Using ResNet to Improve the Casper Network in Detecting Stress

Jiale Wang

Research School of Computer Science, Australian National University

Abstract.
Stress is one of the body responses to the pressure. If people are under anxiety for a long time, it will have negative impacts on human body. Thus, it is crucial to recognize the stress in an effective way. Because of the COVID-19 pandemic, more and more people prefer contact-less meeting. Students are learning remotely, companies are holding on-line meetings, etc. Therefore, building a good system to monitor people's stress can better understand students or employee's feeling. In this paper, the aim of the experiment is to figure out that if pretrained data can improve the Casper Network or not. Therefore, an improved version of the Casper Network was performed, that is using the finetuning and feature extraction first to detect the stress with pictures, then put the pre-trained data into the Casper Network. The results of the improved version end up with the best 83.60% accuracy which is much better than the 60.52% Residual Neural Network (ResNet) only and 51% normal Casper Network with 30 epochs. Regarding the results of the experiment, we can conclude that using Resnet to pre-process the data first can improve the performance of the Casper Network in Stress detection.

Keywords: Resnet, Casper Network, Stress, Calm

1 Introduction

Human stress is produced when people face difficulties especially from school, family and work etc. It is an unhealthy signal if we keep under the anxiety for a long time. Thus, create a system to monitor the stress is very important.

Traditional stress monitoring methods such as physical symptoms monitoring (Jennifer Casarella, 2019) and using distributed wireless intelligent sensor [4] are all highly rely on the contact with people. The limitation of these methods is that they highly rely on the people's introspection. The result of the self-reporting method may produce bias and uncertainties. What's more, because of the Covid-19 pandemic, people are highly preferred to stay in home to work or study. A contact-less method to detect the stress level will be more suitable in nowadays. Therefore, nowadays researchers prefer to use RGB and thermal cameras to produce contact-less monitoring in stress detecting.

There are lots of researches and systems that used RGB or thermal camera to gather data and predict the stress. In this paper, I will use an improved Casper Network that is using Resnet to pretrained the data then put it into the Casper Network. The goal of this work is to help identify if Resnet can help improve the performance of the Casper Network. Besides the dataset in the spread sheet, the experiment will extract data from the video directly. This also requires to use finetuning to process the image of the data. With better performance of the Casper Network, we can have better result and accuracy when we trying to predict the stress. This will help people to monitor their own stress level and can manage their health as soon as possible.

2 Methodology

2.1 CasPer Algorithm

The Casper Algorithm is to be designed to solve the bad generalization property of Cascor. They share the same architecture but Casper has smaller networks and better generalization. The main difference of Casper and Cascor is that Casper does not use the correlation measure or weight freezing, the method that Casper used is resilient backpropagation. Figure 2 shows that how the Casper architecture works. At beginning, new hidden neuron learns remaining error with little interference from other neurons. There is no weight freezing and the Casper algorithm will use weight decay to enhance the generalization.

dE/dwij=dE/dwij-D*sign(wij)*wij**2*2^-0.01*Hepoch

Hepoch is the number of epochs elapsed when the last hidden neuron was added, sign is the sign of the operand, and k is a parameter that effects the weight decay that we used.

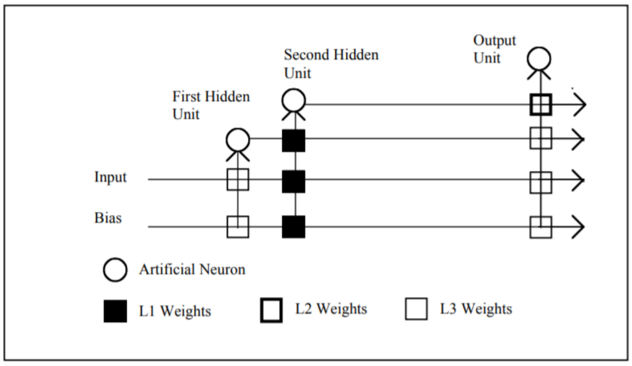

Figure 1. The Casper Algorithm
Source: Cascade Networks

From Figure 1, the relationship of the values of L1, L2 and L3 is that L1 >> L2 > L3. The high value of L1 let the new neuron to learn the error of the remaining network, then L2 allows the new neuron decrease the network error. The interference is called 'herd effect' (Fahlman, 1990).

2.2 The Resilient Backpropagation

The resilient backpropagation improves the normal backpropagation by using the sign of the gradient to overcome the poor and noisy estimator of the weight updates. The resilient backpropagation will increase the step size if the error gradient for the weight had the same sign in two consecutive epochs. If the sign is different, then decrease the step size. Because of the shallow slope in the sigmoid function, we will not be stopped with hard change weight.

$$\Delta_{ij}^{(t)} = \begin{cases} \eta^{+} \cdot \Delta_{ij}^{(t-1)}, & \text{if } \frac{\partial E^{(t-1)}}{\partial w_{ij}} \cdot \frac{\partial E^{(t)}}{\partial w_{ij}} > 0 \\ \eta^{-} \cdot \Delta_{ij}^{(t-1)}, & \text{if } \frac{\partial E^{(t-1)}}{\partial w_{ij}} \cdot \frac{\partial E^{(t)}}{\partial w_{ij}} < 0 \\ \Delta_{ij}^{(t-1)}, & \text{otherwise} \end{cases}$$

$$w_{ij}^{(t)} = \begin{cases} -\Delta_{ij}^{(t)}, & \text{if } \frac{\partial E^{(t)}}{\partial w_{ij}} > 0 \\ +\Delta_{ij}^{(t)}, & \text{if } \frac{\partial E^{(t)}}{\partial w_{ij}} < 0 \\ 0, & \text{otherwise} \end{cases}$$

The rule of the step size updates and the weight updates give us a great concept of how to change the weight while training.

2.3 Residual Neural Network

Residual Neural Network builds on constructs known from pyramidal cells in the cerebral cortex [6]. In order to perform this, the Residual Neural Networks will skip connections or shortcuts to jump over layers. The reasons that why the network want to add skip connections is to avoid gradients vanishing. The more layers that we add in deep model will cause higher training error. Normally, we consider the residual learning as a building block in Figure 2.

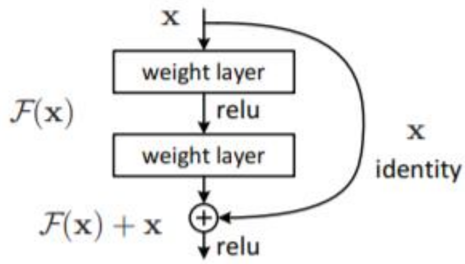

Figure 2. Building Block
Source: Deep Residual Learning for Image Recognition
(Kaiming He Xiangyu Zhang Shaoqing Ren Jian Sun)

The building block defined as:

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{W_i\}) + \mathbf{x}.$$

In the formula, x and y are the input and output vectors. The function F shows the mapping relationship that need to be learned. In the Figure 2, we have two layers, the F + x is showed by a shortcut connection. The shortcut connections in the equation above shows that we must have the same dimensions of x and F. The form of the function F is not stiff. It can have three or four layers. Skipping simplifies the network, it will speed up the learning by decreasing the vanishing gradients, then it will restore the skipped layers when it learns the feature space. At the end of training, it will stay very close to the manifold, therefore, it increases the learning speed. A neural network without residual parts explores more of the feature space. [6] This needs extra training data to trained and to cover. Thus, we consider to put ResNet into our Casper Network.

2.4 Fine Tune and Feature Extraction

Fine-tuning is to change all the weights when training on new tasks and feature extraction is to change the weights of the newly added last layers. So, in this paper, I will also perform the result using Fine Tune strategy and Feature Extraction strategy to see the performance of the network.

3 Data Analysis

3.1 Video Preprocessed

The data we have are lists of videos that record the reaction of people when give different kind of videos to watch. The camera is a thermal camera which shows people in a thermal image like Figure 3 shows.

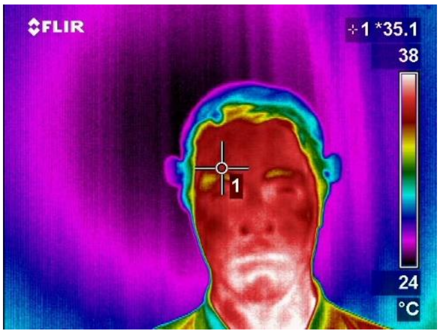

Figure 3: Thermal Image
Source: Author Analysis

There are 36 videos in the dataset, and 31 of these were used to match the dataset in the excel form. Through the videos, we know that there is 5 to 8 seconds setup at the beginning of the video and then there are 20 different kinds of videos (stressful movies or calm movies) are given in the experiment and each of the video is 90 seconds long. So, in order to separate the image of the video into 2 classes, stressful and calm. We need to first preprocess the images.
As the description above, for each video, we need to cut: firstly, the setup seconds for each video, Secondly, we also need to cut the gap between each video that is given in experiment. Finally, in each video, we extract the images of the people when watch different kinds of videos. Finally, we have 2349 images of stressful and 2322 images of calm for our training data. Then we have 828 images of calm and 901 images of stressful as our test data.

3.2 Excel Sheet data

The data that we have, listed 621 rows and 12 columns which shows that we have 31 subjects. For each subject, it represents a person that get tested. In order to predict the relationship between RGB fists and Thermal fists, we do not need to care about which person are we going to test. Thus, we can delete the first column. The remaining data has 11 columns, the first column shows the label of the data which indicate the output data in our datasets. The rest of 10 columns show the input data in our data.
For different subjects, there are some reasons that may have effect on the RGB and Thermal data. For example, gender and skin color etc. Thus, in order to increase the accuracy of the data, we first need to normalize the data. The table below shows the mean and standard deviation of the data.

Mean	Standard Deviation
221085.47851403317	332951.77871553676

Table 1. Mean and Standard Deviation

In order to have high performance on the algorithm and good testing performance, I chose to have train data, test data and validation data. The validation data can provide a good and unbiased evaluation of the Casper model fit on the training dataset. Because we only have two classes, the data is well balanced, we do not need to re-construct the data.
The dataset is quite small. Thus, the small datasets will lead to lower precision. In order to overcome the overfitting problem, when we train the data, we need to add bias on each weight and layers.

4 Results and Discussion

4.1 Optimization

The paper tries three methods to detect the stress, using Casper Network only, using ResNet only and the combination of
the Casper Network and ResNet. While setting up the parameters in the Casper Network, the experiment chose to have fixed L1, L2 and L3 with 0.1, 0.005 and 0.001 respectively. According to the paper A Cascade Network Algorithm Employing Progressive RPROP (N.K. Treadgold and T.D. Gedeon), this is the ideal learning rate. In the experiment, we chose to use Relu activation in the Casper Network because it is the fastest one among other functions. With 10 epochs, 20 epochs and 30 epochs for each experiment, the results are shown below.

Accuracy	10 epochs	20 epochs	30 epochs
Casper Network	51.61%	51.38%	50.23%
ResNet	58.31%	60.52%	62.00%
Casper + ResNet	73.78%	82.01%	83.60%

Table 2. The results of different Networks with 10, 20 and 30 epochs

According to the table 2, we can conclude that using ResNet and put the pretrained data into the Casper Network will enhance the performance. For each epochs shows above, the accuracy of the Casper and ResNet together is much better than the Casper or ResNet only. The result meets our expectation. Besides that, using fine tuning to extract the data from the video directly performs a better result than using the data excel sheet. The Casper Network only performs the prediction badly, only around 50%.

As there may be different when we use fine-tune or the feature extraction, therefore, Table 3 shows the result of the combination of using the Casper and ResNet in 20 epochs with fine-tune and feature extraction. From the result we can see that using fine-tune performs better results. The dataset is more suitable for the fine-tune strategy.

Accuracy	1	2	3	4	5
Fine-Tune	82.01%	81.79%	80.41%	82.31%	83.44%
Feature-Extraction	80.12%	80.34%	81.22%	79.85%	80.76%

Table 3. Casper and Resnet combination using Fine-Tune and Feature extraction

4.2 Training Accuracy

Taking look at the experiment at 20 epochs. Now I am going to discuss about the accuracy trend in each network. From figure 4 we can see that the training accuracy trend of Casepr + ResNet and ResNet is increasing, but for the Casper only it reduced. In the Casper Network, for each neuron that the network adds, the accuracy of the training data will change accordingly. The validation accuracy also increases in three networks. We can see from figure 3 that when we using Casper and ResNet together, the training accuracy and validation accuracy increase dramatically. This shows that the combination of the two networks is working very well. The trend also shows that the Casper Network only is not good for training the dataset that we have.

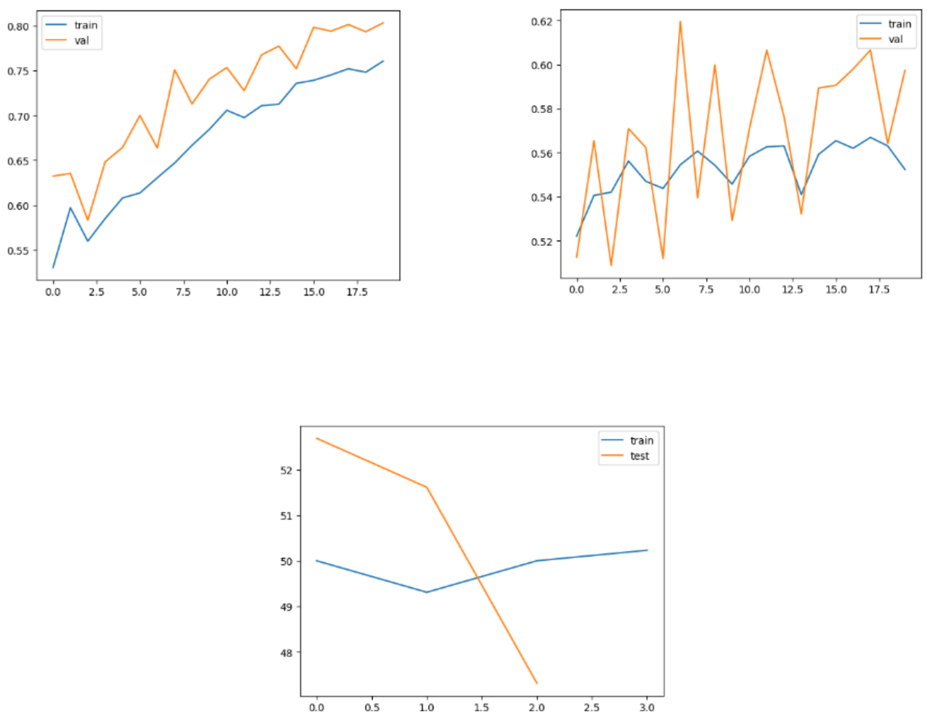

Figure 4. The accuracy of the training and validation data.
Top left is Casper + ResNet
Bottom is Casper only
Top right is ResNet only
Source: Author analysis

4.4 Loss Trend

Figure 5 shows the loss trend for each network, the loss of the Casper and ResNet combination is much better than the others. We can see that the trend is holistic decreasing in ResNet only and Casper and ResNet combination. However, the loss in Casper Network only shows that the trend is not stable. This shows that the dataset that we have is very poorly predicted. This shows that the accuracy of the Casper only network and ResNet only Network have poor prediction. This is may due to the firstly, small size of the dataset in the excel sheet. Secondly, the images that are pretrained is not suitable for the ResNet.

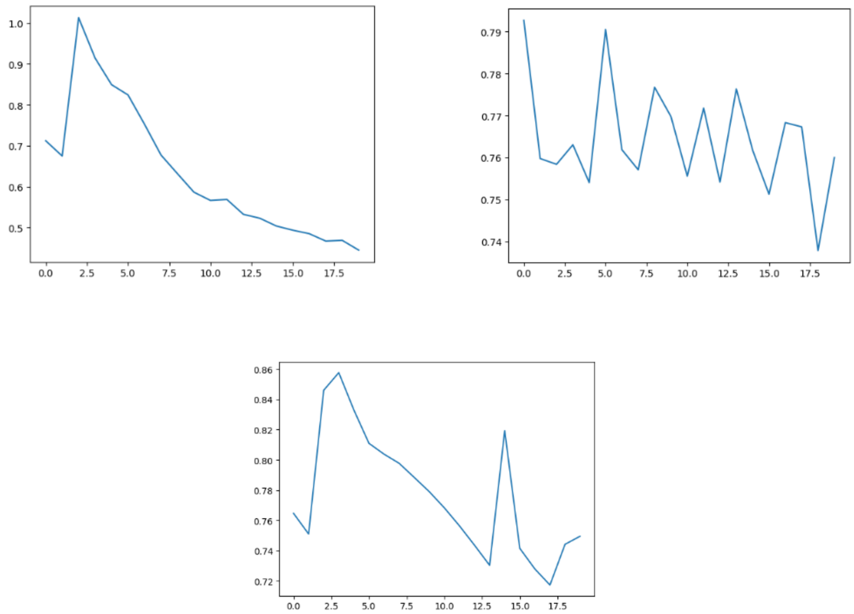

Figure 5. Loss Trend in 20 epochs for each Network
Top left is Casper + ResNet
Bottom is Casper only
Top right is ResNet only
Source: Author analysis

5 Conclusion, Limitations and Future Work

In conclusion, the combination of the Casper and ResNet enhance the stress prediction. The datasets that we have is not suitable for the Casper Network only, and the Casper Network performs very badly in stress prediction. The dataset we have is more suitable for the fine-tune rather than feature extraction. The limitation of this experiment is that the time to process the training takes too long. It took 30 minutes to train 30 epochs which is too long. What's more, the experiment only using a simple approach of the feature extraction, the result maybe different when use another approach of the feature extraction. In the future work, the feature extraction can be improved by using Evolutionary Algorithm such as genetic algorithm. Or we can implement fuzz logic to cluster the dataset to have a better performance.

6 References

[1] W. Liao, W. Zhang, Z. Zhu, and Q. Ji, "A real-time human stress monitoring system using dynamic bayesian network," in Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on. IEEE, 2005, pp. 70–70.

[2] F. Bousefsaf, C. Maaoui, and A. Pruski, “Remote assessment of the heart rate variability to detect mental stress,” in Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2013 7th International Conference on. IEEE, 2013, pp. 348–351
[3] Jennifer Casarella “Stress Symptoms” in WebMd, 2019
[4] E. Jovanov, A. O'Donnell Lords, D. Raskovic, P. G. Cox, R. Adhami and F. Andrasik, "Stress monitoring using a distributed wireless intelligent sensor system," in *IEEE Engineering in Medicine and Biology Magazine*, vol. 22, no. 3, pp. 49-55, May-June 2003, doi: 10.1109/MEMB.2003.1213626.
[5] Fahlman, S.E., and Lebiere, C. (1990) The cascade-correlation learning architecture. In Advances in Neural Information Processing II, Touretzky, Ed. San Mateo, CA: Morgan Kauffman, 1990, pp. 524-532.
[6] He, Kaiming; Zhang, Xiangyu; Ren, Shaoqing; Sun, Jian (2015-12-10). "Deep Residual Learning for Image Recognition". arXiv:1512.03385