Deeply learned face representations are sparse, selective, and robust

Yi Sun[1], Xiaogang Wang[2,3], Xiaoou Tang[1,3]
[1]Department of Information Engineering, The Chinese University of Hong Kong. [2]Department of Electronic Engineering, The Chinese University of Hong Kong. [3]Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences.

Face recognition achieved great progress thanks to extensive research effort devoted to this area. While pursuing higher performance is a central topic, understanding the mechanisms behind it is equally important. When deep neural networks begin to approach human on challenging face benchmarks [3, 4, 5] such as LFW [2], people are eager to know what has been learned by these neurons and how such high performance is achieved. In cognitive science, there are a lot of studies [7] on analyzing the mechanisms of face processing of neurons in visual cortex. Inspired by those works, we analyze the behaviours of neurons in artificial neural networks in a attempt to explain face recognition process in deep nets, what information is encoded in neurons, and how robust they are to corruptions.

Our study is based on a high-performance deep convolutional neural network (deep ConvNet), referred to as DeepID2+, proposed in this paper. It is improved upon the state-of-the-art DeepID2 net [3] by increasing the dimension of hidden representations and adding supervision to early convolutional layers. The best single DeepID2+ net (taking both the original and horizontally flipped face images as input) achieves 98.70% verification accuracy on LFW (vs. 96.72% by DeepID2). Combining 25 DeepID2+ nets sets new state-of-the-art on multiple benchmarks: 99.47% on LFW for face verification (vs. 99.15% by DeepID2 [3]), 95.0% and 80.7% on LFW for closed- and open-set face identification, respectively (vs. 82.5% and 61.9% by Web-Scale Training (WST) [6]), and 93.2% on YouTubeFaces [8] for face verification (vs. 91.4% by DeepFace [5]).

With the state-of-the-art deep ConvNets and through extensive empirical evaluation, we investigate three properties of neural activations crucial for the high performance: sparsity, selectiveness, and robustness. They are naturally owned by DeepID2+ after large scale training on face data, and we did NOT enforce any extra regularization to the model and training process to achieve them. Therefore, these results are valuable for understanding the intrinsic properties of deep networks.

It is observed that the neural activations of DeepID2+ are moderately sparse. As examples shown in Fig. 1, for an input face image, around half of the neurons in the top hidden layer are activated. On the other hand, each neuron is activated on roughly half of the face images. Such sparsity distributions can maximize the discriminative power of the deep net as well as the distance between images. Different identities have different subsets of neurons activated. Two images of the same identity have similar activation patterns. This motivates us to binarize the neural responses in the top hidden layer and use the binary code for recognition. Its result is surprisingly good. Its verification accuracy on LFW only slightly drops by 1% or less. It has significant impact on large-scale face search since huge storage and computation time is saved. This also implies that binary activation patterns are more important than activation magnitudes in deep neural networks.

Related to sparseness, it is also observed that neurons in higher layers are highly selective to identities and identity-related attributes. When an identity (who can be outside the training data) or attribute is presented, we can identify a subset of neurons which are constantly excited and also can find another subset of neurons which are constantly inhibited. A neuron from any of these two subsets has strong indication on the existence/non-existence of this identity or attribute, and our experiment show that the single neuron alone has high recognition accuracy for a particular identity or attribute. In other words, neural activations have sparsity on identities and attributes, as examples shown in Fig. 1. Although DeepID2+ is not taught to distinguish attributes during training, it has implicitly learned such high-level concepts. Directly employing the face representation learned by DeepID2+ leads to much higher classification accuracy on identity-related attributes than widely used handcrafted features such as high-dimensional LBP [1].

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

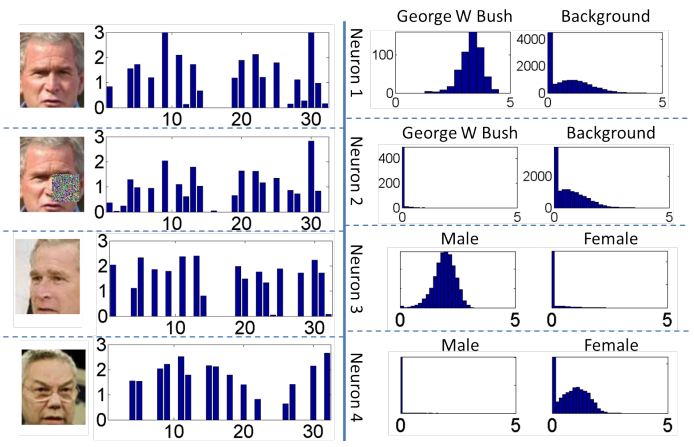

Figure 1: Left: neural responses of DeepID2+ on images of Bush and Powell. The second face is partially occluded. There are 512 neurons in the top hidden layer of DeepID2+. We subsample 32 for illustration. Right: a few neurons are selected to show their activation histograms over all the LFW face images (as background), all the images belonging to Bush, all the images with attribute Male, and all the images with attribute Female. A neuron is generally activated on about half of the face images. But it may constantly have activations (or no activation) for all the images belonging to a particular person or attribute. In this sense, neurons are sparse, and selective to identities and attributes.

Our empirical study shows that neurons in higher layers are much more robust to image corruption in face recognition than handcrafted features such as high-dimensional LBP or neurons in lower layers. As an example shown in Fig. 1, when a face image is partially occluded, its binary activation patterns remain stable, although the magnitudes could change. We conjecture the reason might be that neurons in higher layers capture global features and are less sensitive to local variations. DeepID2+ is trained by natural web face images and no artificial occlusion patterns were added to the training set.

[1] Dong Chen, Xudong Cao, Fang Wen, and Jian Sun. Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification. In *Proc. CVPR*, 2013.

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

[3] Yi Sun, Yuheng Chen, Xiaogang Wang, and Xiaoou Tang. Deep learning face representation by joint identification-verification. In *Proc. NIPS*, 2014.

[4] Yi Sun, Xiaogang Wang, and Xiaoou Tang. Deep learning face representation from predicting 10,000 classes. In *Proc. CVPR*, 2014.

[5] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. DeepFace: Closing the gap to human-level performance in face verification. In *Proc. CVPR*, 2014.

[6] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Web-scale training for face identification. Technical report, arXiv:1406.5266, 2014.

[7] Doris Y. Tsao and Margaret S. Livingstone. Neural mechanisms for face perception. *Annu Rev Neurosci*, 31:411–438, 2008.

[8] Lior Wolf, Tal Hassner, and Itay Maoz. Face recognition in unconstrained videos with matched background similarity. In *Proc. CVPR*, 2011.