Facial Recognition with Deep Learning

Bekhzod Umarov BUMAR1@UNH.NEWHAVEN.EDU
University of New Haven, 300 Boston Post Rd., West Haven, CT 06516 USA

Abstract

This paper proposes a simple deep convolutional neural network system for facial recognition. This deep learning facial recognition system aims to reach human level accuracy while keeping the network as simple as possible. The paper also discusses different approaches in designing neural networks for facial recognition by using two different datasets and comparing their results

1. Introduction

Facial recognition is considered to be among the classical challenges in computer vision. It recently became one of the hot topics in the study of artificial neural networks and has attracted a great deal of research interest (Lu & Tang, 2014; Parmar & Mehta, 2014; Zhou et al., 2015). Although there are no known best setups for every type of neural network, there are specific types of neural networks that work the best on a given problem. According to the research done by researchers at Microsoft convolutional neural networks seem to work best with visual documents (Simard et al., 2003). Therefore, I have decided to use CNN for image recognition.

For this project I have used LFW (Labeled Faces in the Wild) benchmark (Huang et al., 2007) which is considered to be one of the most difficult constrained datasets for a machine to learn according to Zhou, Cao, and Yin (2015). Unfortunately, there was a small issue working with the dataset. Due to lack of access to high performance machines, I was forced to cut down my dataset size. The approach used was inspired by scikit-learn's (Pedregosa et al., 2011) example[1] of using PCA for facial recognition. Their example suggests using a minimum of 70 pictures per person. This limits the dataset to only 1288 images total without any tuning.

Figure 1. Sample faces from LFW dataset

The neural network proposed in this paper is mainly aimed to outperform "out-of-the-shelf" machine learning algorithms for facial recognition such as PCA Eigenfaces, SVMs, and etc. This project design is not intended to outperform any professional designed facial recognition neural network systems such as Megvii, DeepID, or DeepFace. The intent of this research is to understand the relationship between neural networks and datasets, while aiming for a human level accuracy in recognizing faces of people in a given dataset.

In order to better evaluate the network's performance it was also tested with a dataset containing look alike faces. This would help in understanding how the network will perform against the real world problems.

2. Face Recognition System

2.1. Simple CNN

The simple CNN system with one two-dimensional convolution and localization, followed by one two-by-two max-pooling and fully connected layers was able to achieve 93.78% accuracy rate. This result already shows a huge improvement compared to the out of the box face recognition algorithm discussed earlier that uses eigenfaces and SVMs taken from Scikit-Learn website which is only able to achieve 86% accuracy.

However, the same face recognition system using simple CNN fails significantly when it comes to differentiating

[1] http://scikit-learn.org/stable/auto_examples/applications/face_recognition.html

CSCI 6660: Artificial Intelligence. Bekhzod Umarov.

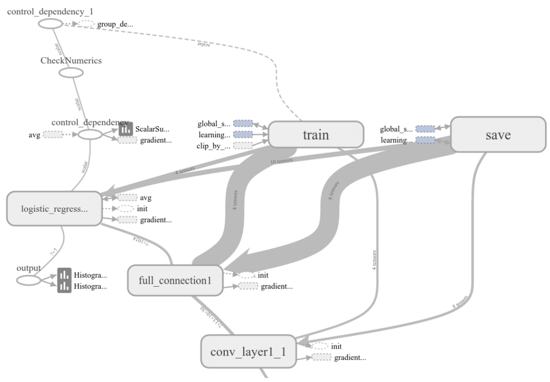

Figure 2. Simple Convolutional Neural Network graph generated by Tensorflow

look alike faces from the real people achieving approximately 10% accuracy. The evaluation test was performed against the "Look Alike Face Database"[2] acquired from the Image Analysis and Biometrics Lab at IIIT Delhi website[3] (Lamba et al., 2011).

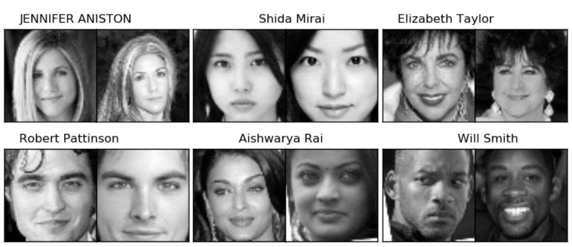

Figure 3. Sample faces from Look Alike People Dataset with impostors positioned right of the original person's picture

This raises two different issues. The high error rate could be due to the dataset being considerably smaller than the previously used LFW dataset (500 vs. 1288). An alternative solution to this error could be the shallow network restricting its ability to learn necessary features of each person. In order to identify the reason for the high error rate, both issues were addressed.

2.2. Deep CNN

The Deep CNN approach assumes that a shallow network loses the ability to learn enough features to be able to differentiate between people. Therefore, we assume that adding more layers perhaps could improve the existing results. While designing the deep CNN, the final prediction rate fluctuated a lot depending on the way the neural network was designed, sometimes even going as low as 70%. This indicated that with the increase of convolution layers the risk of overtraining also increases. Therefore, in order to be able to have better control over the results previous neural network was increased just by one extra layer. Still, with the LFW dataset the performance of the neural network did not improve as was expected. The final prediction rate for the LFW dataset ended up being 87%. The same deep CNN resulted in an accuracy rate fluctuating between 7% and 15%.

These results show that deeper convolution networks depending on the dataset can improve the results, but it can also decrease them. Most possible explanation is that with deeper neural networks it is very easy to overtrain the network with the way it is being set up.

2.3. Simple CNN + Artificial Dataset

The second approach assumes the dataset size is directly related to the available amount of information about each person in the dataset. Since the Look Alike People dataset size is very small, the dataset was extended artificially by randomly picking images from the dataset and flipping them horizontally or adding a bit of noise to the image. This approach extends the dataset from having a total of 500 visual documents to 1250, which is very close to the previously used LFW dataset's size.

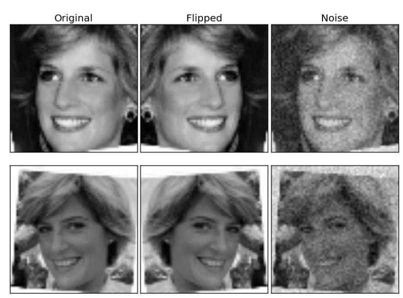

Figure 4. Preprocessed LAP sample images of Princess Diane and her look alike that were used to extend the LAP dataset

With this new artificially extended dataset, our face recognition system using shallow CNN is able to achieve an accuracy rate of 66.67% when LAP dataset is used.

This result supports the argument that large amounts of information regarding each class must be fed into the system during the training process. This is must be done in order to achieve strong confidence in a neural network. This indicates the more features the neural network learns, the more confident it will be in identifying those features in the real world. This also supports the previous results that we were received from the LFW dataset. Since minimum of 70 images were used per person, the dataset had a lot more

[2] http://iab-rubric.org/databases/LADatabase.rar

[3] http://www.iab-rubric.org/resources.html

information per person compared to the people in the LAP dataset.

The LFW dataset with 1288 images consists of only 7 classes, whereas LAP dataset with 1250 images consists of 100 classes. Therefore, the extended LAP dataset is still significantly smaller in size compared to the LFW dataset. The ratio of features per class between datasets is extremely low. Hence, the facial recognition system is only able to achieve a 66.67% prediction rate. However, when the LFW dataset is extended the same way, neural network is able to result in 95.56% accuracy.

2.4. Deep CNN + Artificial Dataset

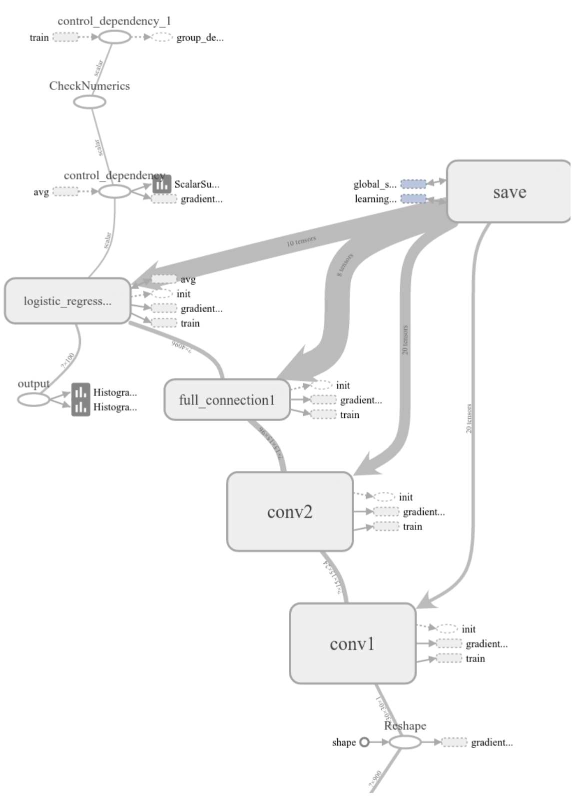

Figure 5. Deep Convolutional Neural Network graph generated by Tensorflow

This approach is used to get the most out of the both previous approaches in order to achieve higher confidence rate. By using this approach the facial recognition system is able to achieve 71.99% accuracy rate with the extended LAP dataset. However, the same neural network is only able to achieve 91.08% accuracy with the extended LFW dataset. Although, with a little tweaking of convolution layers' of the deep neural network the accuracy goes back up to 95.78% accuracy.

These results show that deeper neural networks' performance is directly related to the actual dataset's features, i.e. sample sizes, dataset size, or any other possible features.

3. Future Research

3.1. Deeper CNN + Artificial Dataset

As the research shows, increasing the number of layers as well as the information per class will result in the ability of the neural network distinguishing between features of people with higher accuracy. This can also be supported by the results of DeepID convolutional neural network (Ouyang et al., 2014). However, increase of layers in a neural network can result in a higher risk of overtraining. Therefore, more research needs to be done in order to build a better deep convolutional neural network.

3.2. CNN+RNN

Recent studies show that by using Recurrent Neural Networks (RNNs) machines are able to generate images ("dream") (Theis & Bethge, 2015; Graves, 2013; Gregor et al., 2015) with log-likelihood scores (Oord et al., 2016). With the use of these neural networks, it could be possible to achieve even better results. Especially with images that are taken in non-standard conditions and/or missing some of features.

References

Graves, Alex. Generating sequences with recurrent neural networks. *CoRR*, abs/1308.0850, 2013. URL http://arxiv.org/abs/1308.0850.

Gregor, Karol, Danihelka, Ivo, Graves, Alex, and Wierstra, Daan. DRAW: A recurrent neural network for image generation. *CoRR*, abs/1502.04623, 2015. URL http://arxiv.org/abs/1502.04623.

Huang, Gary B., Ramesh, Manu, Berg, Tamara, and Learned-Miller, Erik. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.

Lamba, H., Sarkar, A., Vatsa, M., Singh, R., and Noore, A. Face recognition for look-alikes: A preliminary study. In *Biometrics (IJCB), 2011 International Joint Conference on*, pp. 1–6, Oct 2011. doi: 10.1109/IJCB.2011.6117520.

Lu, Chaochao and Tang, Xiaoou. Surpassing human-level face verification performance on LFW with gaussianface. *CoRR*, abs/1404.3840, 2014. URL http://arxiv.org/abs/1404.3840.

Oord, Aäron Van Den, Kalchbrenner, Nal, and Kavukcuoglu, Koray. Pixel recurrent neural networks. *CoRR*, abs/1601.06759, 2016. URL http://arxiv.org/abs/1601.06759.

Ouyang, Wanli, Luo, Ping, Zeng, Xingyu, Qiu, Shi, Tian, Yonglong, Li, Hongsheng, Yang, Shuo, Wang, Zhe, Xiong, Yuanjun, Qian, Chen, Zhu, Zhenyao, Wang, Ruohui, Loy, Chen Change, Wang, Xiaogang, and Tang, Xiaoou. Deepid-net: multi-stage and deformable deep convolutional neural networks for object detection. *CoRR*, abs/1409.3505, 2014. URL http://arxiv.org/abs/1409.3505.

Parmar, Divyarajsinh N. and Mehta, Brijesh B. Face recognition methods & applications. *CoRR*, abs/1403.0485, 2014. URL http://arxiv.org/abs/1403.0485.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Simard, Patrice Y., Steinkraus, Dave, and Platt, John C. Best practices for convolutional neural networks applied to visual document analysis. Institute of Electrical and Electronics Engineers, Inc., August 2003. URL http://research.microsoft.com/apps/pubs/default.aspx?id=68920.

Theis, L. and Bethge, M. Generative image modeling using spatial lstms. In *Advances in Neural Information Processing Systems 28*, Jun 2015. URL http://arxiv.org/abs/1506.03478.

Zhou, Erjin, Cao, Zhimin, and Yin, Qi. Naive-deep face recognition: Touching the limit of LFW benchmark or not? *CoRR*, abs/1501.04690, 2015. URL http://arxiv.org/abs/1501.04690.