Survey on Convolutional Neural Networks for Image Semantic Segmentation

COP500 Research Methods Presentation

Zhihao DAI[1] Second Author[1]

Loughborough University[1]

COP500 Research Methods
6th March, 2020

Loughborough University

Outline

Outline

Convolutional Neural Networks (CNN)

In deep learning, a **Convolutional Neural Network (CNN)** is a class of deep neural networks, most commonly applied to analyzing visual imagery.

A convolutional neural network consists of an input and an output layer, as well as multiple hidden layers, they are

- Convolutional Layer
- Pooling Layer
- Fully-connected Layer

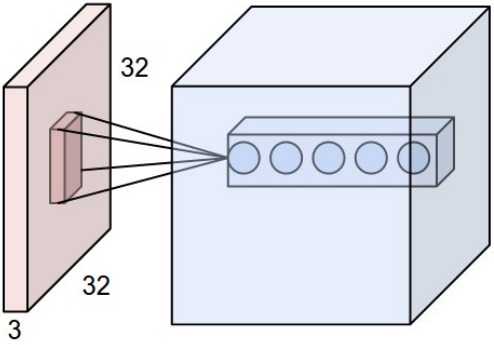

Figure: An example of convolutional layer [1].

[1] http://cs231n.github.io/convolutional-networks/

Solutions and Performance

Table: Summary of Solutions to Challenges and Performance on PASCAL VOC 2012 Challenge (Everingham et al. 2010) within Each Method Reviewed.

Challenges	Reduced Features Resolution	Global Context	Limited Receptive Fields	Spatial Invariance	Boundary Recovery	VOC 2012 mean IU
FCN	Upsampling through bilinear interpolation	×	Aggregation of low-level features	×	Aggregation of low-level features	67.2
PSPNet	Upsampling through deconvolution	Global average pooling	Region-based pooling	×	×	85.4
U-Net	Upsampling through deconvolution	×	×	×	Encoder-decoder	×
DeepLab v2	ASPP	×	ASPP	CRF	×	79.7
DeepLab v3	ASPP	Global average pooling	ASPP	×	×	86.9
DeepLab v3+	ASPP	Global average pooling	ASPP	×	Encoder-decoder	89.0

Solutions and Performance

Table: Summary of Solutions to Challenges and Performance on PASCAL VOC 2012 Challenge (Everingham et al. 2010) within Each Method Reviewed.

Challenges	Reduced Features Resolution	Global Context	Limited Receptive Fields	Spatial Invariance	Boundary Recovery	VOC 2012 mean IU
FCN	Upsampling through bilinear interpolation	×	Aggregation of low-level features	×	Aggregation of low-level features	67.2
PSPNet	Upsampling through deconvolution	Global average pooling	Region-based pooling	×	×	85.4
U-Net	Upsampling through deconvolution	×	×	×	Encoder-decoder	×
DeepLab v2	ASPP	×	ASPP	CRF	×	79.7
DeepLab v3	ASPP	Global average pooling	ASPP	×	×	86.9
DeepLab v3+	ASPP	Global average pooling	ASPP	×	Encoder-decoder	89.0

1. DeepLab aggregates great ideas from other methods along its way.

Solutions and Performance

Table: Summary of Solutions to Challenges and Performance on PASCAL VOC 2012 Challenge (Everingham et al. 2010) within Each Method Reviewed.

Challenges	Reduced Features Resolution	Global Context	Limited Receptive Fields	Spatial Invariance	Boundary Recovery	VOC 2012 mean IU
FCN	Upsampling through bilinear interpolation	×	Aggregation of low-level features	×	Aggregation of low-level features	67.2
PSPNet	Upsampling through deconvolution	Global average pooling	Region-based pooling	×	×	85.4
U-Net	Upsampling through deconvolution	×	×	×	Encoder-decoder	×
DeepLab v2	ASPP	×	ASPP	CRF	×	79.7
DeepLab v3	ASPP	Global average pooling	ASPP	×	×	86.9
DeepLab v3+	ASPP	Global average pooling	ASPP	×	Encoder-decoder	89.0

1. DeepLab aggregates great ideas from other methods along its way.
2. DeepLab v3+ achieves the best performance.

Solutions and Performance

Table: Summary of Solutions to Challenges and Performance on PASCAL VOC 2012 Challenge (Everingham et al. 2010) within Each Method Reviewed.

Challenges	Reduced Features Resolution	Global Context	Limited Receptive Fields	Spatial Invariance	Boundary Recovery	VOC 2012 mean IU
FCN	Upsampling through bilinear interpolation	×	Aggregation of low-level features	×	Aggregation of low-level features	67.2
PSPNet	Upsampling through deconvolution	Global average pooling	Region-based pooling	×	×	85.4
U-Net	Upsampling through deconvolution	×	×	×	Encoder-decoder	×
DeepLab v2	ASPP	×	ASPP	CRF	×	79.7
DeepLab v3	ASPP	Global average pooling	ASPP	×	×	86.9
DeepLab v3+	ASPP	Global average pooling	ASPP	×	Encoder-decoder	89.0

1. DeepLab aggregates great ideas from other methods along its way.
2. DeepLab v3+ achieves the best performance.
3. The more challenges are addressed, the higher performance is.

References I

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J. & Zisserman, A. (2010), 'The Pascal Visual Object Classes (VOC) Challenge', *International Journal of Computer Vision* **88**(2), 303–338.
URL: *http://link.springer.com/10.1007/s11263-009-0275-4*

Thank you!