Convolutional Neural Networks for Facial Expression Recognition

Shima Alizadeh, Azar Fazel
(shima86@stanford.edu, azarf@stanford.edu)
Stanford University

Introduction

Motivation:

- The automatic recognition of facial expressions has been an active research topic since the early nineties.
- It has several applications in Human Computer Interaction (HCI), psychology and neuroscience.
- Development of an automated system to accomplish the expression classification task with high accuracy is challenging.

Dataset:

- Provided by Kaggle website
- Consists of ~37000 well-structured 48x48$ pixel grayscale images of faces.
- Each image is categorized in one of the seven classes that express different facial emotions (Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral)

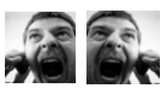
Goal:

- Develop a Convolutional Neural Networks (CNN) model to classify facial expressions into 7 different classes.

Methodology

Data Preprocessing and Augmentation:

- Normalizing images by subtracting the mean
- Random image flipping



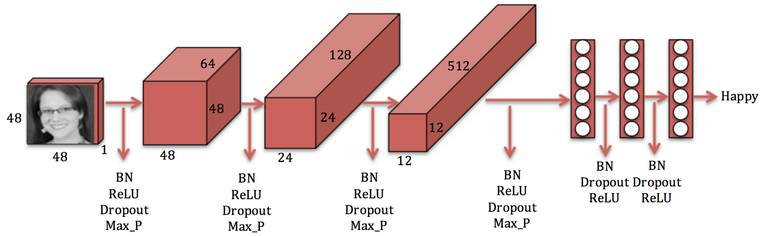
An example of image flipping

Features:

- Generated by convolution layers using the raw pixel data
- Extracted Histogram of Oriented Gradient (HOG) features for each image and using them along with raw pixels. For this part, we built a new learning model containing two neural networks: the first one contains convolutional layers, and the second one fully connected layers only. The developed features by the first network are concatenated with the HOG features and the resultant hybrid features are fed in to the second network.

Network Properties:

- Batch normalization to achieve high accuracy in fewer training steps
- Dropout to prevent the network from overfitting.
- ReLu as the activation function to add non-linearity to the network
- Zero-padding to control the spatial size of the output volume of each layer
- Different strides to control the depth columns around the spatial dimensions
- Flexible network architecture: different number of convolutional layers and hidden layers with user-specified settings (number of neurons, filter size, max-pooling, etc.)
- Log Softmax loss function
- Utilizing the GPU accelerated deep learning facilities on Torch to make the model training process faster

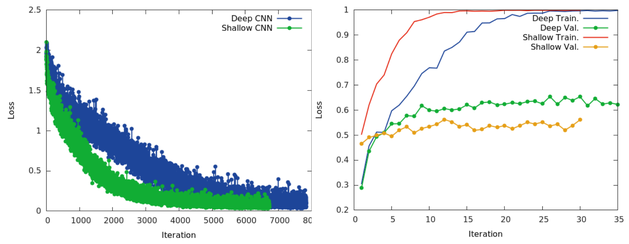


The architecture of the network: 3 convolutional layers and 2 fully connected layers

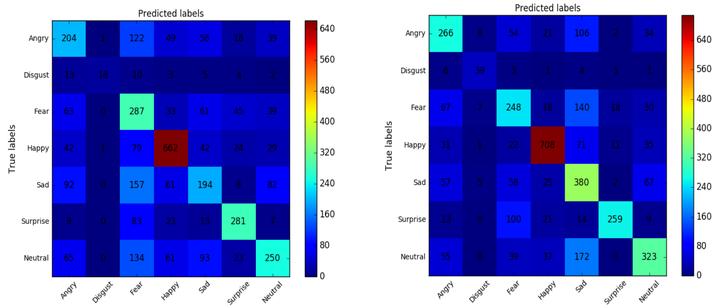
Model Performance

Shallow Network vs. Deep Network:

Deep network enables us to increase the validation accuracy by %18.46. According to the right plot, one can observe that the deep network has reduced the overfitting behavior of the learning model by adding more nonlinearities and hierarchical usage of anti-overfitting techniques such as dropout and batch normalization in addition to L2 regularization. From the left plot we can also see that the shallow network has been converged faster and the training accuracy has quickly reached to its highest value.



The confusion matrices were computed for the shallow and deep networks. The result demonstrated that the deep network results in higher true predictions for most of the labels. It is interesting to see that both models have performed well in predicting the happy label, which implies that the texture of a happy face is learnt easier than other expressions. This visualization technique also reveals which labels are likely to be confused by the trained networks. For instance, we can see the correlation of angry label with fear and sad labels, which makes sense based on their facial textures.

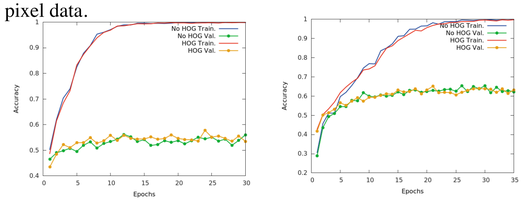


Confusion matrix for shallow CNN model

Confusion matrix for deep CNN model

CNN Models with Hybrid Features:

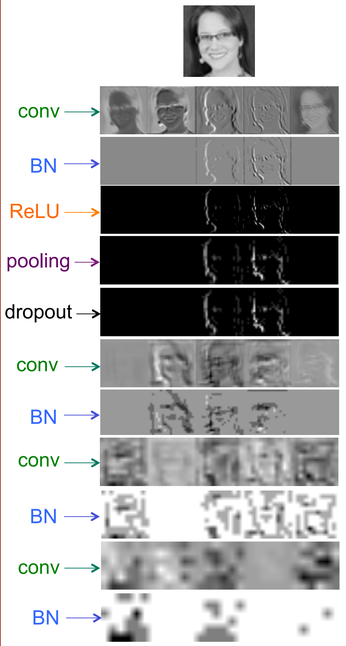
We developed learning models that concatenate the HOG features with those generated by convolutional layers and give them as input features to FC layers. Using this model, we trained one shallow and one deep network. As seen in the figures below, the accuracy of the model is very close to the accuracy we got from the model that has no HOG features. This concludes that CNN is strong enough to extract enough information including those coming from HOG features by using only raw pixel data.



Hyper parameters obtained by cross for the model

Parameter	Learning rate	Regularization	Hidden neurons
Values	0.01	1e-7	256, 512

Activation Map Visualization



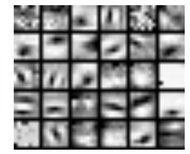
Layer Visualization:

- To see the output of each layer, we visualized the activation maps of different layers during the forward pass (figure to the right). As the training progresses the activations become more sparse and localized.

Weight Visualization:

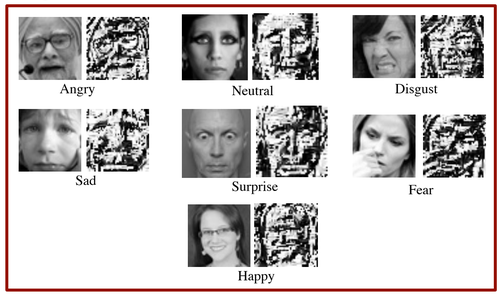
- To see the qualification of the trained network, we also visualized the weights of the first layer. As seen in the figure below, we have smooth filters without any noisy pattern.

The first 30 7x7 filters in the first convolutional layer



DeepDream!

- We applied DeepDream idea to our best predictive model to find enhance pattern in our images. The figures below displays the results for each emotion along with the real image.



Summary and Future Works

- We developed various CNNs for a facial expression recognition problem and evaluated their performances using different post-processing and visualization techniques.
- The results demonstrated that deep CNNs are capable of learning facial characteristics and improving facial emotion detection.
- The hybrid feature sets did not assist in further improving the model accuracy, which concludes that the convolutional networks can intrinsically learn the key facial features per se by only using raw data.

✧ As a future work, we would like to extend our model to color images. This will allows us to investigate the efficacy of pre-trained models such as AlexNet or VGGNet for facial emotion recognition.
✧ Another extension would be the implementation of a face detection process, which is followed by the emotion prediction.

References

[1] Nicu Sebe, Michael S. Lew, Ira Cohen, Yafei Sun, Theo Gevers, Thomas S. Huang (2007) Authentic Facial Expression Analysis. Image and Vision Computing 25.12: 1856-1863
[2] Bettadapura, Vinay (2012). Face expression recognition and analysis: the state of the art. arXiv preprint arXiv:1203.6722
[3] Lonare, Ashish, and Shweta V. Jain (2013). A Survey on Facial Expression Analysis for Emotion Recognition. International Journal of Advanced Research in Computer and Communication Engineering 2.12
[4] Dalal, Navneet, and Bill Triggs (2005). Histograms of oriented gradients for human detection. Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society Conference on. Vol. 1