nobrega, T. Nepomuceno, G. Bin Bian, V. H. C. De Albuquerque, and P. P. R. Filho, "Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications," *IEEE Access*, vol. 6, pp. 61677–61685, 2018, doi: 10.1109/ACCESS.2018.2874767.

[6] B. Bayar and M. C. Stamm, "A deep learning approach to universal image manipulation detection using a new convolutional layer," *IH MMSec 2016 - Proc. 2016 ACM Inf. Hiding Multimed. Secur. Work.*, pp. 5–10, 2016, doi: 10.1145/2909827.2930786.

[7] R. S. Khudeyer and N. M. Almoosawi, "Combination of machine learning algorithms and Resnet50 for Arabic Handwritten Classification," *Informatica*, vol. 46, no. 9, pp. 39–44, 2023, doi: 10.31449/inf.v46i9.4375.

[8] E. Alpaydin, "Neural Networks and Deep Learning," *Mach. Learn.*, 2021, doi: 10.7551/mitpress/13811.003.0007.

[9] M. S. Rana, M. N. Nobi, B. Murali, and A. H. Sung, "Deepfake Detection: A Systematic Literature Review," *IEEE Access*, vol. 10, pp. 25494–25513, 2022, doi: 10.1109/ACCESS.2022.3154404.

[10] M. Louis, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks Mingxing," *Can. J. Emerg. Med.*, vol. 15, no. 3, p. 190, 2013, doi: 10.2310/8000.2013.131108.

[11] H. C. Shin *et al.*, "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016, doi: 10.1109/TMI.2016.2528162.

[12] N. M. Al-Moosawi and R. S. Khudeyer, "ResNet-34/DR: A Residual Convolutional Neural Network for the Diagnosis of Diabetic Retinopathy," *Inform.*, vol. 45, no. 7, pp. 115–124, 2021, doi: 10.31449/inf.v45i7.3774.

[13] W. L. Alyoubi, W. M. Shalash, and M. F. Abulkhair, "Diabetic retinopathy detection through deep learning techniques: A review," *Informatics Med. Unlocked*, vol. 20, p. 100377, 2020, doi: 10.1016/j.imu.2020.100377.

[14] Y. Wang, V. Zarghami and S. Cui, "Fake Face Detection using Local Binary Pattern and Ensemble Modeling," 2021 IEEE International Conference on Image Processing (ICIP), Anchorage, AK, USA, 2021, pp. 3917-3921, doi: 10.1109/ICIP42928.2021.9506460.

[15] S., K. and V., M. (2022) 'Image quality assessment based fake face detection', Multimedia Tools and Applications, 82(6), pp. 8691–8708. doi:10.1007/s11042-021-11493-9.

[16] F. M. Salman and S. S. Abu-Naser, "Classification of Real and Fake Human Faces Using Deep Learning," Int. J. Acad. Eng. Res., vol. 6, no. 3, pp. 1–14, 2022, [Online]. Available: https://philpapers.org/rec/SALCOR-3.

[17] X. Chen *et al.*, "Application of EfficientNet-B0 and GRU-based deep learning on classifying the colposcopy diagnosis of precancerous cervical lesions," *Cancer Med.*, no. July 2022, pp. 1–10, 2023, doi: 10.1002/cam4.5581.

[18] R. Gupta, K. K. Bhardwaj, and D. K. Sharma, "Transfer Learning," *Mach. Learn. Big Data Concepts, Algorithms, Tools Appl.*, pp. 337–360, 2020, doi: 10.1002/9781119654834.ch13.

[19] V. L. Helen Josephine, A. P. Nirmala, and V. L. Alluri, "Impact of Hidden Dense Layers in Convolutional Neural Network to enhance Performance of Classification Model," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1131, no. 1, p. 012007, 2021, doi: 10.1088/1757-899x/1131/1/012007.