VISUAL EXPLANATIONS FOR DEEP NEURAL NETWORKS

Prateek Garg, Siddhant Midha
Department of Electrical Engineering, IIT Bombay
{prateekgarg,siddhantm}@ee.iitb.ac.in

Abstract

With the evolution of Machine Learning and Deep Learning, we have come up with more and more complex networks which perform very well on various tasks. Through this development, one caveat remains. Increasing complexity of the network leads to decreasing interpretability. In this paper, we review methods such as .. and implement them.

Index Terms- Convolutional Neural Networks, Saliency Maps, Occlusion Senstivity Map, Class Activation Mapping

1. Introduction

There are some prediction tasks where the output labels are discrete and are periodic. For example, consider the problem of pose estimation. Although pose can be a continuous variable, in practice, it is often discretized *e.g.*, in 5-degree intervals. Because of the periodic nature of pose, a 355-degree label is closer to 0-degree label than the 10-degree label. Thus it is important to consider the periodic and discrete nature of the pose classification problem.

2. Methodology

We consider learning a pose es

2.1. Saliency Maps

The one-hot encoding

2.2. Occlusion Senstivity Maps

The outlier noise exists in most of da

2.3. Class Activation Maps

With the conservative target label

3. Experiments

In this section, we show the implementation details and experimental results on Saliency Maps, Occlusion Senstivity Map, Class Activation Mapping.

3.1. Head pose

Following [1, 4], we choose the occluded version

3.2. Pedestrians orientation

The TUD multi-view pedestrians

3.3. 3D object pose

ddsdsdsd

4. Conclusions

We have introduced a simple yet efficient loss function for pose estimation, based on the Wasserstein distance. Its ground metric represents the class correlation and can be predefined using an increasing function of arc length or learned by alternative optimization. Both the outlier and inlier noise in pose data are incorporated in a unimodal-uniform mixture distribution to construct the conservative label. We systematically discussed the fast closed-form solutions in one-hot and conservative label cases. The results show that the best performance can be achieved by choosing convex function, Binomial distribution for smoothing and solving its exact solution. Although it was originally developed for pose estimation, it is essentially applicable to other problems with discrete and periodic labels. In the future, we plan to develop a more stable adaptive ground metric learning scheme for more classes, and adjust the shape of conservative target distribution automatically.

5. Acknowledgement

The funding support from Youth Innovation Promotion Association, CAS (2017264), Innovative Foundation of CIOMP, CAS (Y586320150) and Hong Kong Government General Research Fund GRF (Ref. No.152202/14E) are greatly appreciated.

References

[1] Lucas Beyer, Alexander Hermans, and Bastian Leibe. Biternion nets: Continuous head pose regression from discrete

training labels. In *German Conference on Pattern Recognition*, pages 157–168. Springer, 2015. 1

[2] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in neural information processing systems*, pages 2292–2300, 2013. 1

[3] Robert Fisher, Jose Santos-Victor, and James Crowley. Caviar: Context aware vision using image-based active recognition, 2005. 1

[4] Sergey Prokudin, Peter Gehler, and Sebastian Nowozin. Deep directional statistics: Pose estimation with uncertainty quantification. *ECCV*, 2018. 1

[5] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 1