A Review on Deep Learning Techniques Applied to Semantic Segmentation

Yuan An

In the last article, semantic segmentation is introduced in brief. In this article, terminology and background concepts will be introduced [1]. In the section 2, the authors introduce terminology and background concepts of semantic segmentation. Firstly, there is common networks.

Common Deep Network Architectures

Some deep networks have made important contributions to the field, such as AlexNet, VGG-16, GoogLeNet, and ResNet.

AlexNet

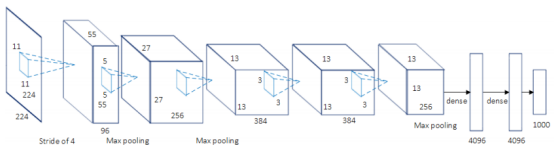

Figure 1. AlexNet Convolutional Neural Network architecture. Figure reproduced from [4]

Figure 1 shows the architecture presented by Krizhevsky *et al.* [4] that consist of 5 convolutional layers, max-pooling ones, Rectified Linear Units (ReLUs), 3 fully-connected layers, and dropout.

VGG

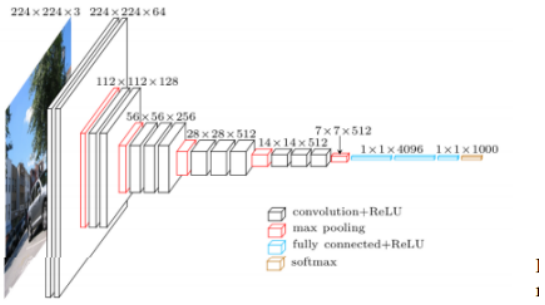

Figure 2. VGG-16 CNN architecture. Figure extracted from Matthieu Cord's talk with his permission.

Visual Geometry Group (VGG) is a CNN model introduced by the Visual Geometry Group from the University of Oxford who proposed various models and configurations of deep CNNs [5]. Figure 2 shows the configuration of VGG-16. The difference between VGG-16 and its predecessors is the use of a stack of convolution layers with small receptive fields in the first layers instead of few layers with big receptive fields.

GoogLeNet

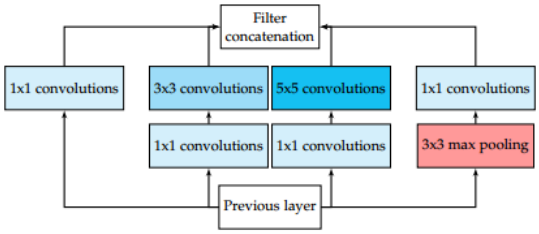

Figure 3. Inception module with dimensionality reduction from the GoogLeNet architecture. Figure reproduced from [6]

GoogLeNet is a network introduced by Szegedy *et al.* [6] which won the ILSVRC-2014 challenge. This CNN architecture is famous of its complexity, emphasized by the fact that it is composed by 22 layers and a newly introduced building block called inception module (see Figure 3).

ResNet

Microsoft's ResNet [3] is famous of high accuracy (won ILSVRC-2016 with 96.4% accuracy) and its depth (152 layers) and the introduce of residual blocks (see Figure 4). The residual blocks solve the problem of training a really deep architecture by introducing identity skip connections so that layers can copy their inputs to the next layer.

ReNet

Graves *et al.* [2] proposed a Multi-dimensional Recurrent Neural Network (MDRNN) architecture for extending Recurrent Neural Networks (RNNs) architectures to mulit-dimensional tasks.

In ReNet, each convolutional layer is replaced with 4 RNNs sweeping the image vertically and horizontally in both directions (see Figure 5).

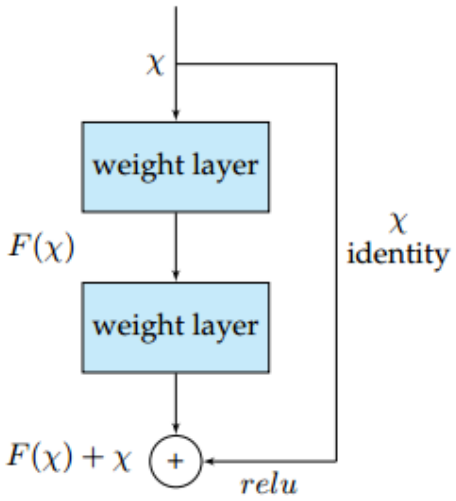

Figure 4. Residual block from the ResNet architecture. Figure reproduced from [3].

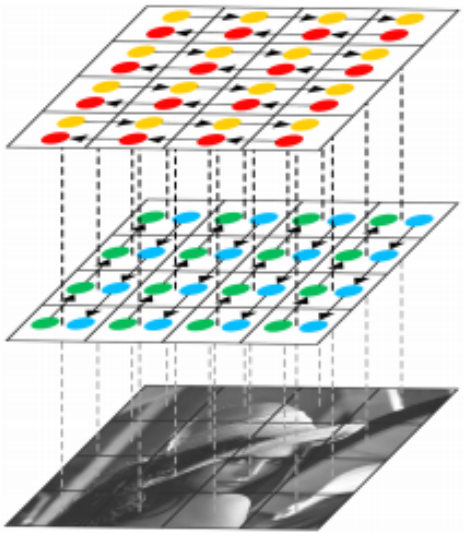

Figure 5. One layer of ReNet architecture modeling vertical and horizontal spatial dependencies. Extracted from [7].

References

[1] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. G. Rodríguez. A review on deep learning techniques applied to semantic segmentation. *arXiv preprint arXiv:1704.06857*, 2017. 1

[2] A. Graves, S. Fernández, and J. Schmidhuber. Multi-dimensional recurrent neural networks. In *ICANN*, pages 549–558, 2007. 1

[3] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016. 1, 2

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. In *NIPS*, pages 1097–1105, 2012. 1

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

[6] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *CVPR*, pages 1–9, 2015. 1

[7] F. Visin, K. Kastner, K. Cho, M. Matteucci, A. C. Courville, and Y. Bengio. ReNet: A recurrent neural network based alternative to convolutional networks. *arXiv preprint arXiv:1505.00393*, 2015. 2