Araştırma Makalesi
https://doi.org/10.35234/fumbd.1219227

Classification of Real and Fake Face Data Using Capsule Networks

Ayşe ÇOBAN[1*], Fatih ÖZYURT[2]

[1] Department of Software Engineering, Graduate School of Natural and Applied Sciences, Fırat University, Elazig, Turkey.
[2] Department of Software Engineering, Faculty of Engineering, Firat University, Elazig, Turkey.
[*1] aysecoban9603@gmail.com, [2] ozyurtfatih@gmail.com

(Geliş/Received: 14/12/2022; Kabul/Accepted: 01/04/2023)

Abstract: Recently, with the advancement of technology, artificial intelligence has begun to be used in many areas. It is used in many fields, such as artificial intelligence, image processing, natural language processing, and recommended systems. The increase in the use of artificial intelligence has revealed the need for data, which has led to the production of new data from existing data. Today, generative adversarial networks (GAN) synthesize a wide variety of data inspired by existing data. These fake data produced from real image data can sometimes cause undesirable situations. It is essential to know whether the images, crucial for security, are fake or not. In this study, the classification of real human face data and fake face data generated from these data has been made. Fake face data were generated with the help of StyleGAN2-ADA from a small-sized dataset created by collecting the facial data of a famous person. It is aimed to classify the generated fake face data and real face data with the capsule network model.

Keywords: Capsule Networks, StyleGAN, Deep Learning.

Kapsül Ağları Kullanılarak Gerçek ve Sahte Yüz Verilerinin Sınıflandırılması

Öz: Son zamanlarda teknolojinin ilerlemesiyle birlikte yapay zekâ birçok alanda kullanılmaya başlandı. Yapay zekâ, görüntü işleme, doğal dil işleme, öneri sistemleri gibi birçok alanda kullanılmaktadır. Yapay zekâ kullanımının artması, veriye olan ihtiyacı ortaya çıkarmıştır. Bu durum mevcut verilerden yeni verilerin üretilmesine yol açmıştır. Günümüzde, çekişmeli üretici ağlar (GAN), mevcut verilerden ilham alarak çok çeşitli verileri sentezler. Gerçek görüntü verilerinden üretilen bu sahte veriler bazen istenmeyen durumlara neden olabilmektedir. Özellikle güvenlik açısından önem arz eden görüntülerin sahte olup olmadığının bilinmesi önemli bir konudur. Bu çalışmada, gerçek insan yüzü verilerinin ve bu verilerden üretilen sahte yüz verilerinin sınıflandırılması yapılmıştır. Ünlü bir kişinin yüz verileri toplanarak oluşturulan küçük boyutlu bir veri setinden StyleGAN2-ADA yardımıyla sahte yüz verileri üretilmiştir. Kapsül ağ modeli ile üretilen sahte yüz verileri ile gerçek yüz verilerinin sınıflandırılması amaçlanmıştır.

Anahtar kelimeler: Kapsül Ağlar, StyleGAN, Derin Öğrenme.

1. Introduction

Facial data classification is one of the deep learning problems studied extensively. Recently, with the popularization of GANs, fake data has started to be generated from real data. GANs consist of two structures: Generator and Discriminator. The Generator is the part that produces fake image data from related image data. On the other hand, Discriminator distinguishes the fake images generated by the Generator from the real images [1]. Thus, it is aimed that the image data generated is similar to the real data.

Obtaining fake data from real data can be used for malicious purposes. Fake data, such as face and signature images, essential for security, can cause problems in many areas. From this point of view, it is crucial to learn whether the image data is fake or not.

Convolutional neural networks are one of the most preferred methods in image data classification processes. Convolutional neural networks generally consist of convolution, pooling, and fully connected layers. The structure of convolutional neural networks is shown in Figure 1. The reason why convolutional neural networks are preferred is; They are easy to implement, and these networks perform classification with high performance.

[*] Sorumlu yazar: aysecoban9603@gmail.com. Yazarların ORCID Numarası: [1] 0000-0002-9922-8616, [2] 0000-0002-8154-6691.

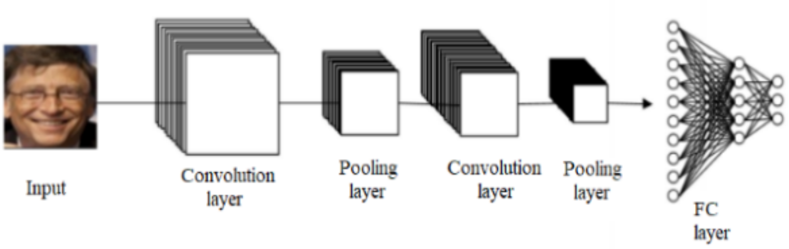

Figure 1. Convolutional neural networks layer

However, convolutional neural networks may only be able to perform somewhat efficient classification in some cases. The pooling layer in the structure of convolutional neural networks is a layer that provides the desired attributes from the data. This layer can cause information loss in the data [2]. In addition, convolutional neural networks may not be able to make successful classifications in datasets with insufficient samples with different angles, poses, and states [3]. For each sample, the need for a sufficient number of similar samples affects the performance of convolutional neural networks. Capsule networks are one of the preferred methods to solve these problems. Capsule networks receive data as vectors, not scalars. The vector parameters contain information about the state of the components in the data. Therefore, capsular networks; are the networks that perform the training process by considering the features of the components in the data, such as position, angle, and orientation relative to each other. Capsule networks learn more information from data than convolutional neural networks. This shows that capsule networks can also classify small data sets successfully [4].

In recent years, many studies have been carried out to solve such problems. One of these studies was done using a 9000-image real-fake human face dataset from the Kaggle depository. The study was conducted to classify the data in question with VGG16, ResNet50, MobileNet, InceptionV3, and a model they recommend and compare the obtained performance. As a result of the study concluded the proposed model with 95%, VGG19 93%, ResNet50 99%, MobileNet 98%, and InceptionV3 99% test accuracy[5].

Another study proposed an approach called FakeSpotter[6]. This proposed approach incorporates the Neuron coverage technique. In this study, the authors obtained fake face images from the CelebA-HQ and FFHQ datasets using InterFaceGAN [7] and StyleGAN [8] models. 12,000 images, 10,000 train, and 2,000 test images were used in the study. In this research, the FakeSpotter approach achieved fake face detection accuracy of 78.23%, 80.54%, and 84.78% in VGG-Face, OpenFace, and FaceNet, respectively.

Another study is based on data from the UW dataset[9], which consists of Google Earth satellite images and fake images produced with CycleGAN [10]. They used ResNet50, InceptionV3, Xception, and VGG16 models in the study. As a result, they achieved the best result among these models with the ResNet50 model.

Studies on this subject continue to be carried out. This study includes the classification of real face image data and fake face image data on a small data set. The study aims to investigate how the capsule networks will obtain results on this dataset.

2. StyleGAN

The general structure of generative adversarial networks consists of Generator and Discriminator. The Generator first performs a generation operation with latent space, and the Discriminator distinguishes whether the generated data is real. Gradients are calculated with backpropagation in each epoch, and the difference between the real and generated images is found [11]. According to the error, it is aimed that the Generator generates data more similar to the real data. This structure is shown in Figure 2.

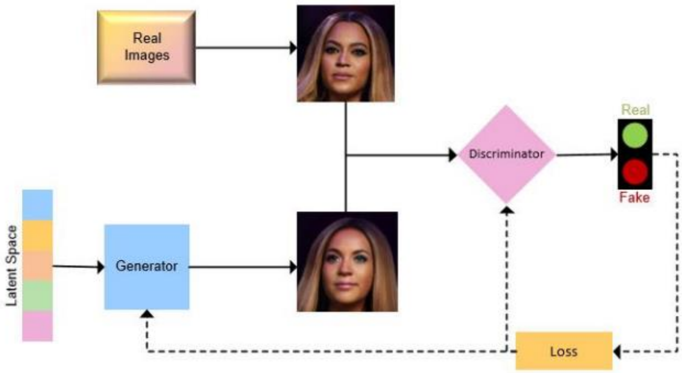

Figure 2. GAN structure

The primary purpose of GANs is to generate fake data very similar to real data. There are many types of GANs used in this sense. One advanced network that generates high-resolution fake image data from the given high-resolution real image data is StyleGAN [8]. It needs to be more manageable how the latent space evolves in traditional GANs. StyleGAN has included a nonlinear mapping network in this latent space. Thus, the generation process can be concluded more successfully. StyleGAN generates the fake image data incrementally, starting from low resolution to high resolution. Another feature is that it combines image styles using adaptive instance normalization (AdaIN) at each layer [12]. Adaptive Instance Normalization causes blobs (Blob artifacts) resembling water droplets to form in the generated images. In the proposed StyleGAN2, adaptive instance normalization is restructured as weighted demodulation [13]. On the other hand, it has been observed that the approach of generating high-resolution data leads to feature degradation in images. In StyleGAN2, the Generator and Discriminator are connected using multi-hop connections at each resolution level [12].

StyleGAN2-ADA is very similar to StyleGAN2 in structure. However, it applies the Adaptive Discriminator Augmentation approach to the inputs. This technique considerably reduces the amount of data needed for training [13], and this enables successful data synthesis with small datasets.

3. Capsule Networks

The first layer of capsule networks is a standard convolution layer. After the convolution layer comes the second layer called the primary capsule layer. The data is resized and passed through the activation function in the primary capsule layer. At the end of these processes, vectors called DigitCaps are obtained from the data as much as the number of classes. The layers of the capsule networks are given in Figure 3.

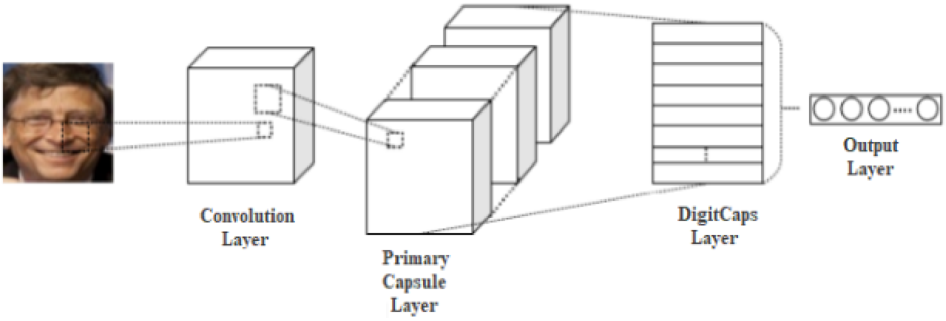

Figure 3. Capsule networks layer

Capsule networks, unlike convolutional neural networks, receive data in vector form. It contains a dynamic routing algorithm in its structure instead of a pooling layer, and the squash function is used as the activation function. A dynamic routing algorithm is an algorithm that matches lower-level capsules with similar inputs to higher-level capsules. It is stated that the dynamic routing algorithm will be more efficient than the pooling layer [14].

In areas where the desired object component is present in the image, the vector size is small in areas where it is not large [3]. The squash function ensures that vector lengths take a value between zero and one, depending on the size of the vectors. The vectors here represent the probability of the received data component being in the target class and its state, namely its parameters [15]. The direction of the vector in question represents the state of the component. (angle, direction, thickness, etc.).

4. Method And Material

In this study, the face data classification process was carried out with Tesla K80 hardware in a Google Collaboratory working environment using a capsule networks model. The data used in this study were created by collecting facial images of a celebrity person from the internet. The dataset consists of 200 face image data, and 200 new data was synthesized from the aforementioned data set using StyleGAN2-ADA. Thus, a dataset of 200 real and 200 fake face images was created to perform the classification study. The data has been reduced to 64x64 dimensions to complete the classification process with less cost using capsule networks. An example of the prepared data set is shown in Figure 4. 60% of the prepared data set was used for training, 20% for validation, and 20% for testing.

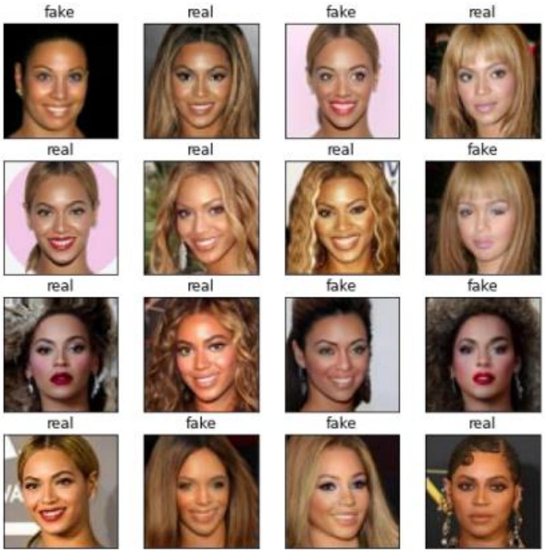

Figure 4. Sample of dataset

5. Results

The results of the capsule networks model from the experiment on the data are shown in Table 1. As a result of the classification study, the capsule networks model obtained 87.9% training accuracy, 86.2% validation accuracy, and 85% test accuracy. The training and validation graph of the model is given in Figure 5.

Table I. Capsule networks model results

Capsule networks model results	
Training	87.9%
Validation	86.2%
Test	85%

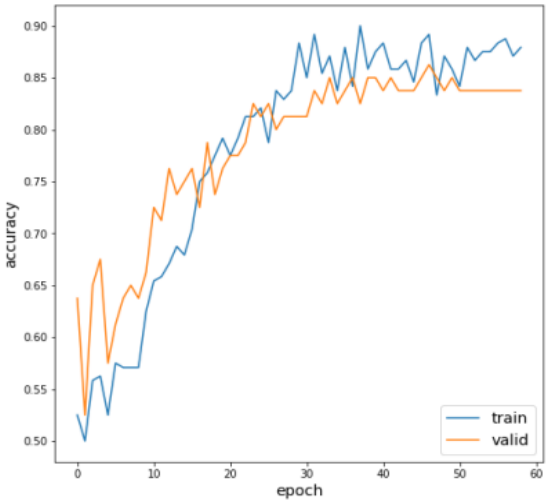

Figure 5. Graph of the training and validation process of the model

No significant change was observed between the training and validation accuracy results. Since the experiment's validation accuracy result remained constant after a while, the study resulted in 59 epochs. The confusion matrix of the model is given in Figure 6.

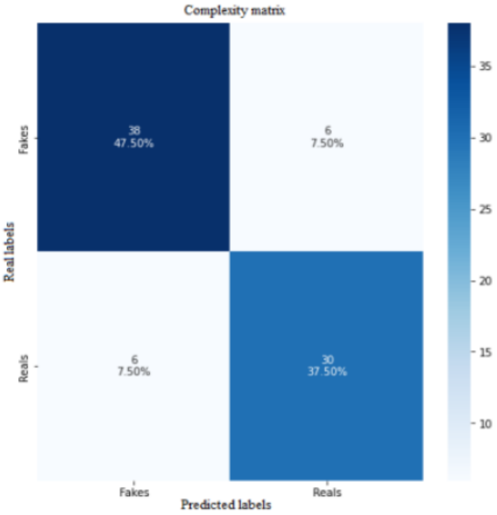

Figure 6. The confusion matrix of the model

An example of the capsule networks model to predict the classes of some of the face data used is as in Figure 7.

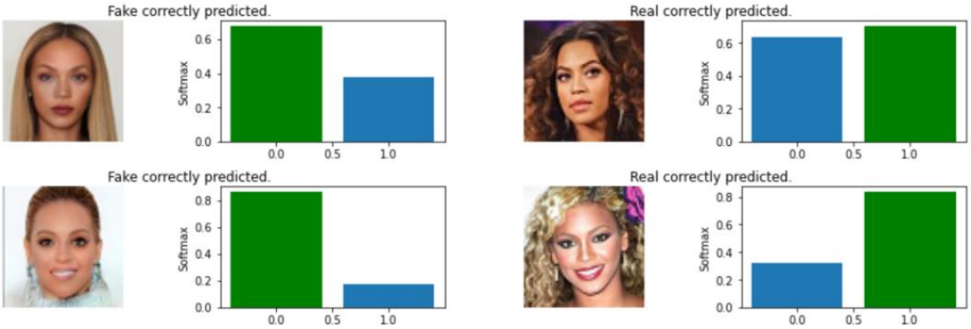

Figure 7. Prediction example of the model

Capsule networks successfully performed the classification process, and the performance is expected to increase if the dataset is prepared larger in size. The disadvantage encountered in the classification is that the capsule network model performs the training process for a long time. The training process, which takes about 20 minutes, is likely to be costly in terms of time compared to other networks.

6. Conclusion and Recommendations

Today, GANs produce similar data sets inspired by image data. These fake data produced from real image data are used in many areas where artificial intelligence is used. The dataset used in the study consisted of facial images of a famous person and fake images produced with StyleGAN2-ADA. This data is a relatively small dataset consisting of 200 real and 200 fake images. On this dataset, the capsule networks reached 85% test accuracy. Capsule networks have performed a classification process that can be considered successful with the dataset it works with. It is estimated that the study has a high probability of being examined in the future. The fake face dataset obtained using real face data shows that the proposed framework is effectively different from the real face data. In addition, the comparisons between the proposed and other methods show that the proposed model is sufficient in fake detection capacity.

In the future, data can be generated and classified with different GANs. Different models can be used for the classification process. The models used can be developed, and performance comparisons can be made. The results can be examined by increasing or decreasing the dataset size. With different data types, fake and real classification processes can be performed and the results can be examined.

References

[1] Bahar MS, Buluş E. Derin Öğrenme Teknikleri Kullanılarak Sahte Yüz Fotoğrafı ve Videosu Sentezi. Düzce Üniversitesi Bilim ve Teknoloji Dergisi 2021; 9(6): 354-369. DOI: 10.29130/dubited.1017584.

[2] Sabour S, Frosst N, Hinton GE. Dynamic Routing Between Capsules. arXiv preprint 2017; arXiv:1710.09829.

[3] Kizrak MA, Beser F, Bolat B, Yildirim T. Kapsül Ağları ile İşaret Dili Tanıma Recognition of Sign Language using Capsule Networks. 26th Signal Processing and Communications Applications Conference (SIU) 2018; 1-4. doi:10.1109/SIU.2018.8404385.

[4] Osman AA, Face Identification Using Capsule Network with Small Data Set. Master Thesis, Tallin University of Technology, 2020.

[5] Salman, FM, Abu-Naser, SS. Classification of Real and Fake Human Faces Using Deep Learning. International Journal of Academic Engineering Research (IJAER) (2022); 6 (3):1-14.

[6] Wan R, Ma L, Juefei-Xu F, Xie X, Wang J, Liu Y. FakeSpotter: A Simple Baseline for Spotting AI-Synthesized Fake Faces. 2019; arXiv:1909.06122.

[7] Shen Y, Gu J, Tang X, Zhou B, Interpreting the Latent Space of GANs for Semantic Face Editing. in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition 2020.

[8] Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2019; 4401–4410.

[9] Zhao B, Zhang S, Xu C, Sun Y, Deng C. Deep fake geography? when geospatial data encounter artificial intelligence. Cartography and Geographic Information Science 2021; 48(4): 338–352.

[10] Fezza SA, Ouis M, Bachir K, Hamidouche W, Hadid A. Evaluation of Pre-Trained CNN Models for Geographic Fake Image Detection. arXiv preprint 2022; arXiv:2210.00361.

[11] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. In Advances in neural information processing systems 2014; 2672–2680.

[12] Situ Z, Teng S, Liu H, Luo J, Zhou Q. Automated Sewer Defects Detection Using Style-Based Generative Adversarial Networks and Fine-Tuned Well-Known CNN Classifier. in IEEE Access 2021; 9: 59498-59507. doi: 10.1109/ACCESS.2021.3073915.

[13] Karras T, Aittala M, Hellsten J, Laine S, Lehtinen J, Aila T. Training generative adversarial networks with limited data. Proc. Adv. Neural Inf. Process. Syst. 2020; 33: 1-15.

[14] Kınlı F, Kıraç F. FashionCapsNet: Clothing Classification with Capsule Networks. Bilişim Teknolojileri Dergisi 2020; 13(1): 87-96. doi:10.17671/gazibtd.580222.

[15] Çoban A, Özyurt F. Kapsül Ağları ile Yüz Verilerinin Sınıflandırılması, Avrupa Bilim ve Teknoloji Dergisi 2022; 33: 176-183. doi:10.31590/ejosat.999055.