PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Cat face recognition using Siamese network

Haohuan Li, Wenqing Zhang

Haohuan Li, Wenqing Zhang, "Cat face recognition using Siamese network," Proc. SPIE 12456, International Conference on Artificial Intelligence and Intelligent Information Processing (AIIIP 2022), 124562U (30 November 2022); doi: 10.1117/12.2659645

Event: International Conference on Artificial Intelligence and Intelligent Information Processing (AIIIP 2022), 2022, Qingdao, China

Cat Face Recognition using Siamese Network

Haohuan Li[1,a†], Wenqing Zhang[2*, †]
[1]Computer Science University of St. Andrews St. Andrews, Scotland
[2]Computer Science University of Sheffield Sheffield, England
[a]hl74@st-andrews.ac.uk
[*]wzhang110@sheffield.ac.uk
[†]These authors are equally contributed.

Abstract

As a challenging problem in the field of image analysis and computer vision, face recognition has garnered a lot of attention in recent years due to its many applications across a wide range of areas. In recent years, face recognition technology has been focused on humans, but considering the expansion and improvements of the cat industry, face recognition of cats should also receive some attention. This paper proposes a face recognition system for cats using Siamese and VGG16. We will send two images of cat faces to the Siamese network, which will be converted to a vector by mapping their features to the space and then get their probability of similarity by calculating their losses. Our training model on 13,106 cat images from a dataset of 509 different cats shows that our method can recognize cats' faces with an accuracy of 72.91% on a test dataset with 1702 image pairs containing 851 pairs labelled true, and 851 pairs labelled false. Experiments have demonstrated that this method is convenient and contactless and has a high recognition rate.

Keywords-Cat face recognition; Siamese network; VGGNet

1. Introduction

Face recognition was naturally an essential area of research in the computer vision area. Over many years of continuous development, face recognition has now achieved a high level of accuracy. While researching human face recognition, an exciting idea suddenly occurred to us: could face recognition be used to identify the faces of cats? Cause more than a thousand pet cats are missing, which makes their owner extremely painful. The implantation of microchips has always been our preferred method of locating lost pets, but some health issues (e.g., inflammatory reaction and cancer) can arise. Therefore, the identification of missing pets requires a noninvasive method. There are quite a few works for the cat facial recognition area. Adam Klein [1] uses deep convolutional neural networks and the dataset collected from the Petfinder website to train the model and obtained an accuracy of around 95% on the verification task and a less impressive but still robust 81% on the rank-5 identification task using an open-set protocol with cats not seen during training. Lin and Kuo [2]train a CNN network to perform cat facial feature identification and have achieved an accuracy of 91%. However, whether the reported accuracy excludes individuals or images that have been seen is not reported in the paper. Yu Fan, Chih-Chung Yang and Chin-Ta Chen [3] use the MVCC and GMM to recognize the cat face with an accuracy between 70% and 83.3% with different orders of Gauss and MFCC, which the performance does not always monotonically increase with increment of the order. The pet industry continues to grow and improve as the pet economy grows gradually, with cats already being a crucial part of our families. There has never been a greater need for cat identification and cat profiles than in the management of urban cats, prevention of cat loss, and even in cat medical insurance. Siamese networks are often used to measure the similarity of two inputs [4]. Siamese is usually used in the face identification area and has achieved relatively high performance, which represents the robust competitive ability in the facial recognition area. The use of the Siamese Network should be able to get a better result in the cat facial recognition area considering its robustness in human facial recognition. We use the training dataset from Kaggle [5] and the test dataset from GitHub [6]. In the future, we aim to pre-process it by extracting cat faces in images and using those images extracted from the original image to train our model to achieve better performance. This paper provides the Siamese algorithm with the feature extraction algorithm VGG16 to recognize the different cat faces, thereby obtaining 72.91% precision.

International Conference on Artificial Intelligence and Intelligent Information Processing
(AIIIP 2022), edited by Pavel Loskot, Proc. of SPIE Vol. 12456, 124562U

2. Method

2.1VGGNet

An improvement of VGG16 compared to AlexNet is to use several consecutive 3x3 convolution kernels instead of larger convolution kernels (11x11, 7x7, 5x5) in AlexNet. For a given receptive field (the local size of the input image relative to the output), using stacked small convolution kernels is preferable to using large convolution kernels. The choice of the convolution kernel is led by multiple nonlinear layers that can increase the network depth to ensure a more complex learning mode. The cost is relatively small (fewer parameters). The VGG use three 3x3 convolution kernels instead of 7x7 convolution kernels, and two 3x3 convolution kernels are used instead of 5*5 convolution kernels [7]. The primary purpose of this is to ensure that the depth of the network is improved within the condition of the perception field, and the effect of the neural network is enhanced to a certain extent.

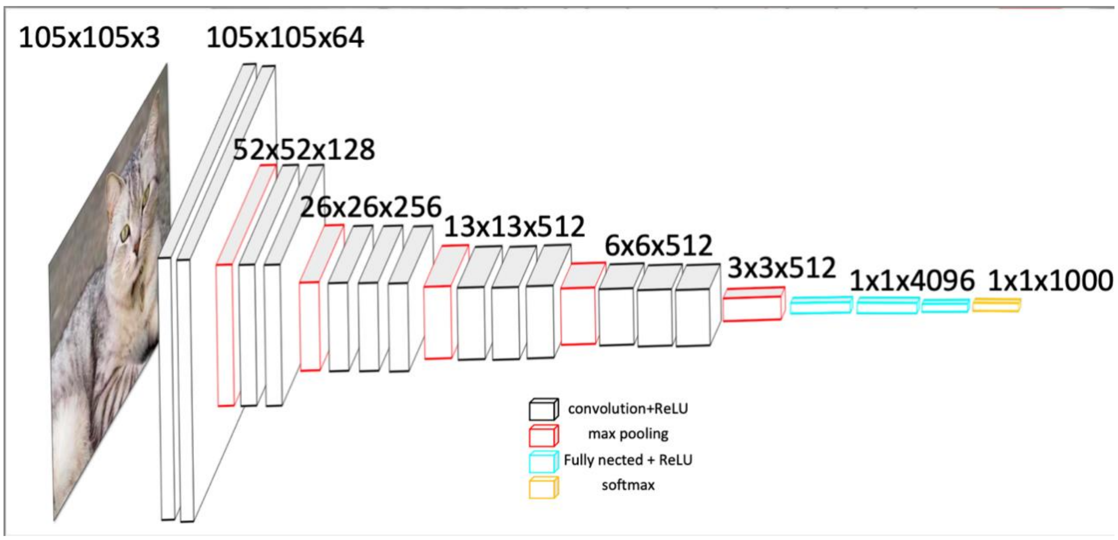

Figure 1. The structure of VGGNet

2.2Siamese Network

Siamese nets were first introduced in the early 1990s by Bromley and LeCun to solve signature verification as an image matching problem [8]. The Siamese neural network consists of twin networks, and the connection between the two networks is achieved by sharing weights. Unlike a model that learns to classify its inputs, this neural network learns to distinguish between the two inputs, which means that it retains the similarities between the two inputs. The function of the twins' networks is to perform feature extraction.

Network architecture. The function of the backbone feature extraction network of the Siamese neural network is to perform feature extraction, and various neural networks can be applied. Here we use the VGG16 neural network, which extracts the input image after resizing by following steps.

1) An original image is resized to the specified size. This article uses 105x105.

2) conv1 includes two [3, 3] convolutional networks, one 2X2 max pooling, and the output feature layer is 64 channels.

3) conv2 includes two [3, 3] convolutional networks, one 2X2 max pooling, and the output feature layer is 128 channels.

4) conv3 includes three [3, 3] convolutional networks, one 2X2 max pooling, and the output feature layer is 256 channels.

5) conv4 includes three [3, 3] convolutional networks, one 2X2 max pooling, and the output feature layer is 512 channels.

6) conv5 includes three [3, 3] convolutional networks, one 2X2 max pooling, and the output feature layer is 512 channels.

After obtaining the backbone feature extraction network, we can obtain a multi-dimensional feature, which could be converted to one dimension vector using flatten.

We subtract these two one-dimensional vectors and then sum their absolute values, which is equivalent to obtaining the L1 norm of the interpolation of the two eigenvectors, to find the distance between two one-dimensional vectors [9]. And then, we make two full connections for this distance so that these two are fully connected to a neuron whose result is sigmoid, making the distance between 0 to 1 represent the similarity of the two input images.

In a word, the Siamese network has two inputs, and these two inputs are passed into a neural network which maps the features to the space, forms a vector, and calculates the loss using uses the "distance" (L1 Norm) between the two vectors to represent the difference between the inputs (the gap in image semantics).

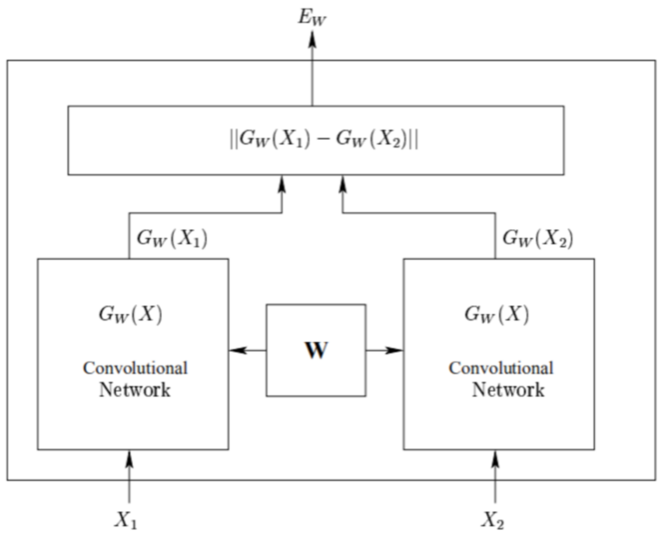

Figure 2. Structure of the proposed network. We introduce binary_crossentropy as our loss to optimize the network.

Loss. A Siamese neural network has two inputs that point to the lable1 if the type of image is the same and point to the lable0 if the kind of image is different. And then, we make the output of the network and the real label do the cross-entropy operation, binary_crossentropy, which can be used as the final loss [10].

$$\text{Loss} = -\frac{1}{\substack{\text{output}\\\text{size}}}\sum_{i=1}^{\substack{\text{output}\\\text{size}}} y_i \cdot \log \hat{y}_i + (1 - y_i) \cdot \log(1 - \hat{y}_i) \tag{1}$$

Implementation details. Adam's name comes from adaptive moment estimation. The Adam method is a combination of the strengths of two recently popular methods: AdaGrad [11], which works well on sparse gradients, and RMSProp [12], which works well on non-stationary and online problems have excellent performance.

And Adam is an efficient stochastic optimization method requiring only a first-order gradient and a small amount of memory [13]. The method calculates the adaptive learning rate of different parameters by estimating the first and second gradients.

Assuming that dW and db are calculated with each mini-batch, at the t-th iteration

$$v_{dW} = \beta_1 v_{dW} + (1 - \beta_1)dW \tag{2}$$

$$v_{db} = \beta_1 v_{db} + (1 - \beta_1)db \tag{3}$$

$$v_{dW}^{corrected}[\ell] = \frac{v_{dW}[l]}{1 - (\beta 1)^t} \tag{4}$$

$$sdW = \beta_2 sdW + (1 - \beta_2)(dW)^2 \tag{5}$$

$$s_{db} = \beta_2 s_{db} + (1 - \beta_2)(db)^2 \tag{6}$$

$$s_{dW}^{corrected}[\ell] = \frac{sdW[l]}{1 - (\beta 2)^t} \tag{7}$$

Parameter update of the Adam algorithm:

$$W := W - \alpha \frac{v_{dW}^{\text{corrected}}}{\sqrt{s_{dW}^{\text{corrected}}} + \epsilon} \tag{8}$$

$$b := b - \alpha \frac{v_{db}^{\text{corrected}}}{\sqrt{s_{db}^{\text{corrected}}} + \epsilon} \tag{9}$$

And we choose our optimizer to be Adam, with a learning rate of 1e-5.

3. Results and Discussion

3.1Dataset

Our training dataset uses the "cat individual snn" dataset from Kaggle, which includes a set of images with corresponding labels where cats with the same label represent the fact that two images belong to the same cat. The dataset consists of 13106 images in total that belong to 509 different cats. For any cat individual, the most miniature images a cat owns are 6, and the most images a cat owns are 170. In all the pictures, the cat appears in the front or side view (e.g., Figure 3). We note that our test set contains multiple cat samples. In training and testing, the samples are randomly selected as a training and test pairs to let the method recognize whether two samples belong to the same cat.

Our test dataset includes 105 images from 7 different cats; each cat owns a number of images in a range from 5 to 24. In terms of data pre-processing, we compare images in datasets two by two and classify output labels by comparing whether their labels are the same. If two animals are of the same kind, the label is to be set to true, and vice versa. After connecting two pairs in this way, there were a total of 5460 image pairs in the test set, with 851 pairs labelled True and 4609 pairs labelled False. We can find out that there happens to have an extreme unbalance in the test set. To make the test data more precise, we reduce False samples in the test set and make the total number of test datasets to 1702 pairs of images, including 851 pairs labelled True, and 851 pairs labelled False. We achieve 82.5% accuracy using the proposed Siamese network.

Figure 3. Sample of the training set.

3.2Result Evaluation and Discussion

For 1702 pairs of cat images that include 50% pairs of images that belong to the same cat and 50% pairs of images that belong to different cats, we set the threshold of the model verification to 0.5, which means that when the possibility is larger than 0.5, the model will judge two images belonging to the same cat. The parameter leading to the accuracy of the cat identification reaches 72.91%. We plotted the relationship between Epoch and Accuracy rate, validation loss and the model loss, as is shown in Figure 1.

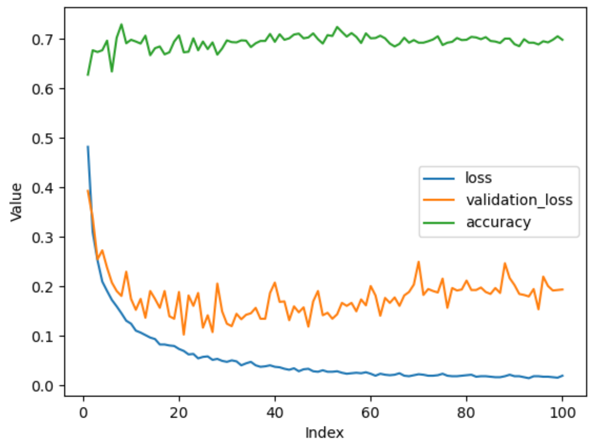

Figure 4. Loss, Validation Loss and Accuracy

3.3Future Work

In this model, we believe that the main reason for the low accuracy of the difference between the two images is the presence of a background in the original dataset, which is considered in the training process of the model. In our future work, features of the cat's face from the data in the dataset will be extracted and then trained in the image using the Siamese network.

4. Conclusion

This paper combines the Siamese algorithm with the feature extraction algorithm VGG16 to analyze different cat faces and achieve 72.91% accuracy. In the experiment, two cat face images as the inputs are passed to the Siamese network, which processes them into a neural network mapping the features to space and creating a vector. And then, we have used this vector as an input to calculate the loss between the two cat face images represented by the two vectors to describe the gap in image semantics. Our final step is to convert the difference between two input vectors of the cat face image to the probability of whether these two images were taken from the same cat by using the sigmoid function. Based on our training model on 13,106 cat images from a dataset of 509 different cats, our method can recognize cats' faces with an accuracy of 72.91% based on 1702 pairs of images, which includes 851 pairs labelled True and 851 pairs labelled False. By using our method, cat identification and cat profiles could be primarily applied to the management of urban cats, the prevention of cat loss, and even the provision of cat medical insurance as the pet economy grows gradually.

References

[1] A. Klein, "Pet Cat Face Verification and Identification," 2019.
[2] T.-Y. Lin and Y.-F. Kuo, "Cat face recognition using deep learning," in 2018 ASABE Annual International Meeting, Michigan, 2018.
[3] Y. Fan, C.-C. Yang and C.-T. Chen, "Cat Face Recognition Based on MFCC and GMM," 2021 6th International Conference on Image, Vision and Computing (ICIVC), pp. 81-84, 2021.
[4] D. Chicco, "Siamese Neural Networks: An Overview," in Artificial Neural Networks, New York, NY, Springer US, 2021, pp. 73--94.
[5] L. Jovi, "cat individual_snn," 8 2021. [Online]. Available: https://www.kaggle.com/code/jovi1018/cat-individual-snn/notebook. [Accessed 2022 4 3].
[6] J. Jiang, "Cat Face Recognition," [Online]. Available: https://github.com/El-Chiang/CatFace. [Accessed 01 04 2022].
[7] H. Qassim, A. Verma and D. Feinzimer, "Compressed residual-VGG16 CNN model for big data places image recognition," 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), pp. 169-175, 2018.

[8] J. Bromley, I. Guyon, Y. LeCun, S. Eduard and R. Shah, “Signature verification using a” siamese” time delay neural network,” Advances in neural information processing systems, p. 6, 1993.

[9] Y. Qiaolin, J. Yang, F. Liu, C. Zhao, N. Ye and Y. Tongming, “L1-Norm Distance Linear Discriminant Analysis Based on an Effective Iterative Algorithm,” IEEE Transactions on Circuits and Systems for Video Technology, pp. 114-129, 2018.

[10] Z. Zhang and M. Sabuncu, “Generalized cross entropy loss for training deep neural networks with noisy labels,” Advances in neural information processing systems, p. 31, 2018.

[11] A. Lydia and F. Sagayaraj, “Adagrad—an optimizer for stochastic gradient descent,” Int. J. Inf. Comput. Sci, pp. 566-568, 2019.

[12] T. Tijmen and H. Geoffrey, "RMSProp," COURSERA: Lecture, p. 7017, 2012.

[13] K. Diederik P and B. Jimmy, “Adam: A method for stochastic optimization “.arXiv preprint arXiv:1412.6980.