The Receptive Field in CNNs

Jonathan Kobold

5.3.2020

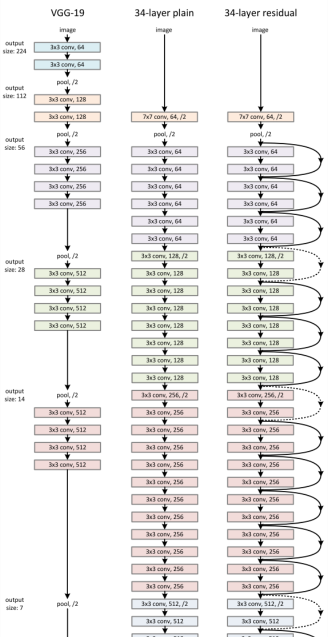
Laboratoire IBISC
Université Paris Saclay, Université Evry Val d'Essonne

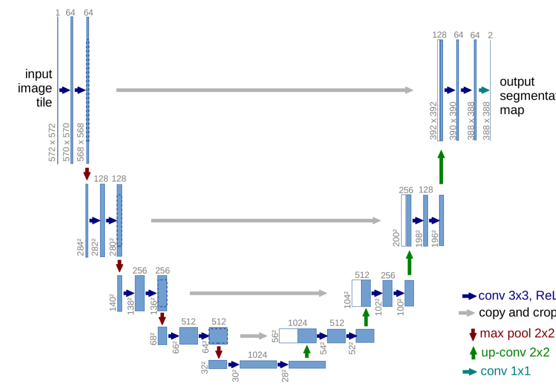
CNNs

Convolutional neural networks (CNNs) are currently the number one tool for pattern recognition

- Spatial patterns
  - Object detection from images
  - Semantic segmentation
  - Image classification
- Sound patterns
  - Natural Language Processing (NLP)
  - Music recommender systems
  - Automatic subtitles
  - Voice control
- Temporal patterns
  - Process control
  - Automatic translation

CNNs





Left: Resnet (Kaiming et al. 2015)

Right: U-Net (Ronneberger et al. 2015)

Operations in a CNN

- Convolutions
  - Standard Convolutions
  - Dilated Convolutions
  - Strided Convolutions
- Pooling Operations
  - Maximum Pooling
  - Average Pooling
- Regularisation Operations
  - Dropout
  - Batch Normalisation
- Activation Functions

Discrete Convolution

Kernel size: 3   Dilation rate: 1   Stride: 1

$$y_3 = x_3 w_1 + x_4 w_2 + x_5 w_3$$

Discrete Convolution

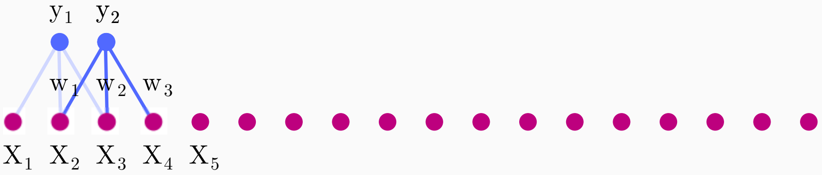
Kernel size: 3 Dilation rate: 1 Stride: 1

$$y_1 = x_1 w_1 + x_2 w_2 + x_3 w_3$$

Discrete Convolution

Kernel size: 3 Dilation rate: 1 Stride: 1

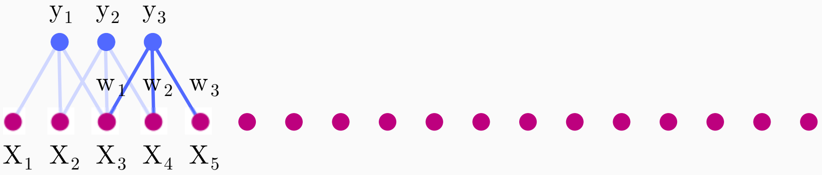
$$y_2 = x_2 w_1 + x_3 w_2 + x_4 w_3$$



Discrete Convolution

Kernel size: 3 Dilation rate: 1 Stride: 1

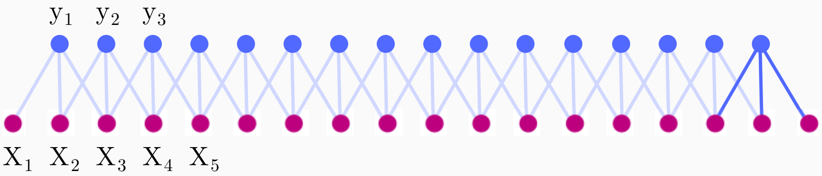
$$y_3 = x_3 w_1 + x_4 w_2 + x_5 w_3$$



Discrete Convolution

Kernel size: 3 Dilation rate: 1 Stride: 1

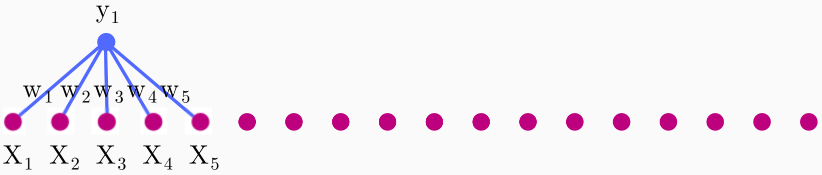
$$y_3 = x_3 w_1 + x_4 w_2 + x_5 w_3$$



Discrete Convolution

Kernel size: **5** Dilation rate: 1 Stride: 1

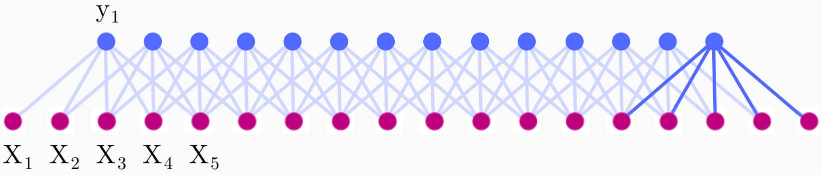
$$y_1 = x_1 w_1 + x_2 w_2 + x_3 w_3 + x_4 w_4 + x_5 w_5$$



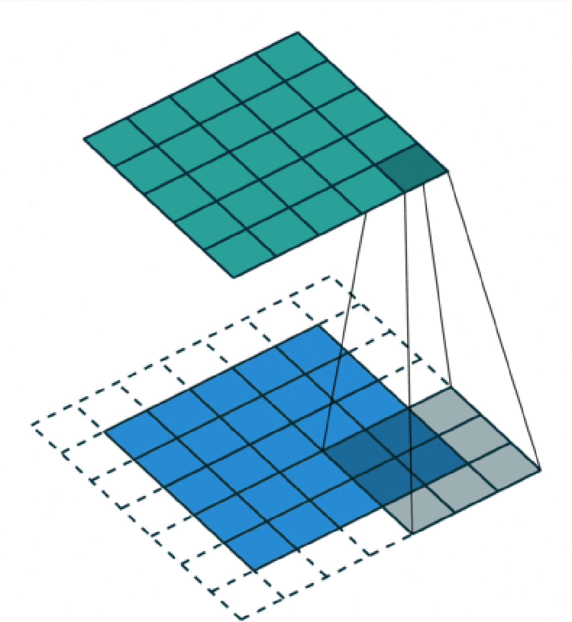
Discrete Convolution

Kernel size: 5 Dilation rate: 1 Stride: 1

$$y_1 = x_1 w_1 + x_2 w_2 + x_3 w_3 + x_4 w_4 + x_5 w_5$$



Convolution



Dilated Convolution

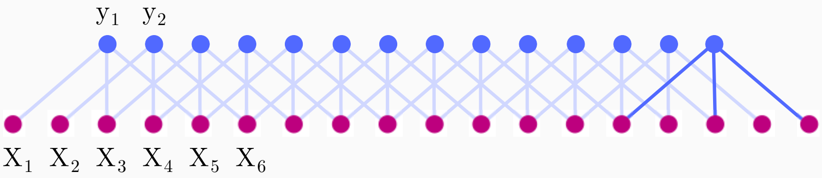
Kernel size: 3 Dilation rate: **2** Stride: 1

$$y_1 = x_1 w_1 + x_3 w_2 + x_5 w_3$$

Dilated Convolution

Kernel size: 3 Dilation rate: 2 Stride: 1

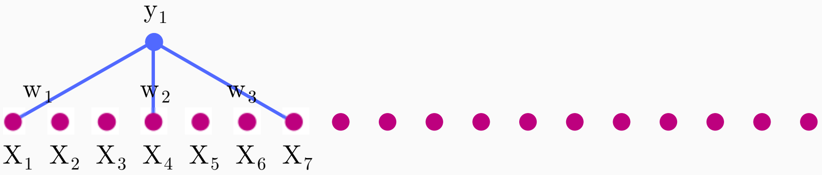
$$y_1 = x_1 w_1 + x_3 w_2 + x_5 w_3$$



Dilated Convolution

Kernel size: 3 Dilation rate: **3** Stride: 1

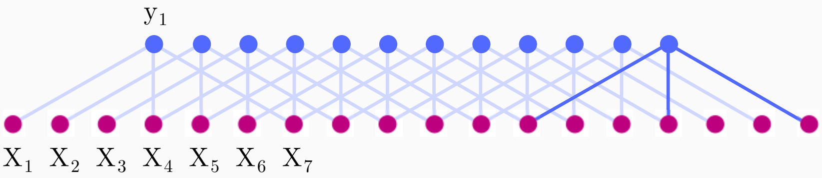
$$y_1 = x_1 w_1 + x_4 w_2 + x_7 w_3$$



Dilated Convolution

Kernel size: 3 Dilation rate: 3 Stride: 1

$$y_1 = x_1 w_1 + x_3 w_2 + x_5 w_3$$



Maximum Pooling

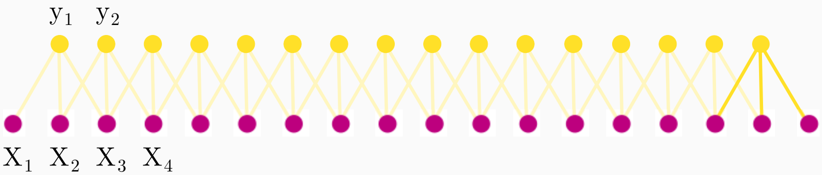
Kernel size: 3 Dilation rate: 1 Stride: 1

$$y_1 = max\{x_1, x_2, x_3\}$$

Maximum Pooling

Kernel size: 3 Dilation rate: 1 Stride: 1

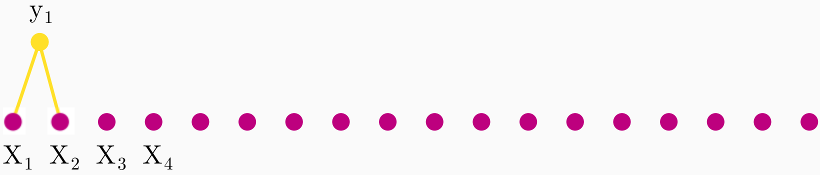
$$y_3 = x_3 w_1 + x_4 w_2 + x_5 w_3$$



Maximum Pooling

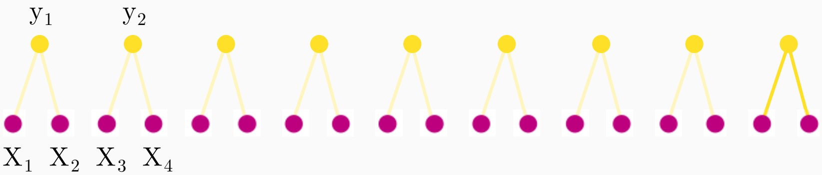
Kernel size: 2 Dilation rate: 1 Stride: 2

$$y_1 = max\{x_1, x_2, x_3\}$$



Maximum Pooling

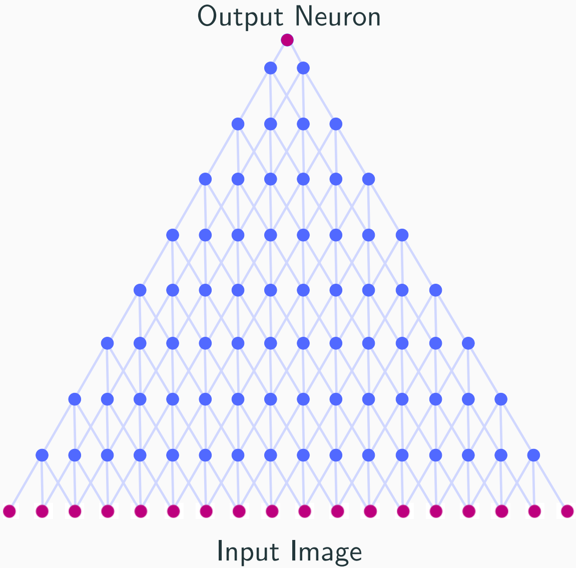
Kernel size: 2 Dilation rate: 1 Stride: 2



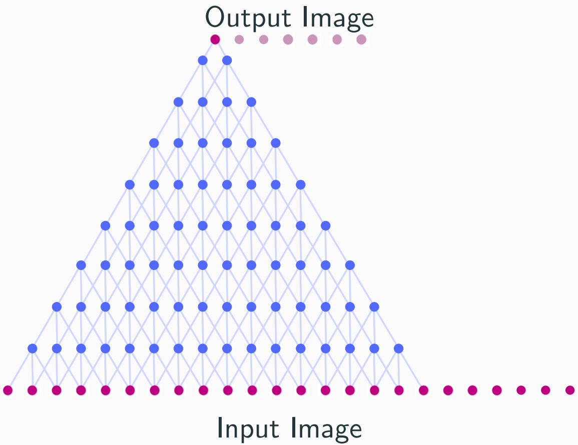
Operations in a CNN

- Sliding windows which perform:
    - Weighted Sum
    - Maximum
    - Average
- Point-wise Operations
    - Dropout
    - Batch Normalisation
    - Activation Functions

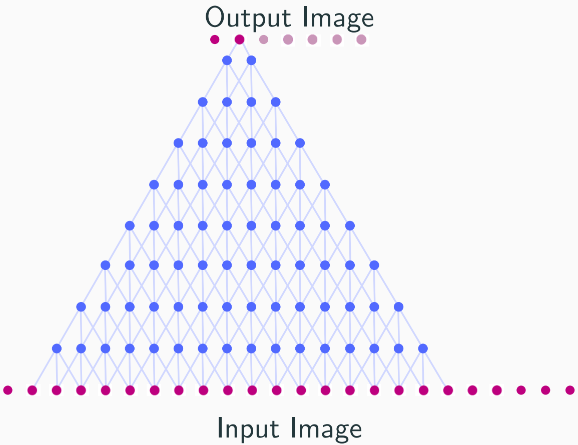
A simple CNN



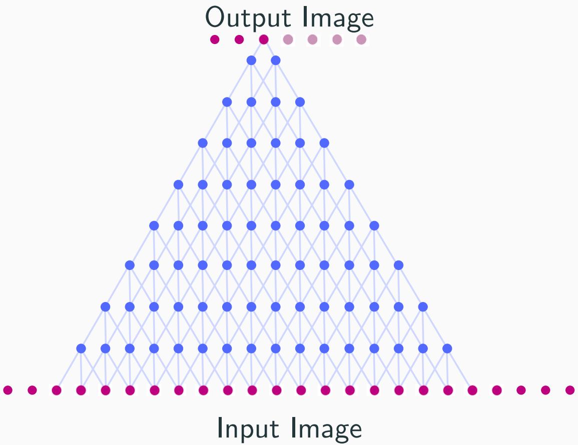
A simple CNN



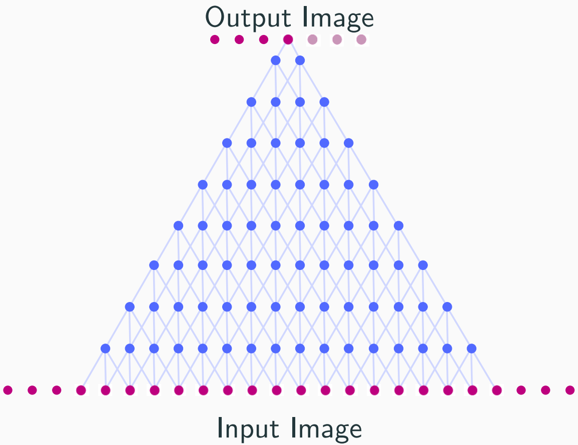
A simple CNN



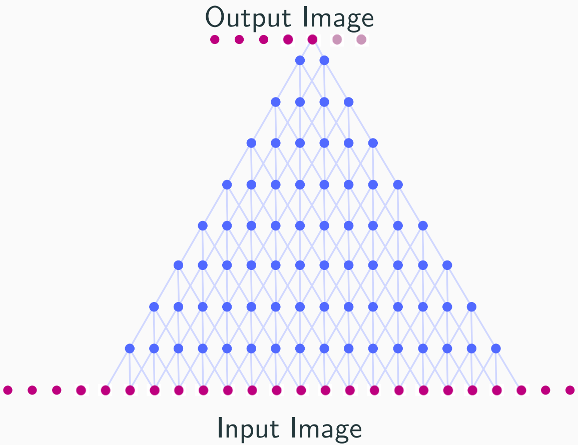
A simple CNN



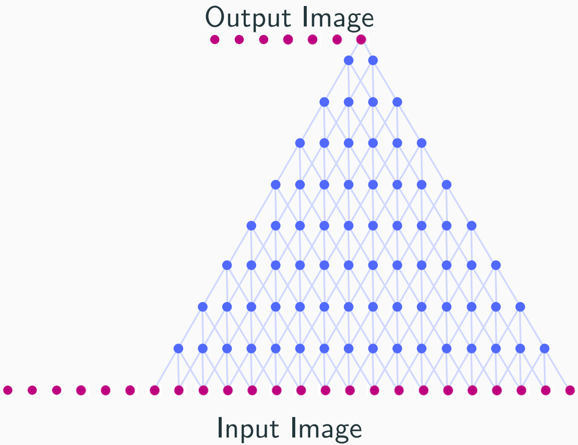
A simple CNN



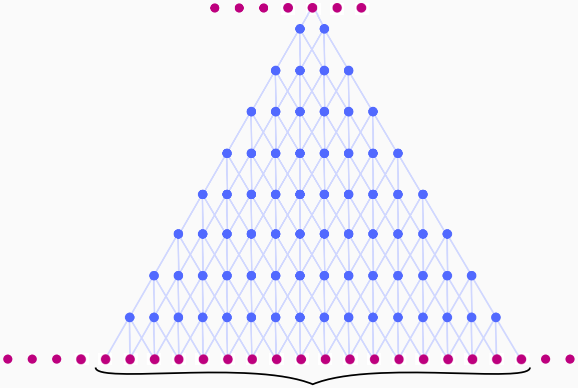
A simple CNN



A simple CNN



Receptive Field



Receptive Field

Receptive Field

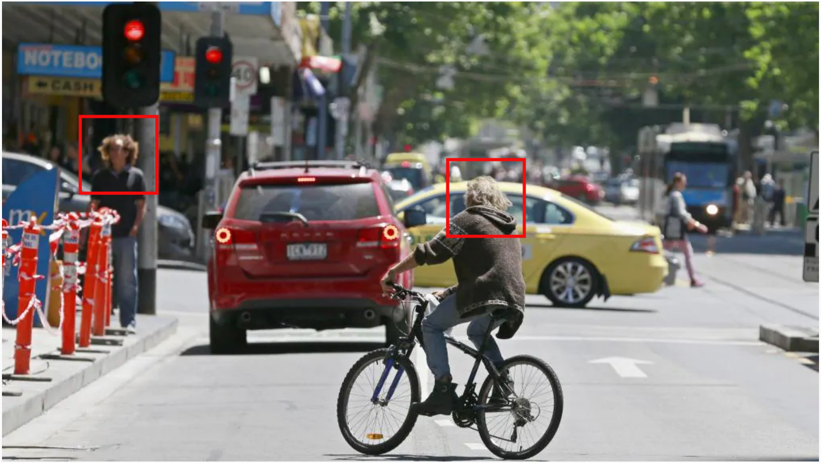
The **receptive field** is the number of pixels in the input which impact the value of the output of a CNN.

Why is the Receptive Field important?

Example: Cyclist and pedestrian detection for autonomous driving

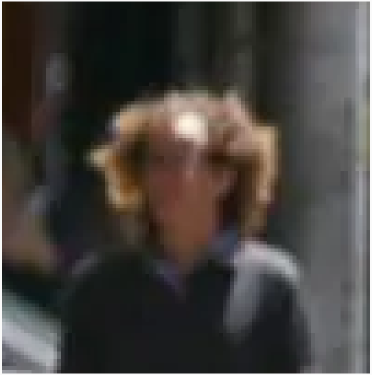
Why is the Receptive Field important?

Receptive Field: 100×100 pixels



Why is the Receptive Field important?

Cyclist and pedestrian are not distinguishable with a 100×100 receptive field. A detection of heads is possible!

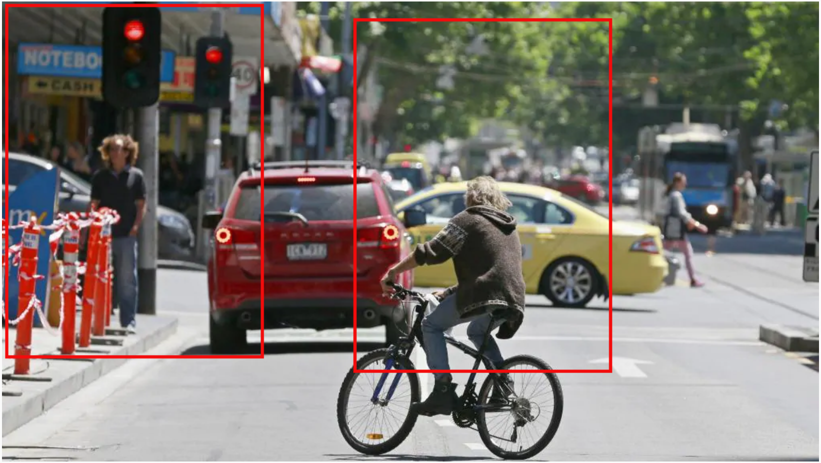


Pedestrian

Cyclist

Why is the Receptive Field important?

Receptive Field: 324×444 pixels



Why is the Receptive Field important?

Cyclist and pedestrian are distinguishable with a 324×444 receptive field!

Pedestrian

Cyclist

Receptive Field

Considerations for choosing the receptive field size:

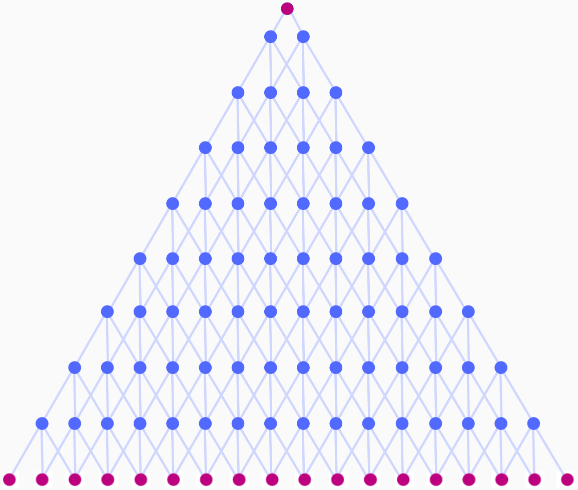
- Decides the size of the input's context available to the model
- Too small receptive fields make detection tasks impossible
- Too large receptive fields may lead to target objects disappearing because they are too small

$\rightarrow$ The right size depends on the application

Choose the receptive field in the same order of magnitude as the objects to detect! Within this range larger receptive field sizes lead to larger models.

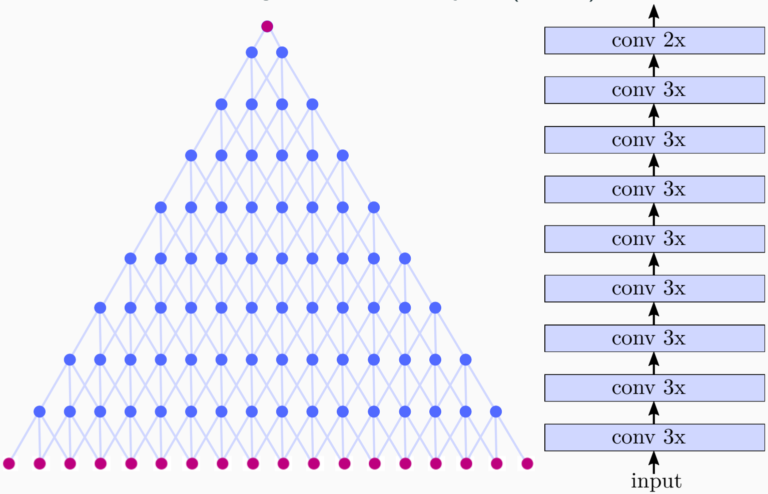
Methods for Extending the Receptive Field

Stacking convolution layers (Stack)

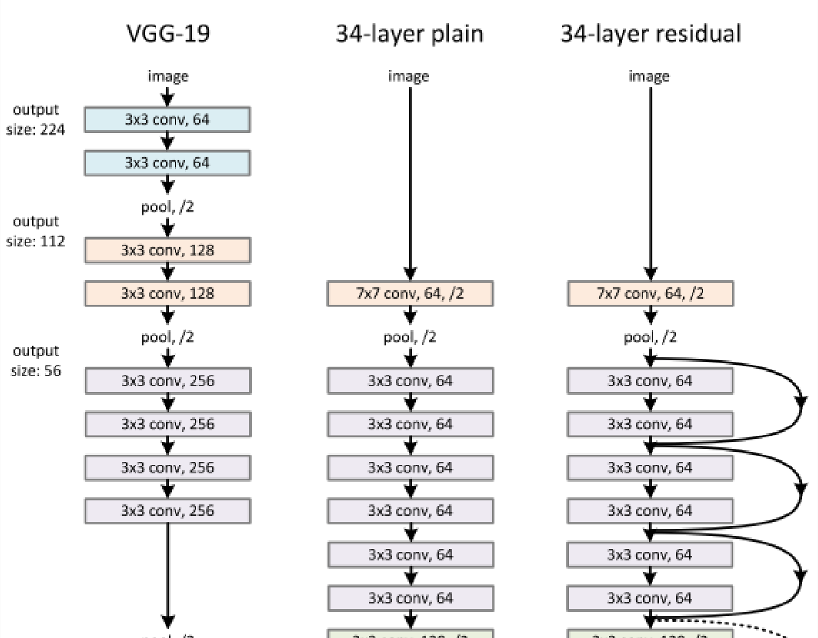


Methods for Extending the Receptive Field

Stacking convolution layers (Stack)

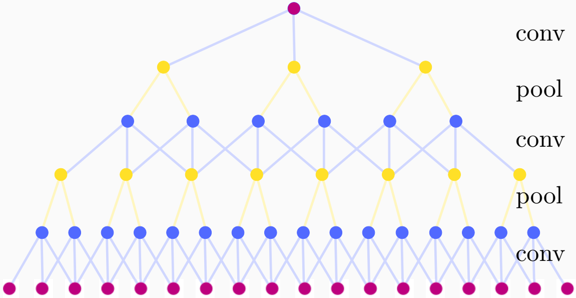


Resnet (Kaiming et al. 2015)



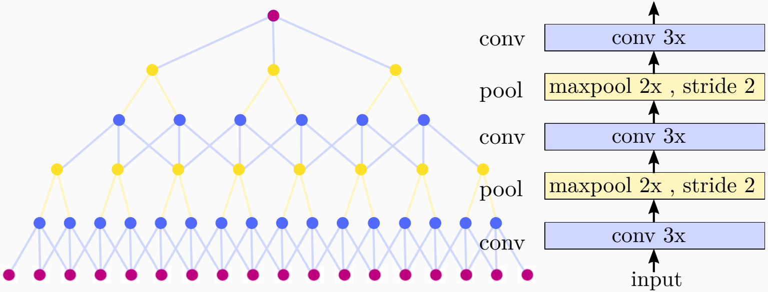
Methods for Extending the Receptive Field

Alternating convolution and maxpool layers (Pyramid)



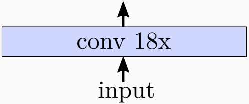
Methods for Extending the Receptive Field

Alternating convolution and maxpool layers (Pyramid)

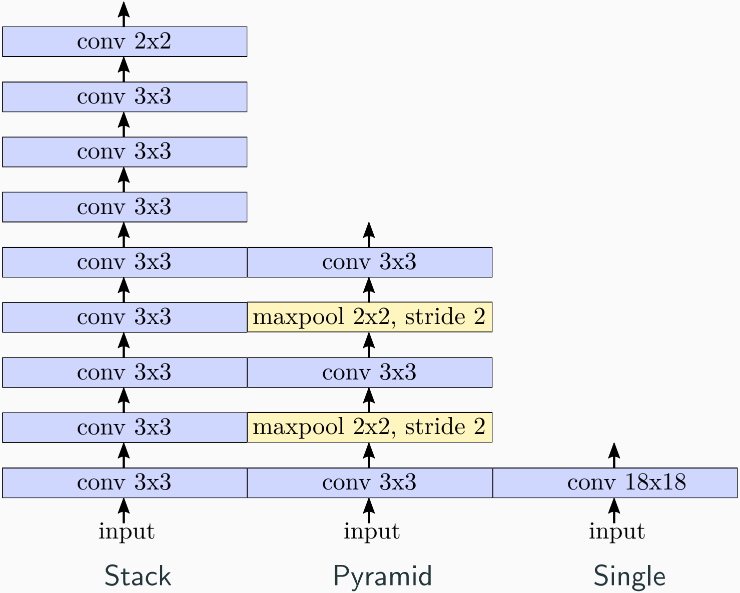


Methods for Extending the Receptive Field

A single convolution with large kernel (Single)



Methods for Extending the Receptive Field



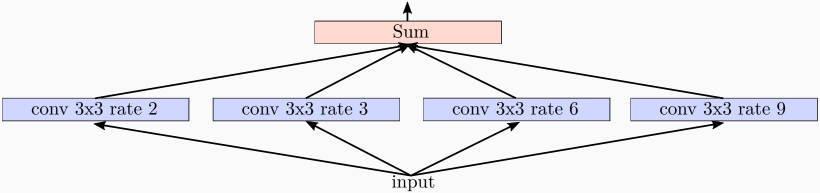
Methods for Extending the Receptive Field

Receptive field: 18×18

Method	Stack	Pyramid	Single
Parameters	76	27	324

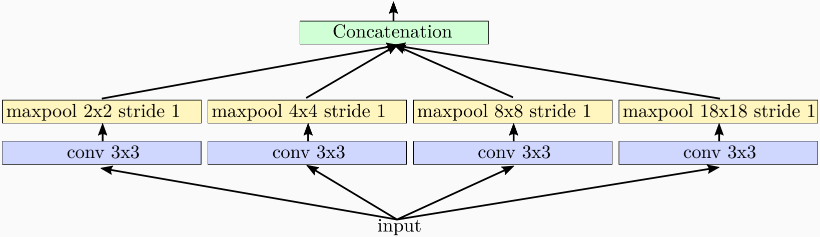
Methods for Extending the Receptive Field

A trous spatial pyramid pooling (ASPP) (Chen et al. 2017)



Methods for Extending the Receptive Field

Transfer block (Transfer) (Kobold 2019)



Methods for Extending the Receptive Field

Sequential methods:

- Stacks
- Pyramids

$\rightarrow$ Rigid structures, no tuning possible

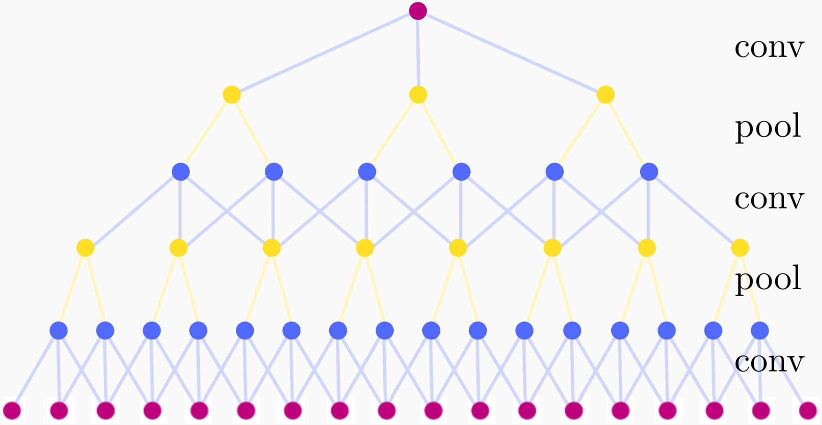
Parallel methods:

- ASPP
- Transfer Block

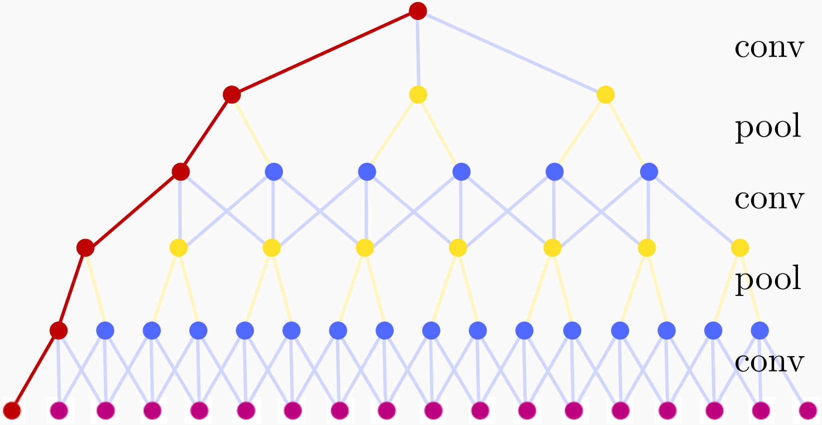
$\rightarrow$ Tunable, number of parameters can be adapted to the actual problem

$\rightarrow$ Scale well with parallel (GPU) computing architectures

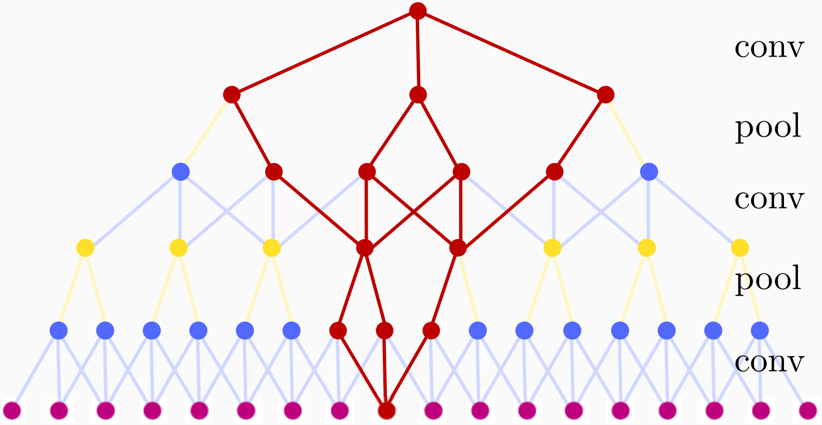
Effective Receptive Field



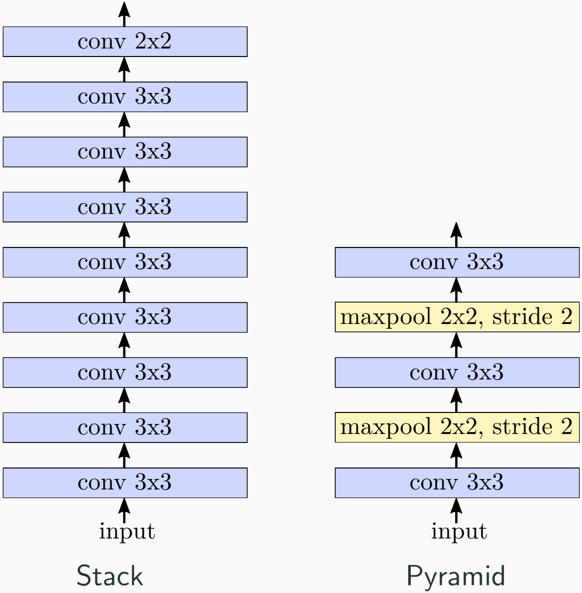
Effective Receptive Field



Effective Receptive Field

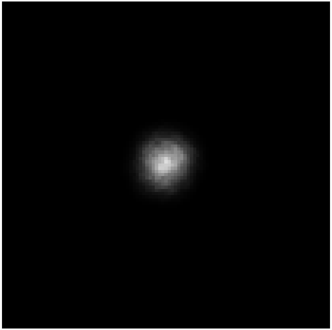


Effective Receptive Field

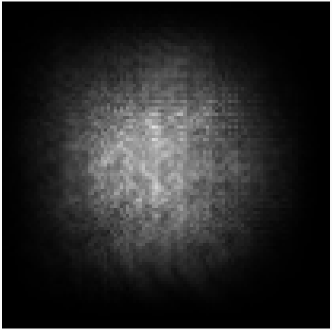


Effective Receptive Field

Effective receptive fields gained from simulations. Networks with 20 layers (Luo et al. 2016)



Stack



Pyramid

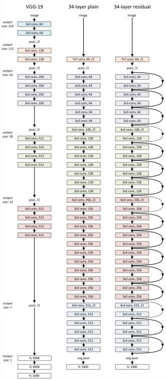
Effective Receptive Field Example: Mushroom Classification

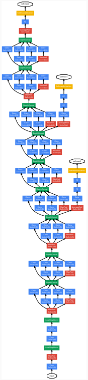

Original


Stack


Pyramid

Effective Receptive Field





Resnet (Kaiming et al. 2015) Inception (Szegedy et al. 2014)

Effective Receptive Field

The effective receptive field (EFR) describes the impact of the input pixels on the output.

$$ERF_i = \frac{\partial \text{Output}}{\partial x_i}$$

Effective Receptive Field

The ERF

- follows a Gaussian distribution
- cannot me modified in sequential models
- can have it's standard deviation modified in parallel model
- is a better estimate of a models capacity than the receptive field alone

Summary

- CNNs
- Receptive Field
    - Importance
    - Stack
    - Pyramid
    - Single
    - ASPP
    - Transfer Block
- Effective Receptive Field

Thank You for Your Attention