Turn Trash into Treasure: Ensemble Learning for Face Classification

Zhefan Xu, Hongyi Zhou, Yidong He

Carnegie Mellon University, Mechanical Engineering Department

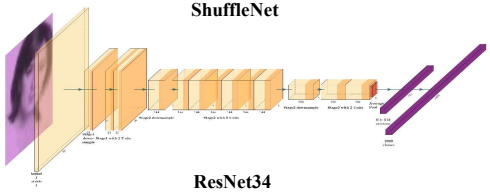
ABSTRACT

Face recognition performs the task of determining the identity of the image of a person's face. This task is essentially a **N-way classification** problem. A series of Convolutional Neural Network classifier has been developed to increase the accuracy of classification. In this project, we aim to further improve the accuracy by using an ensemble of three state-of-the-art CNN architectures **ResNet**, **ShuffleNet v1**, and **MobileNet v2**, each of which has been trained on the same dataset. For every face image, each ensemble member network produce 512 facial embeddings which are then concatenated together and fed into another consensus scheme network to decide the collective classification. Final results show that the ensemble network is able to achieve **15%~20%** higher accuracy than individual CNN network.

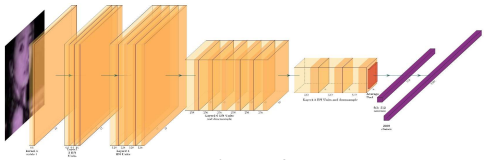
The dataset contains ~820k single-channel training image with size 32x32. The total number of classes is 2300. This image shows some training samples.

NEURAL NETWORKS ARCHITECTURE

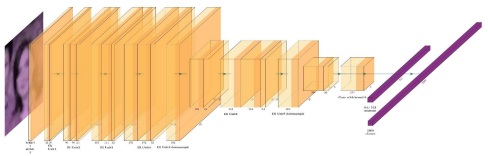
ShuffleNet



ResNet34



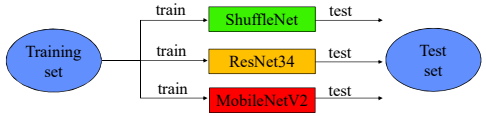
MobileNetV2



METHODOLOGY

Train Three Networks Separately

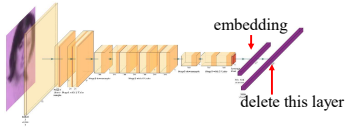
- Train ShuffleNet, ResNet34, MobileNetV2 from scratch using the same training data set.
- Evaluate them on test set. Got relatively low test accuracy.



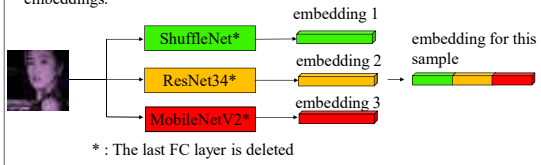
	#of epochs in training	Train accuracy	Validation Accuracy	Test accuracy
MobileNetV2	22	78.6%	73.0%	62.7%
ResNet34	30	82.3%	72.6%	67.3%
ShuffleNet	18	83.5%	68.3%	55.7%

Make Embeddings for Each Image

- Delete the last Fully Connected Layer in three networks. This FC layer makes N-way classification from 512-dimension embedding.



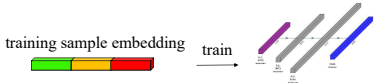
- For each image, use theses three networks to create three 512-dimension embeddings.



* : The last FC layer is deleted

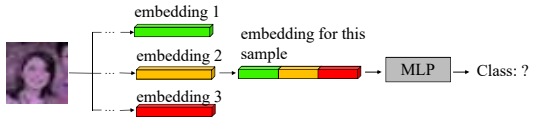
Train an MLP using Embeddings

- Train a multilayer perceptron using the embeddings of each training sample.



Test the results

- Create the embedding for test image in the same way.
- Test the classification results.



EXPERIMENT RESULTS & CONCLUSION

Model	Train Accuracy	Test Accuracy
MobileNetV2	78.6%	62.7%
ResNet34	82.7%	67.3%
ShuffleNet	83.5%	55.7%
Ensemble Net1	**99.8%**	**77.5%**
Ensemble Net2	**95.8%**	**77.2%**
Ensemble Net3	**99.2%**	**74.0%**

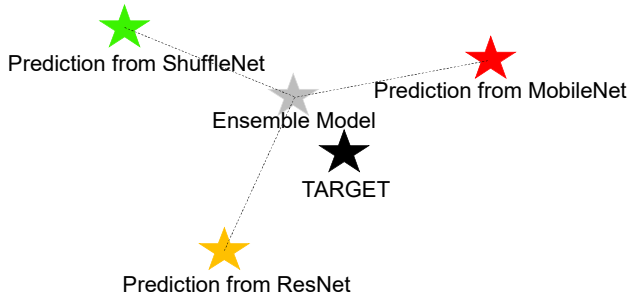
Ensemble Net1: 3 Linear Layer MLP with ReLU activation.
Ensemble Net2: 3 Linear Layer MLP with dropout rate 0.4, 0.3 for the first two linear layers.
Ensemble Net3: 4 Linear Layer MLP with dropout rate 0.2, 0.1 for the first two linear layers.

Conclusion:

Ensemble model outperforms three individual models by combing their predictions. Even though the individual models have poor performance (way below 70%) and are useless, the ensemble technique gets much better result by simply adding a small MLP layer.

ANALYSIS

Ensemble learning reduces the variance of prediction for individual model. We use tried difference MLP models to get better performance than individual model. This diagram shows why ensemble learning performs better:



References:

[1] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. doi: 10.1109/cvpr.2016.90

[2] Zhang, Xiangyu, et al. "ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices." 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, doi:10.1109/cvpr.2018.00716.

[3] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. doi:10.1109/cvpr.2018.00474