

Discriminative Attention: Reducing Visual Confusion and Bridging Multi-tasks of Face Analysis

Lezi Wang
Dept. of Computer Science

11/29/2018 at 02:00 pm
CBIM 22

Abstract

Recent developments in gradient-based attention modeling have led to improved model interpretability by means of class-specific attention maps. First, we address that the key limitation of these approaches is that the resulting attention maps while being well localized, are not class discriminative. We propose a new learning framework that makes class-discriminative attention and cross-layer attention consistency a principled and explicit part of the learning process. Furthermore, our framework provides attention guidance to the model in an end-to-end fashion, resulting in better discriminability and reduced visual confusion. We conduct extensive experiments on various image classification benchmarks with our proposed framework and demonstrate its efficacy by means of improved classification accuracy including CIFAR-100(+3.46%), Caltech-256 (+1.64%), ImageNet (+0.92%), CUB-200-2011 (+4.8%) and PASCAL VOC2012 (+5.78%)

Second, we observe that the intermediate model attention can bridge two different vision tasks. We address this issue by proposing a coupled encoder-decoder network to jointly detect faces and localize facial keypoints. The encoder and decoder generate attention maps for facial landmark localization, while the intermediate feature maps attend to the facial regions, which motivates us to build a unified framework by coupling the attention features for multi-scale cascaded face detection. Experiments on face detection show strongly competitive results against the existing methods on two public benchmarks. The landmark localization further shows consistently better accuracy than state-of-the-art on three face-in-the-wild databases.