Improving neural network performance

by adding direct-connected structures

for image manipulation Classification

Yuchi Liu

Research School of Computer Science,
Australian National University, Canberra, Australia
{u6920312@anu.edu.au}

Abstract. Neural networks (NN) have been used for decades for classification tasks, but their application on constantly updated massive data is still a challenge. With the development of deep learning techniques, attempts have been made to increase the depth of neural networks to improve the accuracy of the models, but deep neural networks (DNN) perform worse when more layers are added. Therefore, the residual network (ResNet) was born. In this paper, we add the ResNet direct-connection structure to a normal deep network to verify the effectiveness of ResNet on a general neural network. We compare the performance of a general deep neural network (a network with three hidden layers), a deep network with the addition of the ResNet direct-connection structure, and the constructcascade neural network (ConstrCasc) proposed in [1] to identify whether a graph has been manipulated or not. These models were trained with datasets containing the number of participants, the number of images, the number of times participants gazed at the whole image/target area of the image and the time. These data were collected using eye tracking techniques [2]. During our experiments we verified that the ResNet direct-connected structure can improve the accuracy of the deep neural network model. First, the deep network model with the addition of the ResNet direct-connect structure achieved an overall accuracy of 85.54% on prediction, which is about 5% better than the normal deep neural network model, about 17% better than the construct casc architecture, and performs better than the results in the manipulated image perception paper [2].

Keywords: DNN, ResNet, cascade neural network, image manipulation detection, classification, performance improvement

1. Introduction

Showing only the image itself relying on the ability of both eyes to determine if the image has been manipulated has been shown to be grossly inadequate in the existing literature because our judgments are based on our logic and people have a better ability to recognize images when they show physically untrustworthy changes, while it is difficult to recognize images when only physically trustworthy changes have occurred [4], which will have an impact on the future of multiple This will have an impact on the implications for the future design of multiple forms of human-computer interaction through the Web [2]. Therefore, assessing whether an image is manipulated using existing techniques has been investigated in many areas. Existing measures include eye tracking and the use of grounded theory analysis to analyze respondents' verbal responses [2].

There are many approaches that can handle the analysis of image manipulation detection data. Among them, neural networks undoubtedly excel in classification tasks. In this paper, we first used a DNN model for prediction, but the degradation problem (Degradation problem) appeared in the deep network when trying to increase the accuracy of the model by adding more hidden layers: the network accuracy saturated and even decreased when the depth of the network increased.

We refer to the Resnet model proposed in [7] and try to add the direct link structure of Resnet to the DNN model to improve the accuracy of the model. Finally, the DNN model with the direct connection structure of Resnet added, the normal DNN model and the network model using construccasc [1] technique are compared in terms of performance.

2.Method

2.1 Neural Network Topology

Deep Neural Network

In this paper, a deep neural network model (with three hidden layers) is used. The input layer has a total of six neurons that capture the six features of its input. The output layer has a total of two output neurons holding two labels (image manipulated or not manipulated). Each layer of the network is completely linearly connected. After comparison, the number of nodes in the hidden layer was chosen to be 52,32,24, respectively, to maintain a certain level of applicability while not easily causing overfitting of the data. The final most likely labels were calculated by sigmoid as predictions.

Deep Neural Network with direct-connected structure

Often deeper networks outperform their shallower counterparts: Suppose we have built an n-layer network and achieved a certain level of accuracy. Then an n+1 layer network should be able to achieve at least the same level of accuracy - simply by replicating the previous n layers and adding a layer of constant mapping to the last layer. In practice, however, these deeper networks will largely perform worse.

Dr. Kaiming He, author of ResNet, attributes these problems to a single assumption: constant mappings are hard to learn. And they propose a correction: instead of learning the basic mapping relationship from x to H(x), they learn the difference between the two, also known as the "residual" [7]. Then, in order to compute H(x), we simply add this residual to the input. Suppose the residual is $F(x)=H(x)-x$, so now our network will not learn H(x) directly, but $F(x)+x$.

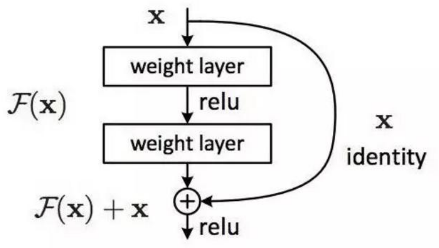

Figure 1. Residual Network Module

However, ResNet is implemented on convolutional neural networks, but image datasets are not used for training and testing in this paper. We refer to its idea of adding constant mapping and residual-based learning to improve the performance of deep neural networks by adding the ResNet direct-connected structure.

ConstrCasc

The ConstrCasc technique is derived from the paper[1], where the paper investigates the generalization performance and structural changes of cascaded neural networks in face recognition tasks [1]. In our constructed cascaded neural network, we intend to modify it to reduce the complexity of the network while making it more suitable for image manipulation datasets. A simple neural network (NN) is more suitable for this task than the convolutional neural network (CNN) used in [1]. the main advantage of the CNN is that it can capture important features (e.g., edges) in an image without human supervision [3]. However, the image manipulation dataset is composed of six features from each participant's observed image and is not an image vector [2]. Therefore, the constructed neural network in which one-dimensional layers will be used instead of two-dimensional layers.

The initial network consists of an input layer and an output layer with a total of 6 neurons in the input layer and 2 output neurons representing the labels. Also, the added hidden neurons are partially connected to the input neurons and fully connected to the output neurons according to the construction cascade algorithm. [3]

Since the dataset is fairly simple, the model only needs to add at most two cascade layers. In this report, two variants of the above architecture are tested: one cascade layer is added and finally two cascade layers are added.

2.2 Dataset

The image manipulation dataset was proposed by Sabrina Caldwell, Tamás Gedeon, Richard Jones, Leana Copeland [2] and includes the number of participants and images, the number of times and times participants observed the images as well as the target area, and the participants' opinions on whether the images were manipulated or not (transformed into numerical features, where 0 = not manipulated, 1 = manipulated, and 2 = not known). The data set was collected from 80 participants observing 5 images. During their observations, their observation area and observation time will be recorded by the eye tracking instrument. Before the data were fed into the neural network, the dataset was preprocessed as follows: 1) participants' opinions on whether the images were manipulated were discarded because it was considered too relevant to the labels we needed to predict. 2) The data were normalized. For each type of network, the dataset was randomly divided into a training set and a test set in a ratio of 8:2.

2.3 Training Methodology

Deep Neural Network

For each training, run 200 epochs. Output loss and accuracy after every 50 epochs of training. The learning algorithm used in the training process is Adam, and the loss of each epoch is calculated as cross-entropy loss. The Relu and Sigmoid function is used together as the activation function. The model was tested for different learning rates and was finally determined to be 0.01.

Deep Neural Network with direct-connected structure

For each training, run 200 epochs. Output loss and accuracy after every 50 epochs of training. The learning algorithm used in the training process is Adam, and the loss of each epoch is calculated as cross-entropy loss. The Relu function is used as the activation function. The model was tested for different learning rates and was finally determined to be 0.01.

ConstrCasc

Because of the lack of symmetric features in the dataset, the weight sharing strategy proposed in [1] was not applied.Resilient Propagation (RPROP) is used as the learning algorithm in all learning processes, reference to paper [1].The model was tested for different learning rates and was finally determined to be 0.01. Starting with no hidden layers and no cascading layers between inputs and outputs, each training is repeated for a maximum of 200 epochs, and the loss is calculated as the cross-entropy loss of each epoch. If the loss is lower than 0.6, the training process ends early and a new layer is added. Both variants were subjected to the above training process.

3 Results and Discussion

3.1 Results

We evaluated the dataset discussed in Section 2.2 using the model presented in Section 2.1. The problem to be solved is to classify whether an image is manipulated or not based on the number of times a person observes the image and the time of day. The experiments were run manually for 10 rounds. Table 1 contains the average performance of these models for comparison purposes.

Table 1. The average performance of each network over 10 runs

Architecture	DNN	DNN with direct-connected structure	1 cascade layer	2 cascade layers
Testing accuracy%	80.31%	**85.54%**	62.50%	68.72%

3.2 Discussion

As shown in Table 1 for the average performance of all four networks, the DNN model shows an improvement of about 10% over the ConstrCasc model. This may be because the size of the dataset limits the generalization of the ConstrCasc model. And DNN with direct-connected structure shows an improvement of about 5% over the DNN model. Compared to the accuracy of 56% using grounded theory analysis in the image manipulation paper [2], the average accuracy of the DNN with direct-connected structure model is 85.54%.

4 Conclusion

The main objective of this study was to see if the direct-connected structure of Resnet improves the ability of deep neural networks to discriminate whether an image is manipulated or not. The inputs to the study model included the number of participants and observed images, as well as the number of times and the time that participants observed the image/image target region. The test results showed that the deep neural network model with the addition of the direct-connect structure improved accuracy by approximately 5%. Given the current dramatic increase in the frequency of image use in society, from the Internet to the news industry, future research could focus on exploring more powerful models to improve recognition accuracy. Also, it is worthwhile to investigate in depth how to use convolutional networks to directly recognize images to determine whether they are manipulated or not.

References

1. Khoo, Suisin, and Tom Gedeon. n.d. "Generalisation Performance vs. Architecture Variations in Constructive Cascade Networks"8.(2008)
2. Caldwell, Sabrina, Tamás Gedeon, Richard Jones, and Leana Copeland. n.d. "Imperfect Understandings: A Grounded Theory And Eye Gaze Investigation Of Human Perceptions Of Manipulated And Unmanipulated Digital Images."9.(2015)
3. Cai, Tong. n.d. "AI as Mendacity Indicators: A Comparative Study on Neural Networks/ Deep Learning Models for Thermal Imaging-Based Deceive Classification." 5.(2020)
4. Nightingale, Sophie J. n.d. "Can People Identify Original and Manipulated Photos of Real-World Scenes?" 21.(2017)
5. Hall, Eric. n.d. "Detecting Depression with Various Neural Networks." 9.(2020)
6. Sadouk, L.n.d. "CNN Approaches for Time Series Classification. In Time Series Analysis-Data, Methods, and Applications."24.(2018)
7. K. He, X. Zhang, S. Ren, and J. Sun. n.d "Deep Residual Learning for Image Recognition," 11.(2015)
8. K. He, X. Zhang, S. Ren, and J. Sun. n.d "Identity Mappings in Deep Residual Networks," 7. (2016)