

# Exploring Features and Attributes in Deep Face Recognition Using Visualization Techniques

Yaoyao Zhong, Weihong Deng

**Abstract—** Deep convolutional neural networks (CNNs) currently have achieved state-of-the-art results on face recognition; yet, the understanding behind the success of the deep face model is still lacking. In particular, it is still unclear the inner workings of deep face model. What effective features does a deep face model learn? What do these features represent and what is the semantic meaning of them? This work explores this problem by analyzing the classic network VGGFace using deep visualization techniques. We first explore features computed by neurons, investigating characters of features like diversity, invariance, discrimination. It's worth noting that the middle layer is the least robust to transform, which contradicts the conventional view that robustness to transform increases as the network going deeper. The most significant phenomenon we find is that high level features are correspond with complex face attributes which human could not describe using a few words. We present a quantitative analysis on these face attributes perceived by deep CNNs, understanding them and the complex relationships between them. Additionally, we also focus on the significant point, the pose invariance in face recognition. Our research is the first work to understand the inner works of deep face models, elucidating some particular phenomena in deep face recognition.

**For the published version of record document, go to:**

**<http://dx.doi.org/10.1109/FG.2019.8756546>**













