

Week 6 paper summary

Sina Mehdinia

05/07/2021

Summary

Zeiler and Fergus [1] introduced some methods to better understand and visualize CNNs. These methods, also result in designing better network architectures. The DeconvNet method is used to reverse convolutions and map the feature maps back to pixels. For pooling layers, they get the locations only since it is irreversible. For filters, they transpose them. With this method, they show feature evolution in each layer and show that how higher layers are more invariant. The next thing, is gaining insight about the architecture. They show that how some early layers in AlexNet [2] were not learning mid-frequency features or a large stride was causing generation of artifacts. They suggest reducing the 1st layer filter size from 11x11 to 7x7 and reducing the stride size to 2. They show that the new architecture learns better features without artifacts, and the classification accuracy is also improved. The next thing they did was sensitivity analysis to occlusion, they occluded various parts of the input image to see how sensitive it is for the correct classification. They showed that the model is truly identifying the location of the object in the image and is highly sensitive to local structures. They also performed an ablation study and showed that the network depth is very important for a good performance. At last, they showed how the ImageNet trained model can generalize well to other datasets.

References

- [1] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014.
- [2] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.