CSC413 Proposal

Keyi Zhang Jiayan Kong Xiang Chen

Abstract

We propose to reproduce multi-CNN deep learning algorithms and re-evaluate the implementations against some standard benchmarks in the realm of facial expression recognition. In this project, we will mainly focus on the ResNet including its variations and AttentionNet. Both of these two deep learning models have shown a huge success on features extractions, latent features detections and object classifications. Therefore, we would like to research more deeply for these two deeping learning algorithms and apply them to facial expression recognition. We will also perform sensitivity analysis on hyper-parameters for both models. In the end of this project, we expect to have comparisons in multi-aspects for these two but not limited to ResNet and AttentionNet models .

1 Intro

Facial expression is a key signal of human emotion ***(Zijun Cui 2020)***. More and more advanced deep learning models have been researched since the early of 21-st century. Research on facial expression recognition in recent years focuses on deep neural networks, such as AlexNet and VGGNet, to obtain powerful representations and classifications. After we have gone through *Knowledge Augmented Deep Neural Networks for Joint Facial Expression and Action Unit Recognition* ***(NIPS 2020)****, Joint Representation and Estimator Learning for Facial Action Unit Intensity Estimation* ***(CVPR 2019)****, and Facial Expression Recognition Based on VGGNet Convolutional Neural Network* ***(IEEE 2018)****,* we find there have been huge research in this open area and we would like to put other advanced and popular deep learning models on the realm of facial expression recognition.

2 Related Works

We collected, analyzed some papers from arXiv.org, NeurIPS and IEEE that released a novel and upgraded implementation of CNN focusing on facial expression recognition. We want to make a comparison between two world-leading technologies in facial expression recognition that have some different points of focus but all improve the FER model in terms of boosting the accuracy.

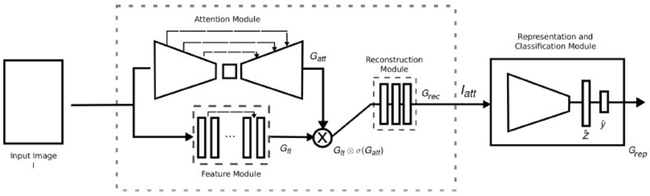
2.1Attention Net: In 2014, Liu et al. [4] proposed a facial expression recognition framework using 3DCNN to jointly localize facial action parts and learn part-based representations for expression recognition. In 2015, Yu and Zhang[5] used stochastic pooling and ensemble of CNNs to achieve great success in the EmotiW challenge. In 2019, based on the FER model, Fernandez[6] and his team at California Institute of Technology contributed to a new network architecture that includes an encoder-decoder style network called attention module and this improved the system classification performance in comparison to other methods from the state-of-the-art.

2.2 Residual Net: Besides the fact that researchers are trying to improve the accuracy of the model considering the feature extraction, using ResNet is also a way to optimize and improve the accuracy from considerably increased depth of CNN. The Inception-ResNet module achieved remarkable performance in image classification and object detection and uses ReLU as the activation function. However, ReLU would result in the gradients of negative inputs becoming zero. After some improvements on ReLU functions being made, in 2020, Peng et al.[7] replaced the ReLU function in Inception-ResNet module with PReLU and trained a new version of the model that also improved the performance as well as the stability.

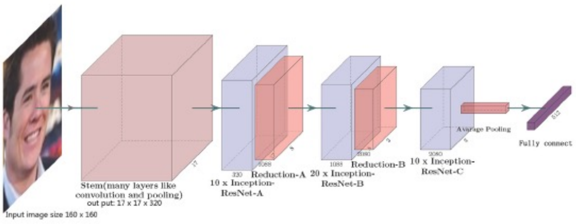
Our goal is to compare the two models in different aspects and try to combine them with the action unit detection model or to apply them to a new field.

3 Algorithm & Method

We reproduce the experimental results of Attention Network and Inception-ResNet on the same list of databases. Moreover, we are going to compare the computational cost, trading time and accuracy between the two models.



Architecture of FERAtt(Facial Expression Recognition with Attention Net)



Architecture of Inception-ResNet

Due to the fact that we have not fully experienced the models yet, we haven’t come up with the specific algorithms. We would really appreciate having a discussion with TAs in this part to get some insight.

4 Summary

In this work, we applied two deep learning models, ResNet including its variations and AttentionNet, to the task of facial expression recognition. We use ResNet as the baseline to compare the advantages and disadvantages of the two methods. We also conducted a sensitivity analysis on the hyperparameters of the two models.

For future work, we plan to conduct more research on how ResNet can gain more expressive power, such as by using skip connections.

Reference

[1] Cui, Z., Song, T., Wang, Y., & Ji, Q. (2020). Knowledge Augmented Deep Neural Networks for Joint Facial Expression and Action Unit Recognition. *Advances in Neural Information Processing Systems*, *33*.

[2] Zhang, Y., Wu, B., Dong, W., Li, Z., Liu, W., Hu, B. G., & Ji, Q. (2019). Joint representation and estimator learning for facial action unit intensity estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3457-3466).

[3] Jun, H., Shuai, L., Jinming, S., Yue, L., Jingwei, W., & Peng, J. (2018, November). Facial expression recognition based on VGGNet convolutional neural network. In *2018 Chinese Automation Congress (CAC)* (pp. 4146-4151). IEEE.

[4] Liu, P., Han, S., Meng, Z., & Tong, Y. (2014). Facial expression recognition via a boosted deep belief network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1805-1812).

[5] Yu, Z., & Zhang, C. (2015, November). Image based static facial expression recognition with multiple deep network learning. In *Proceedings of the 2015 ACM on international conference on multimodal interaction* (pp. 435-442).

[6]Marrero Fernandez, P. D., Guerrero Pena, F. A., Ren, T., & Cunha, A. (2019). Feratt: Facial expression recognition with attention net. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 0-0).

[7] Peng, S., Huang, H., Chen, W., Zhang, L., & Fang, W. (2020). More trainable inception-ResNet for face recognition. *Neurocomputing*, *411*, 9-19.