F-Net Neural Architecture[[1]](#footnote-1)

**input**

**output**

D

C

B

A

Key

Convolutional block (2 Convolutions, 2 batch normalizations, 2 Relu) + MaxPool + Residual

Encoding

Deconvolutional block (2 deconvolutions, 2 batch normalizations, 2 Relu) + Upsampling + Residual

Decoding

|  |  |  |  |
| --- | --- | --- | --- |
|  | **N-filter convolutional block** |  | **N-filter deconvolutional block** |
| 1 | Relu activation | 1 | Relu activation |
| 2 | Conv (filters=N, size=3) | 2 | Deconvolution (filters=N, size=3) |
| 3 | Bach Normalization | 3 | Batch Normalization |
| 4 | Relu activation [[2]](#footnote-2) | 4 | Relu activation |
| 5 | Conv (filters=N, size=3) | 5 | Deconvolution (filters=N, size=3) |
| 6 | Bach Normalization | 6 | Batch Normalization |
| 7 | Maxpooling (size=3, stride=2) | 7 | Upsampling (size=2) |
| 8 | Residual (add output of previous block) | 8 | Residual (add upsampled output of previous block) |

Padding is added for each convolutional and deconvolutional layer such that the layer output has the same size as the layer input. Changes in input sizes are designed to take place at the max-pooling and up-sampling layers.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Example distribution of filters (N = 32, 64, 128, 256) convolutional sizes along different blocks in a 4-level F-Net** | | | | | | | | | |
|  |  | **Encoder blocks** | | | | **Decoder blocks** | | | |
|  | Image size | **A** | **B** | **C** | **D** | **A** | **B** | **C** | **D** |
| Level 1 | 96x96 | 32 |  |  | 32 |  |  | 32 | 32 |
| Level 2 | 48x48 | 64 |  | 64 | 64 |  | 64 | 64 | 64 |
| Level 3 | 24x24 | 128 | 128 | 128 | 128 | 128 | 128 | 128 | 128 |
| Level 4 | 12x12 | 256 | 256 | 256 | 256 | 256 | 256 | 256 | 256 |

1. Inspired by the and Multigrid F-cycle and the U-Net architecture for semantic segmentation [↑](#footnote-ref-1)
2. For the first convolutional block only steps 2-4 are applicable [↑](#footnote-ref-2)