Semantic Segmentation

Introduction

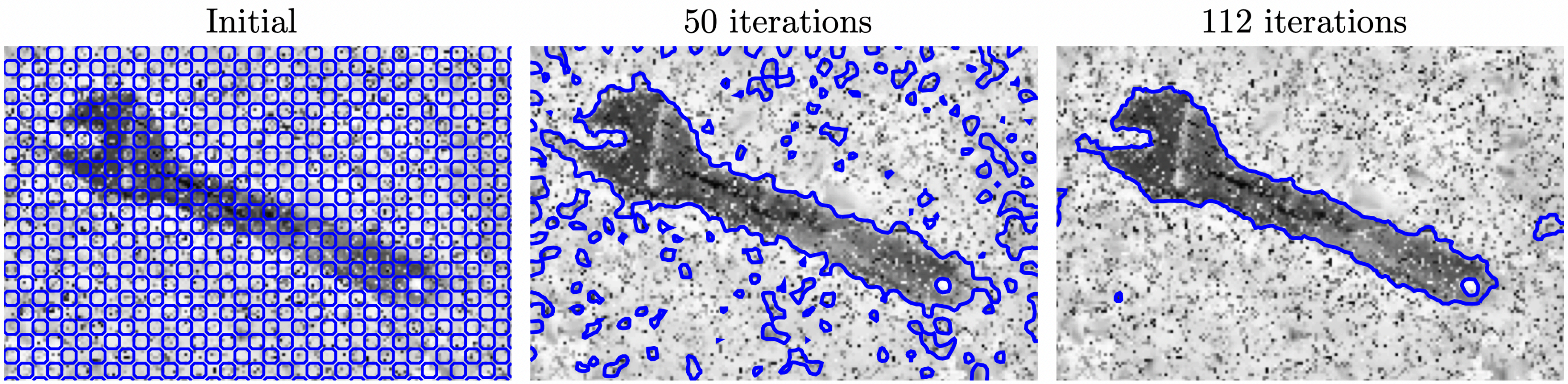
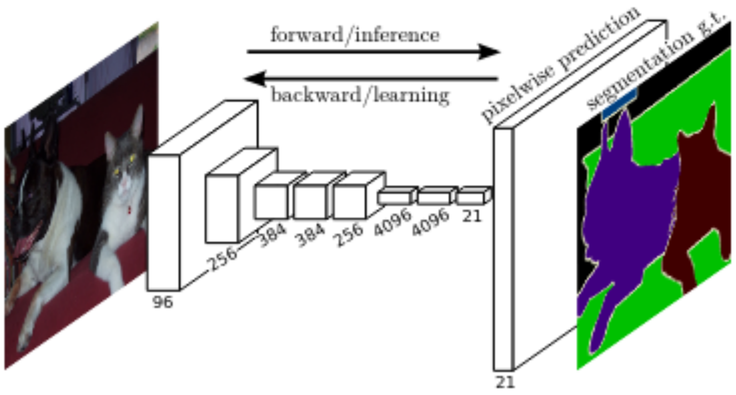


IMAGE PROCESSING ON LINE

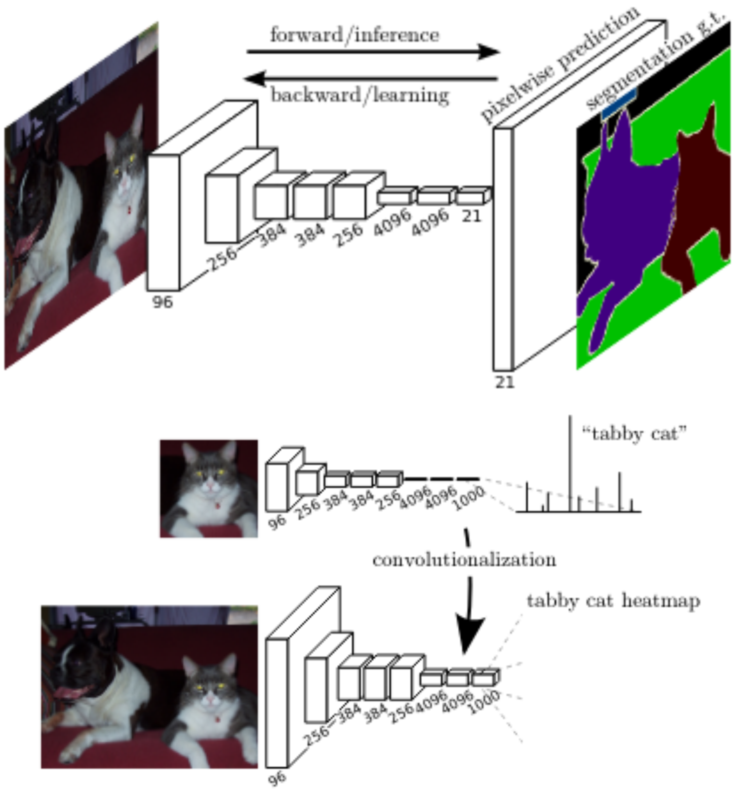
Published in Image Processing On Line on 2012–08–08.
Submitted on 2012–00–00, accepted on 2012–00–00.
ISSN 2105–1232 © 2012 IPOL & the authors CC–BY–NC–SA
This article is available online with supplementary materials, software, datasets and online demo at
http://dx.doi.org/10.5201/ipol.2012.g-cv

Fully Convolutional Networks for Semantic Segmentation



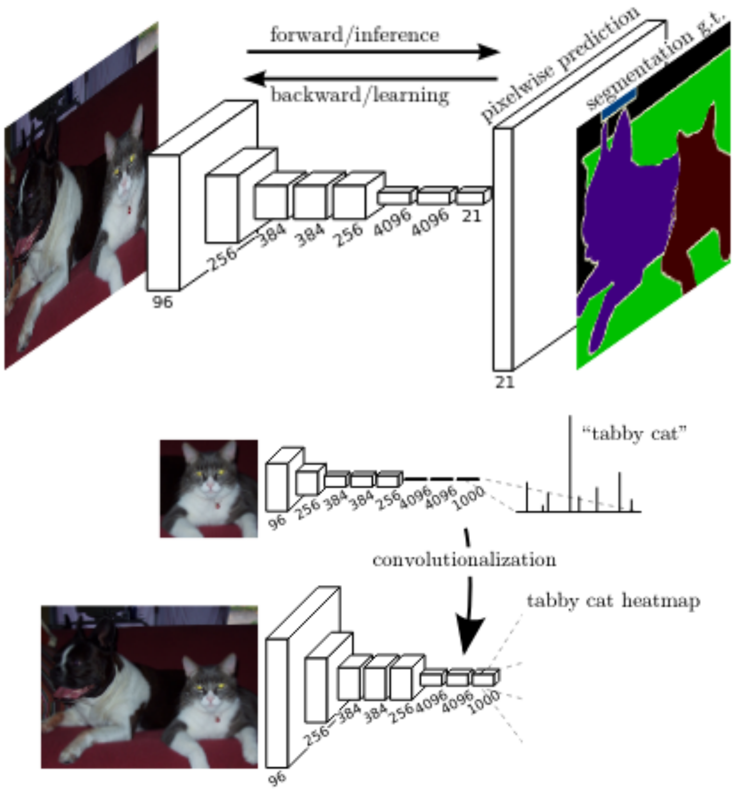
Fully Convolutional Networks for Semantic Segmentation by Jonathan Long, Evan Shelhamer, Trevor Darrell (CVPR 2015)

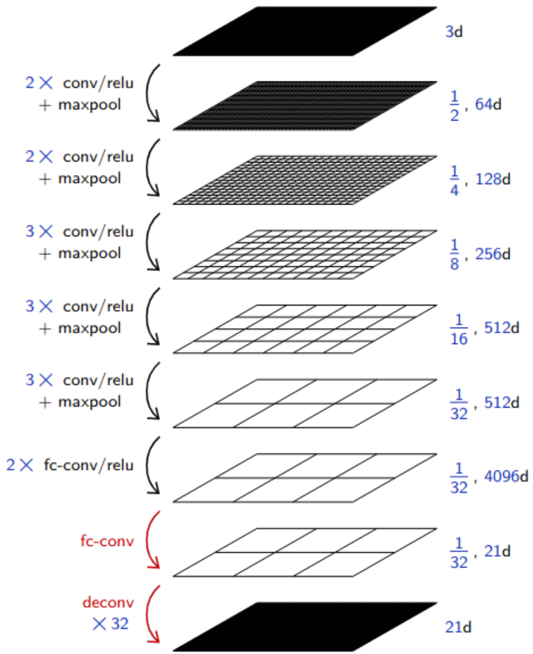
Fully Convolutional Networks for Semantic Segmentation



Fully Convolutional Networks for Semantic Segmentation by Jonathan Long, Evan Shelhamer, Trevor Darrell (CVPR 2015)

Fully Convolutional Networks for Semantic Segmentation



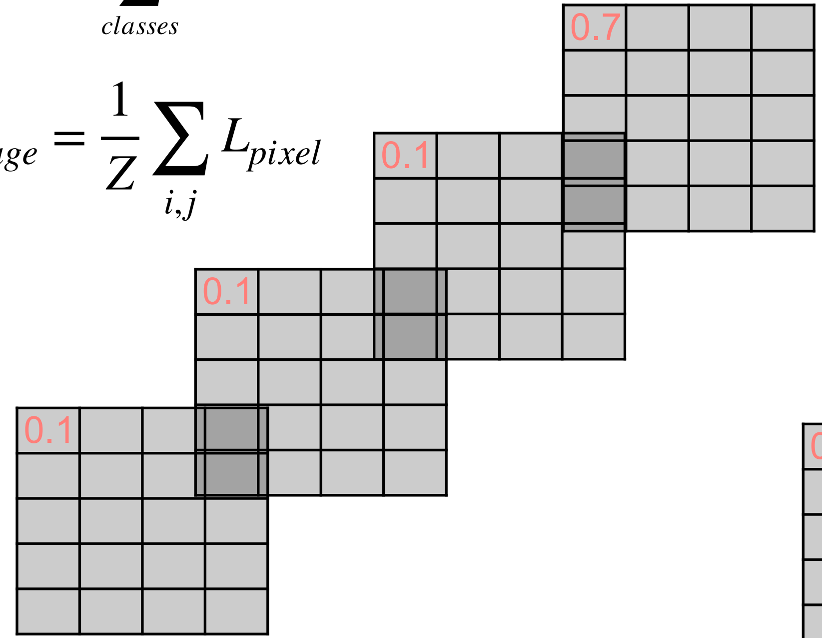


Fully Convolutional Networks for Semantic Segmentation by Jonathan Long, Evan Shelhamer, Trevor Darrell (CVPR 2015)

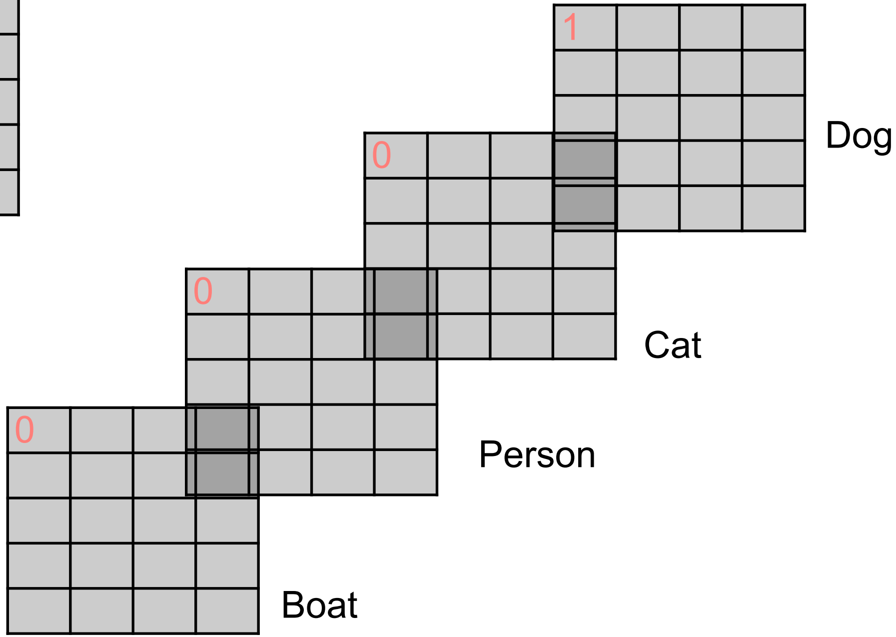
Loss function for semantic

$$L_{pixel} = -\sum_{classes} y log(\hat{y})$$

$$L_{image} = \frac{1}{Z}\sum_{i,j} L_{pixel}$$

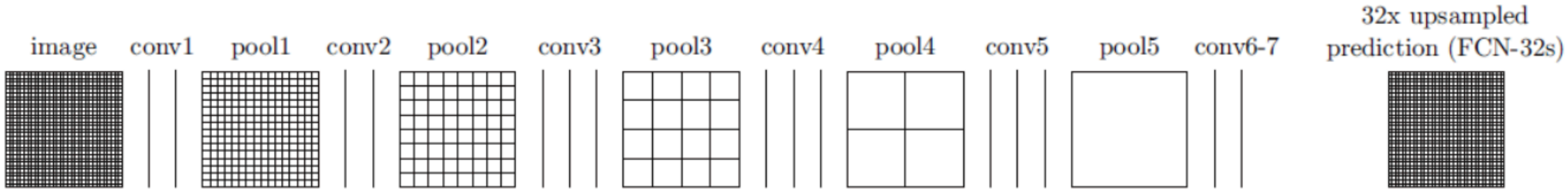


Predicted Pixel wise probabilities

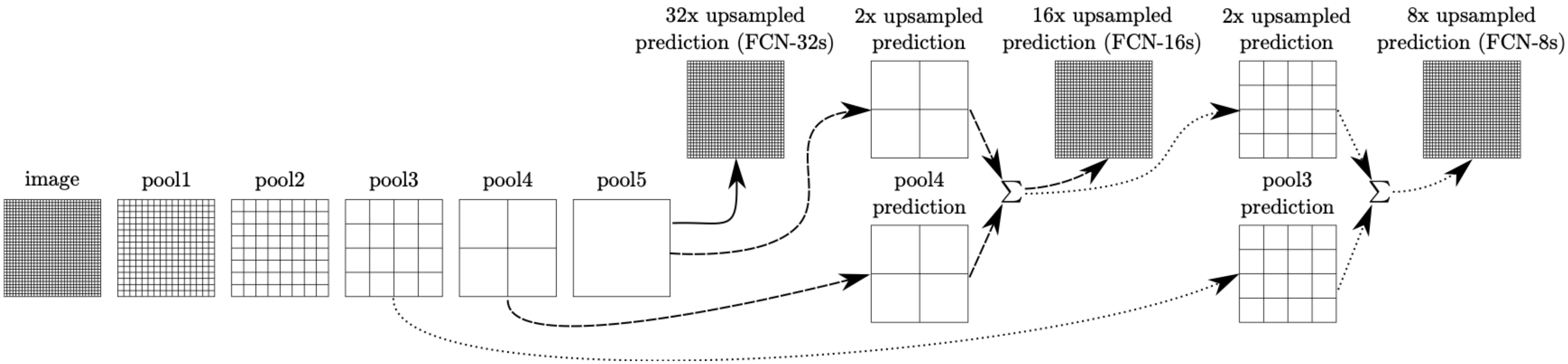


Corresponding pixel targets

Skip connections



FCN-32s



SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation

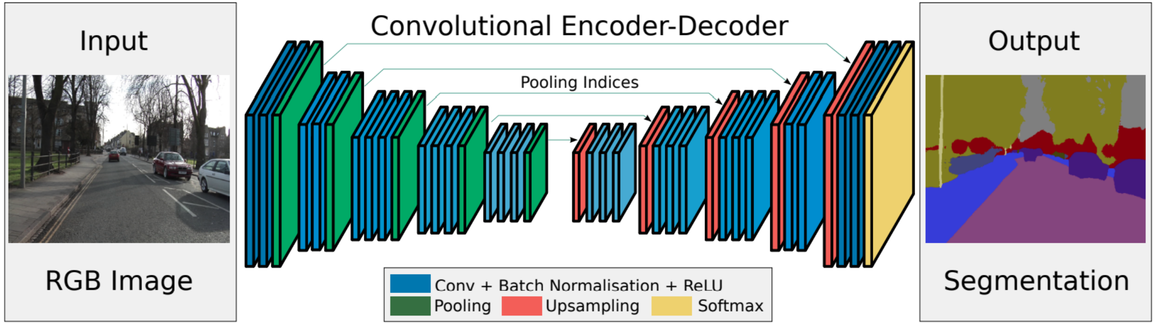


Fig. 2. An illustration of the SegNet architecture. There are no fully connected layers and hence it is only convolutional. A decoder upsamples its input using the transferred pool indices from its encoder to produce a sparse feature map(s). It then performs convolution with a trainable filter bank to densify the feature map. The final decoder output feature maps are fed to a soft-max classifier for pixel-wise classification.

V. Badrinarayanan, A. Kendall and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," in IEEE Transactions on Pattern Analysis and Machine Intelligence,

SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation

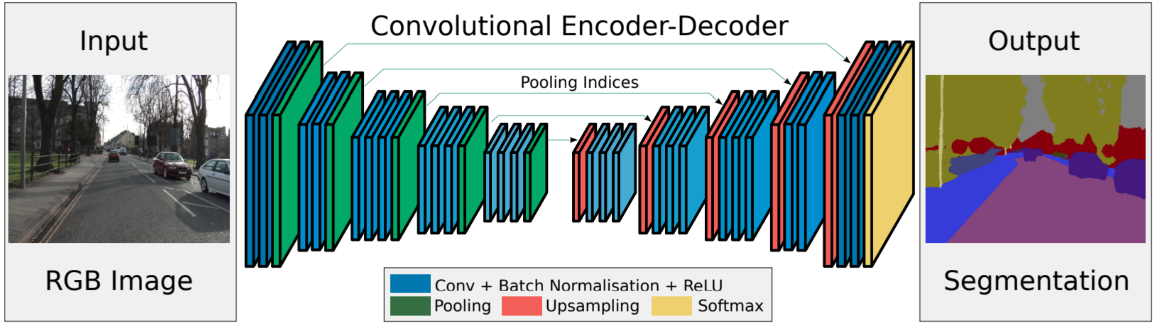
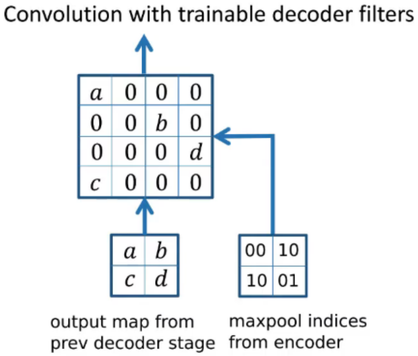


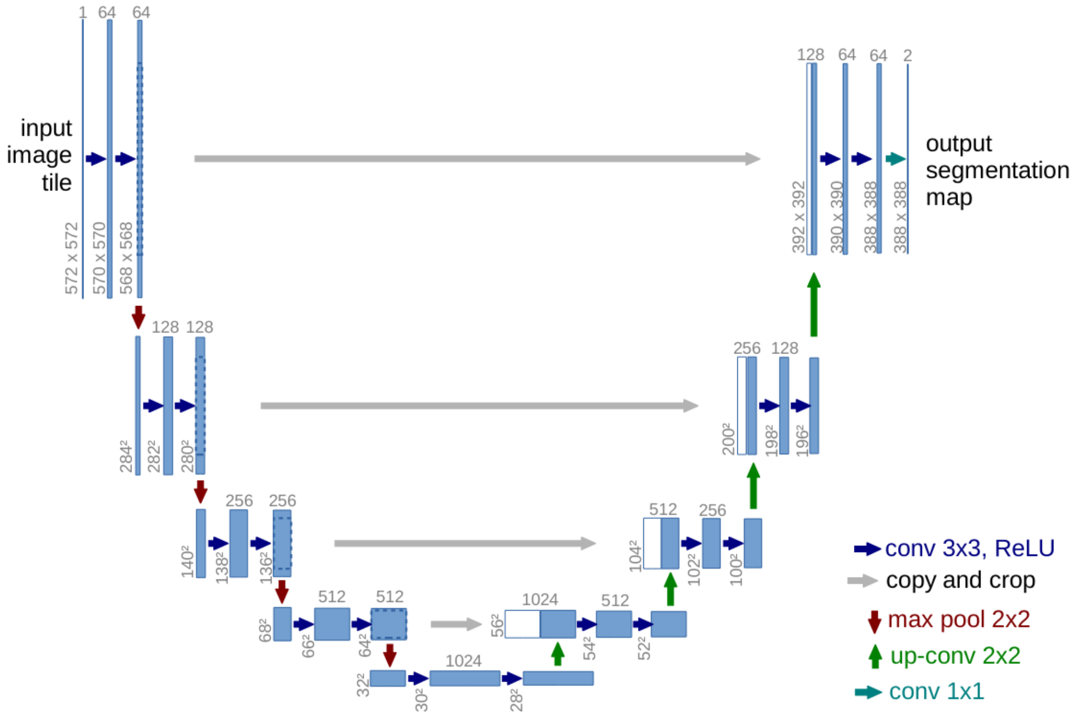
Fig. 2. An illustration of the SegNet architecture. There are no fully connected layers and hence it is only convolutional. A decoder upsamples its input using the transferred pool indices from its encoder to produce a sparse feature map(s). It then performs convolution with a trainable filter bank to densify the feature map. The final decoder output feature maps are fed to a soft-max classifier for pixel-wise classification.

- Encoder - Decoder Based architecture
- For upsampling the dataset, the pooling indices are stored at time of Max-pooling and then are passed again to decoder.



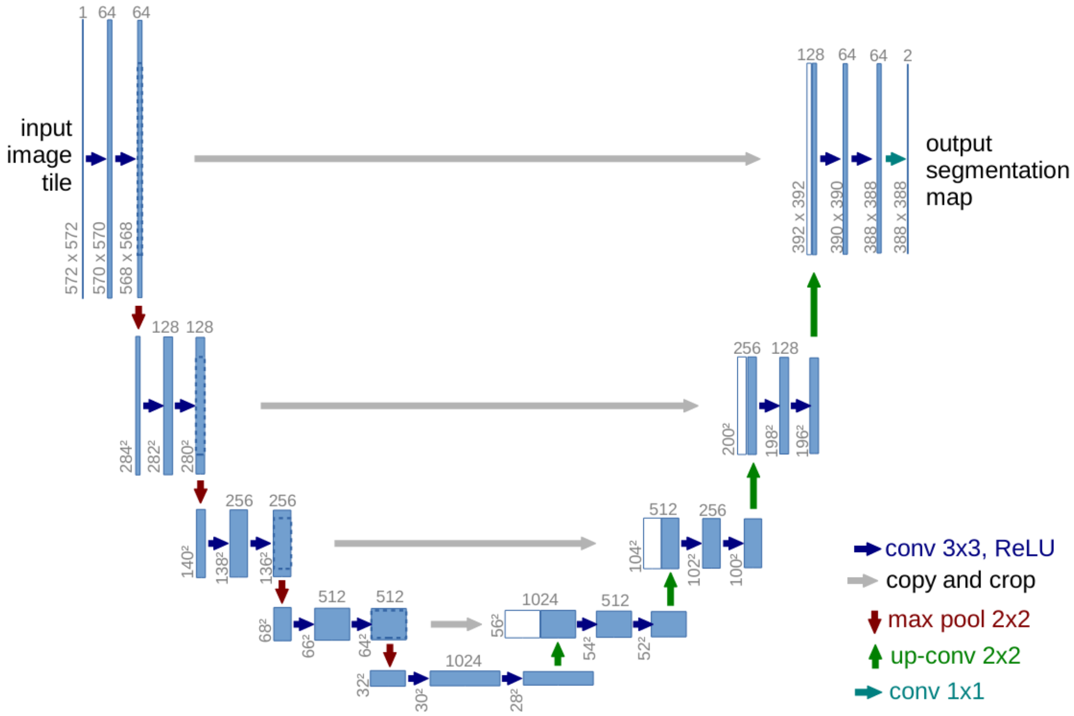
V. Badrinarayanan, A. Kendall and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," in IEEE Transactions on Pattern Analysis and Machine Intelligence,

U-Net: Convolutional Networks for Biomedical Image Segmentation



U-Net: Convolutional Networks for Biomedical Image Segmentation Olaf Ronneberger, Philipp Fischer, Thomas Brox (MICCAI 2015)

U-Net: Convolutional Networks for Biomedical Image Segmentation

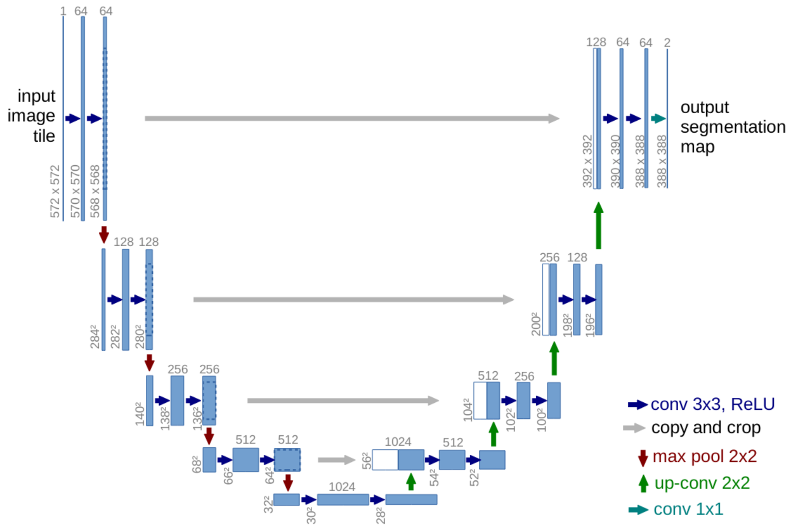


More than 45000 citations

U-Net: Convolutional Networks for Biomedical Image Segmentation Olaf Ronneberger, Philipp Fischer, Thomas Brox (MICCAI 2015)

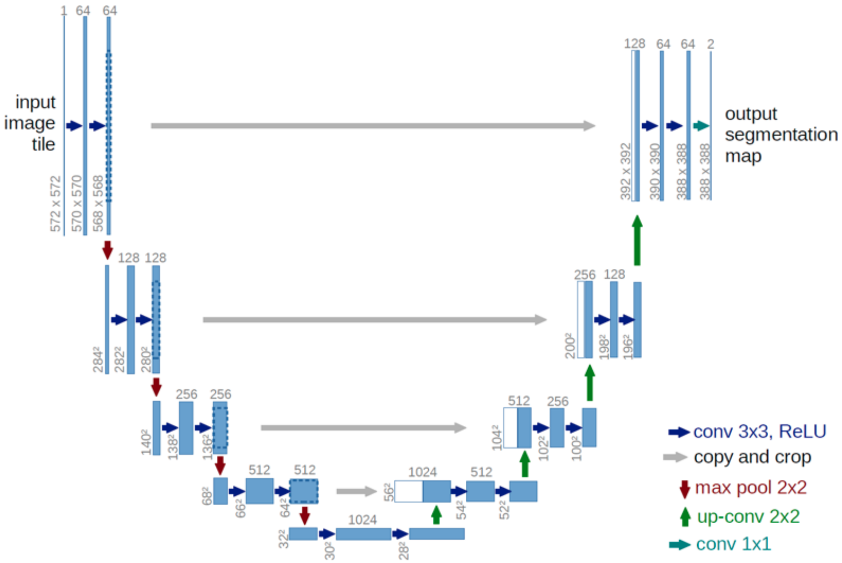
U-Net: Convolutional Networks for Biomedical Image Segmentation

- Fully Convolutional encoder-decoder architecture with skip connections.
- Contracting path is an existing classification network with FC layers removed.
- Upsampling is performed by transposed convolution with s = 2, p= 0, with number of features halved.
- Final layer is 1x1 conv with C channels, C = Number of classes.
- In original architecture, unpadded convolution were performed, making output map smaller than input.

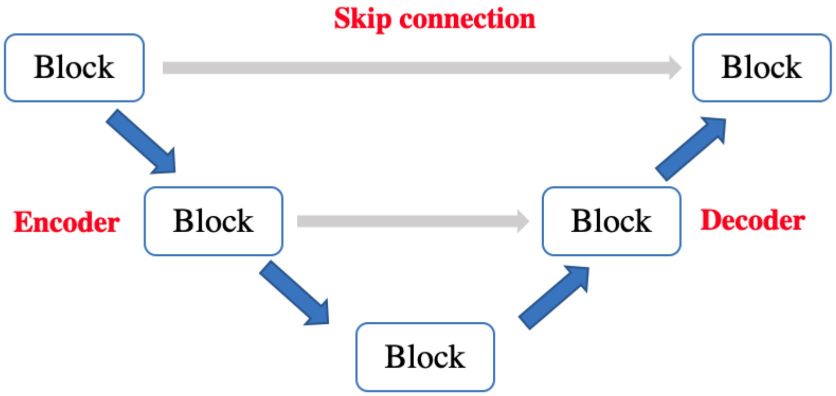


U-Net: Convolutional Networks for Biomedical Image Segmentation Olaf Ronneberger, Philipp Fischer, Thomas Brox (MICCAI 2015)

Explosion of 3D Unet based Architecture



(a) Original UNet architecture



(b) Abstract topological structure of UNet-family models

Explosion of 3D Unet based Architecture

2015

- U-Net: Convolutional Networks for Biomedical Image Segmentation (MICCAI) [paper] [my-pytorch][keras]

2016

- V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation [paper] [caffe][pytorch]
- 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation [paper][pytorch]

2017

- H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation from CT Volumes (IEEE Transactions on Medical Imaging)[paper][keras]
- GP-Unet: Lesion Detection from Weak Labels with a 3D Regression Network (MICCAI) [paper]

Explosion of 3D Unet based Architecture

2018

- UNet++: A Nested U-Net Architecture for Medical Image Segmentation (MICCAI) [paper][my-pytorch][keras]
- MDU-Net: Multi-scale Densely Connected U-Net for biomedical image segmentation [paper]
- DUNet: A deformable network for retinal vessel segmentation [paper]
- RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans [paper]
- Dense Multi-path U-Net for Ischemic Stroke Lesion Segmentation in Multiple Image Modalities [paper]
- Stacked Dense U-Nets with Dual Transformers for Robust Face Alignment [paper]
- Prostate Segmentation using 2D Bridged U-net [paper]
- nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation [paper][pytorch]
- SUNet: a deep learning architecture for acute stroke lesion segmentation and outcome prediction in multimodal MRI [paper]
- IVD-Net: Intervertebral disc localization and segmentation in MRI with a multi-modal UNet [paper]
- LADDERNET: Multi-Path Networks Based on U-Net for Medical Image Segmentation [paper][pytorch]
- Glioma Segmentation with Cascaded Unet [paper]
- Attention U-Net: Learning Where to Look for the Pancreas [paper]
- Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation [paper]
- Concurrent Spatial and Channel 'Squeeze & Excitation' in Fully Convolutional Networks [paper]
- A Probabilistic U-Net for Segmentation of Ambiguous Images (NIPS) [paper] [tensorflow]
- AnatomyNet: Deep Learning for Fast and Fully Automated Whole-volume Segmentation of Head and Neck Anatomy [paper]
- 3D RoI-aware U-Net for Accurate and Efficient Colorectal Cancer Segmentation [paper][pytorch]
- Detection and Delineation of Acute Cerebral Infarct on DWI Using Weakly Supervised Machine Learning (Y-Net) (MICCAI) [paper](Page 82)
- Fully Dense UNet for 2D Sparse Photoacoustic Tomography Artifact Removal [paper]

Explosion of 3D Unet based Architecture

2018

- UNet++: A Nested U-Net Architecture for Medical Image Segmentation (MICCAI) [paper][my-pytorch][keras]
- MDU-Net: Multi-scale Densely Connected U-Net for biomedical image segmentation [paper]
- DUNet: A deformable network for retinal vessel segmentation [paper]
- RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans [paper]
- Dense Multi-path U-Net for Ischemic Stroke Lesion Segmentation in Multiple Image Modalities [paper]
- Stacked Dense U-Nets with Dual Transformers for Robust Face Alignment [paper]
- Prostate Segmentation using 2D Bridged U-net [paper]
- nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation [paper][pytorch]
- SUNet: a deep learning architecture for acute stroke lesion segmentation and outcome prediction in multimodal MRI [paper]
- IVD-Net: Intervertebral disc localization and segmentation in MRI with a multi-modal UNet [paper]
- LADDERNET: Multi-Path Networks Based on U-Net for Medical Image Segmentation [paper][pytorch]
- Glioma Segmentation with Cascaded Unet [paper]
- Attention U-Net: Learning Where to Look for the Pancreas [paper]
- Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation [paper]
- Concurrent Spatial and Channel 'Squeeze & Excitation' in Fully Convolutional Networks [paper]
- A Probabilistic U-Net for Segmentation of Ambiguous Images (NIPS) [paper] [tensorflow]
- AnatomyNet: Deep Learning for Fast and Fully Automated Whole-volume Segmentation of Head and Neck Anatomy [paper]
- 3D RoI-aware U-Net for Accurate and Efficient Colorectal Cancer Segmentation [paper][pytorch]
- Detection and Delineation of Acute Cerebral Infarct on DWI Using Weakly Supervised Machine Learning (Y-Net) (MICCAI) [paper](Page 82)
- Fully Dense UNet for 2D Sparse Photoacoustic Tomography Artifact Removal [paper]

2019

- MultiResUNet : Rethinking the U-Net Architecture for Multimodal Biomedical Image Segmentation [paper][keras]
- U-NetPlus: A Modified Encoder-Decoder U-Net Architecture for Semantic and Instance Segmentation of Surgical Instrument [paper]
- Probability Map Guided Bi-directional Recurrent UNet for Pancreas Segmentation [paper]
- CE-Net: Context Encoder Network for 2D Medical Image Segmentation [paper][pytorch]
- Graph U-Net [paper]
- A Novel Focal Tversky Loss Function with Improved Attention U-Net for Lesion Segmentation (ISBI) [paper]
- ST-UNet: A Spatio-Temporal U-Network for Graph-structured Time Series Modeling [paper]
- Connection Sensitive Attention U-NET for Accurate Retinal Vessel Segmentation [paper]
- CIA-Net: Robust Nuclei Instance Segmentation with Contour-aware Information Aggregation [paper]
- W-Net: Reinforced U-Net for Density Map Estimation [paper]
- Automated Segmentation of Pulmonary Lobes using Coordination-guided Deep Neural Networks (ISBI oral) [paper]
- U2-Net: A Bayesian U-Net Model with Epistemic Uncertainty Feedback for Photoreceptor Layer Segmentation in Pathological OCT Scans [paper]
- ScleraSegNet: an Improved U-Net Model with Attention for Accurate Sclera Segmentation (ICB Honorable Mention Paper Award) [paper]
- AHCNet: An Application of Attention Mechanism and Hybrid Connection for Liver Tumor Segmentation in CT Volumes [paper]
- A Hierarchical Probabilistic U-Net for Modeling Multi-Scale Ambiguities [paper]
- Recurrent U-Net for Resource-Constrained Segmentation [paper]
- MFP-Unet: A Novel Deep Learning Based Approach for Left Ventricle Segmentation in Echocardiography [paper]
- A Partially Reversible U-Net for Memory-Efficient Volumetric Image Segmentation (MICCAI 2019) [paper] [pytorch]
- ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data [paper]
- A multi-task U-net for segmentation with lazy labels [paper]
- RAUNet: Residual Attention U-Net for Semantic Segmentation of Cataract Surgical Instruments [paper]
- 3D U2-Net: A 3D Universal U-Net for Multi-Domain Medical Image Segmentation (MICCAI 2019) [paper] [pytorch]
- SegNAS3D: Network Architecture Search with Derivative-Free Global Optimization for 3D Image Segmentation (MICCAI 2019) [paper]
- 3D Dilated Multi-Fiber Network for Real-time Brain Tumor Segmentation in MRI [paper][pytorch] (MICCAI 2019)
- The Domain Shift Problem of Medical Image Segmentation and Vendor-Adaptation by Unet-GAN [paper]
- Recurrent U-Net for Resource-Constrained Segmentation [paper] (ICCV 2019)
- Siamese U-Net with Healthy Template for Accurate Segmentation of Intracranial Hemorrhage (MICCAI 2019)

DeepLab

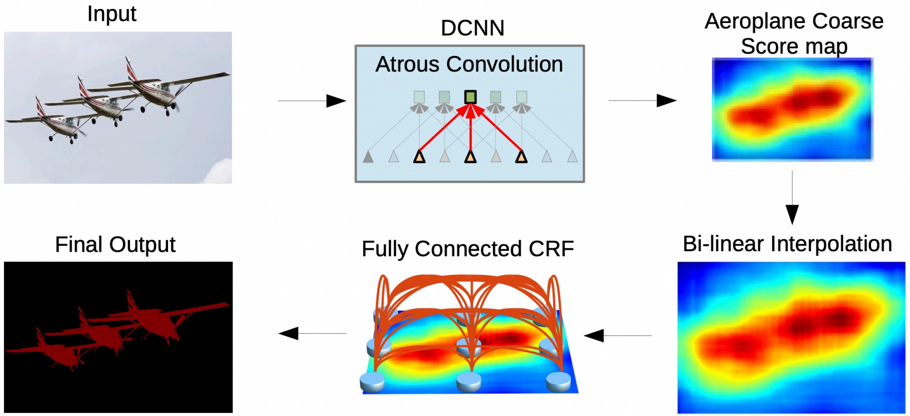
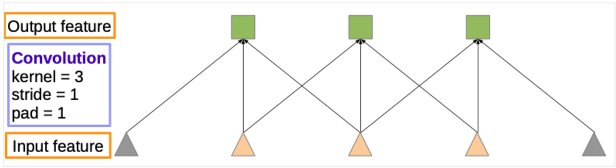


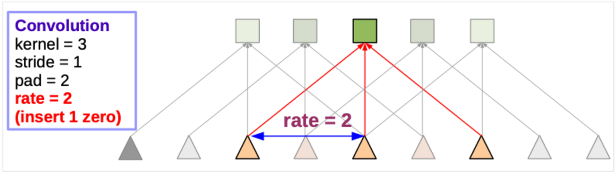
Fig. 1: Model Illustration. A Deep Convolutional Neural Network such as VGG-16 or ResNet-101 is employed in a fully convolutional fashion, using atrous convolution to reduce the degree of signal downsampling (from 32x down 8x). A bilinear interpolation stage enlarges the feature maps to the original image resolution. A fully connected CRF is then applied to refine the segmentation result and better capture the object boundaries.

- Atrous Convolutions
- Atrous Spatial Pyramid Pooling

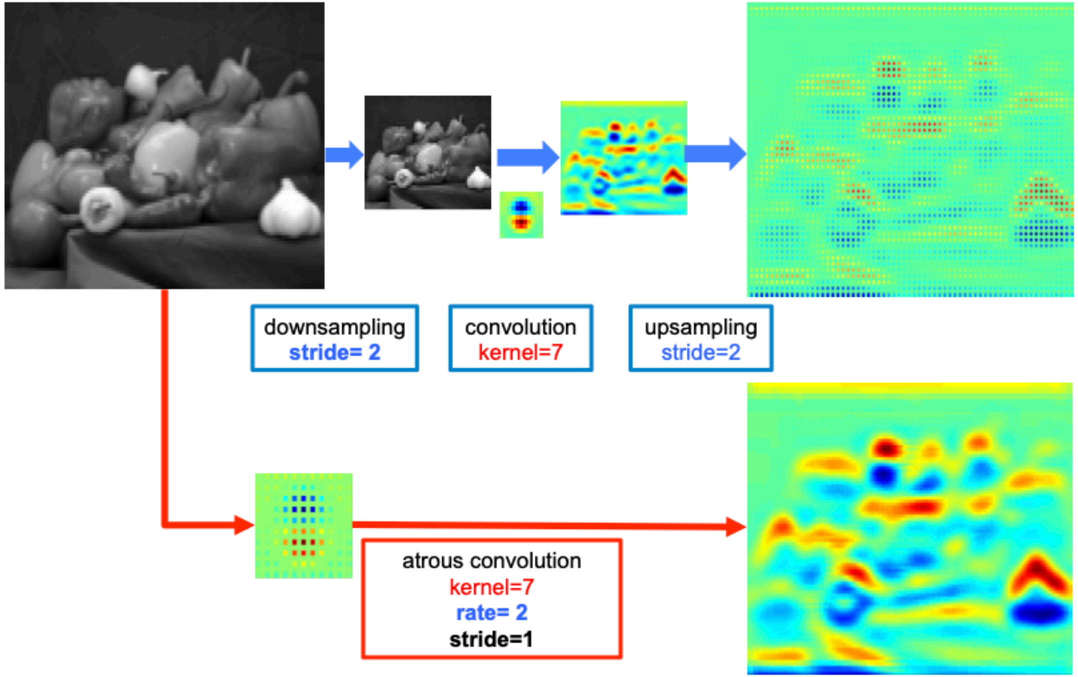
Atrous convolutions- Preserves Resolution



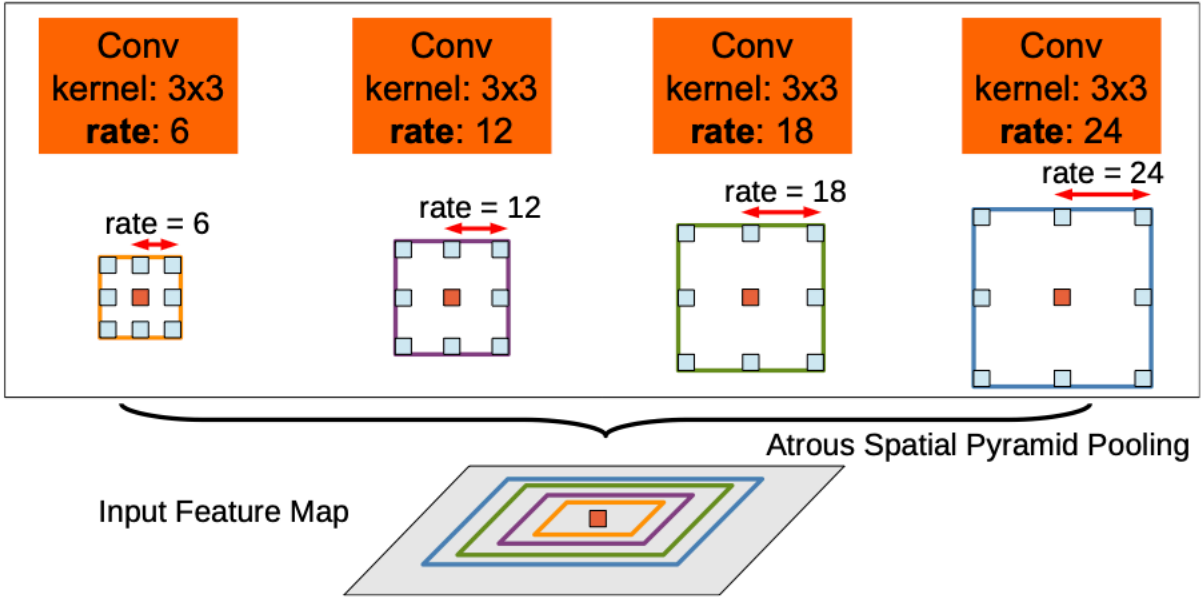
(a) Sparse feature extraction



(b) Dense feature extraction



Atrous Spatial Pyramid Pooling- Detection at multiple scales



Structured Prediction- Localization

$$E(\boldsymbol{x}) = \sum_i \theta_i(x_i) + \sum_{ij} \theta_{ij}(x_i, x_j)$$

$$\theta_i(x_i) = -\log P(x_i),$$

$$\theta_{ij}(x_i, x_j) = \mu(x_i, x_j)\Bigg[w_1 \exp\Big(-\frac{||p_i - p_j||^2}{2\sigma_\alpha^2} - \frac{||I_i - I_j||^2}{2\sigma_\beta^2}\Big) + w_2 \exp\Big(-\frac{||p_i - p_j||^2}{2\sigma_\gamma^2}\Big)\Bigg] \quad (3)$$

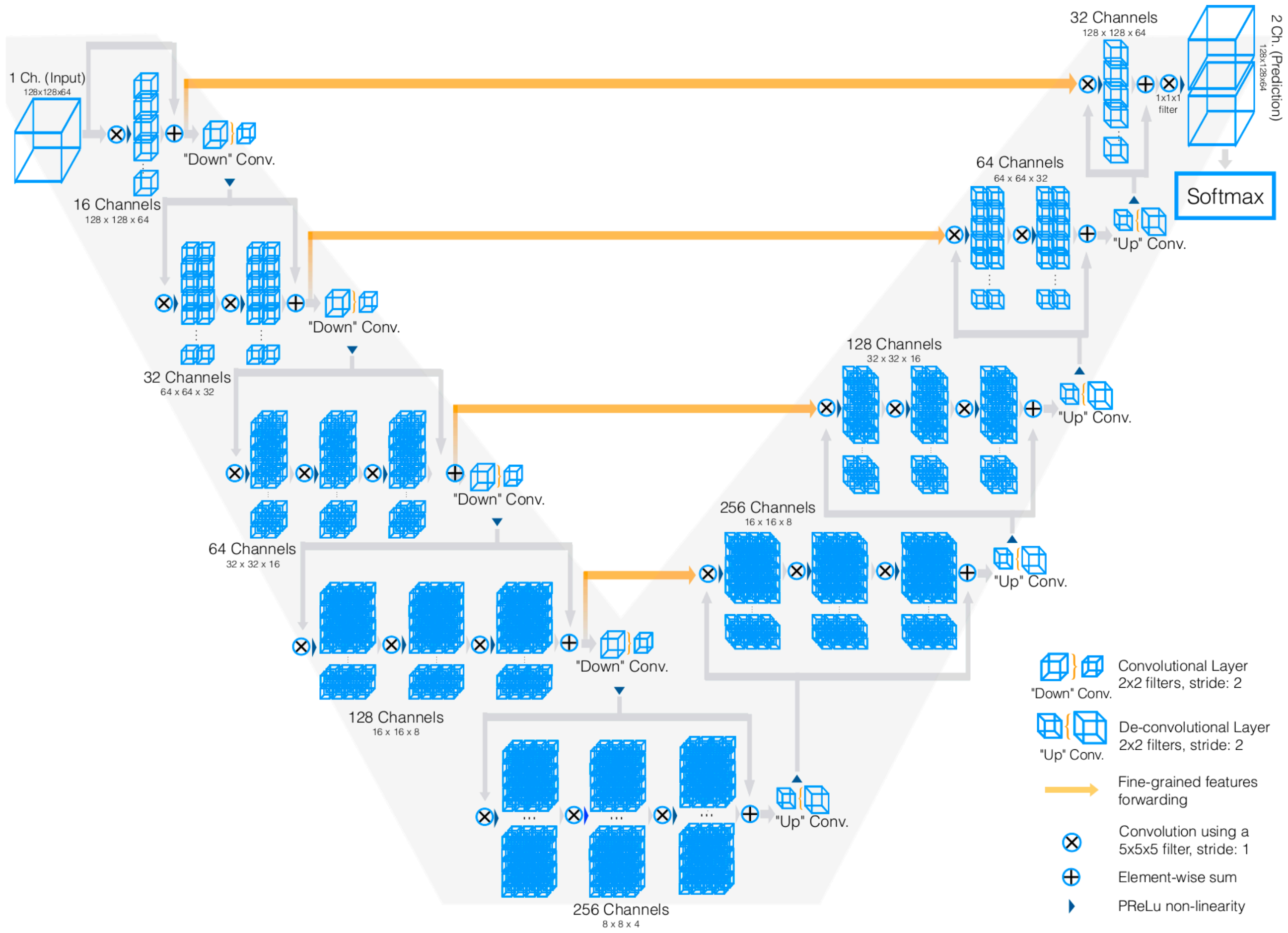
Same colours, are positions should have similar labels

Optimisation problem to find out
The best labelling to minimise this loss function

$\mu(x_i, x_j) = 1$ if $x_i \neq x_j$, and zero otherwise

Spatial proximity to enforce smoothness

V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation

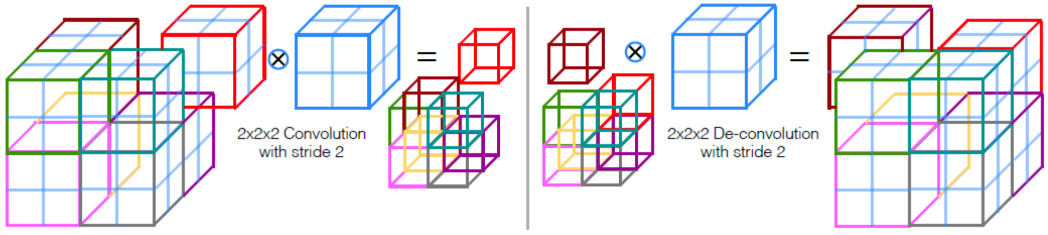


Left Side

1. The left side of the network is divided in different stages that operate at different resolutions. Each stage comprises one to three convolutional layers.
2. **At each stage, a residual function is learnt.** The input of each stage is used in the convolutional layers and processed through the non-linearities and added to the output of the last convolutional layer of that stage in order to enable learning a residual function. This architecture ensures convergence compared with non-residual learning network such as U-Net.
3. The **convolutions** performed in each stage use **volumetric kernels** having size of **5×5×5 voxels**. (A voxel represents a value on a regular grid in 3D space. The term voxel is commonly used in 3D many 3D space just like voxelization in point cloud.)
4. Along the compression path, **resolution is reduced by convolution with 2×2×2 voxels wide kernels applied with stride 2**. Thus, the size of the resulting feature maps is halved, with **similar purpose as pooling layers**. And **the number of feature channels doubles at each stage** of the compression path of the V-Net.
5. Replacing pooling operations with convolutional ones helps to have a smaller memory footprint during training, due to the fact that no switches mapping the output of pooling layers back to their inputs are needed for back-propagation.
6. Downsampling helps to increase the receptive field.
7. **PReLU** is used as non-linearity activation function. (PReLU is suggested in PReLU-Net.)

Right Side

1. The network extracts features and **expands the spatial support of the lower resolution feature maps** in order to gather and assemble the necessary information to output a two channel volumetric segmentation.



Convolution for Downsampling (Left), Deconvolution for Upsampling (Right)

1. At each stage, a **deconvolution** operation is employed in order **increase the size of the inputs followed by one to three convolutional layers, involving half the number of 5×5×5** kernels employed in the previous layer.

1. **Residual function** is learnt, similar to the left part of the network.

1. The two features maps computed by **the very last convolutional layer**, having **1×1×1 kernel size** and producing **outputs of the same size as the input volume**.

1. These two output feature maps are **the probabilistic segmentations of the foreground and background regions by applying soft-max voxelwise.**

Horizontal Connections

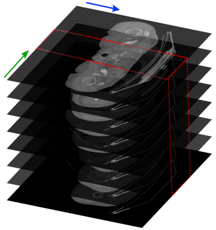
1. Similar to U-Net, **location information is lost in the compression path (left).**

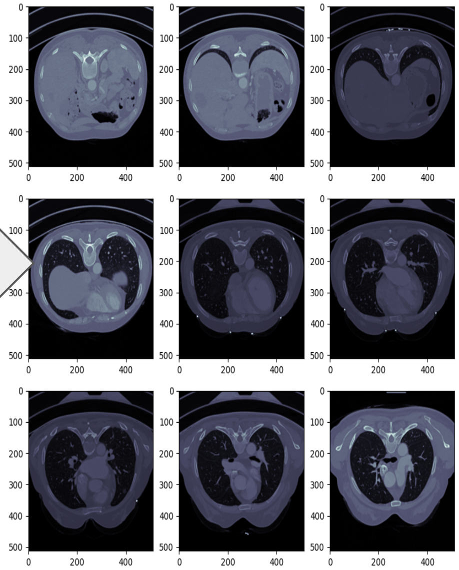
1. Thus, the features extracted from early stages of the left part of the CNN are forwarded to the right part through horizontal connections.

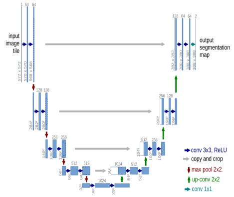
1. This can help to **provide location information to the right part, and improve the quality of the final contour prediction** and these connections improve the convergence time of the model.

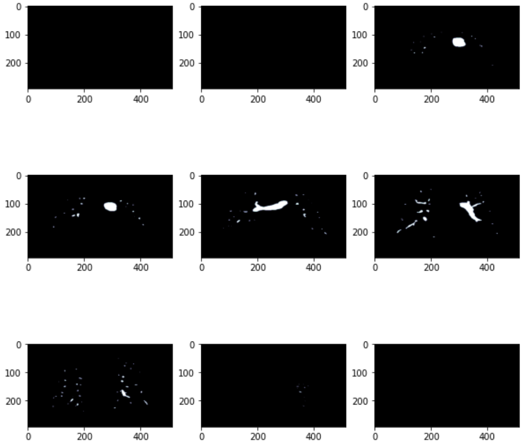
https://towardsdatascience.com/review-v-net-volumetric-convolution-biomedical-image-segmentation-aa15dbaea974

2D Network vs 3D Network

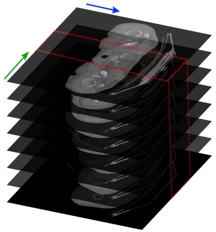


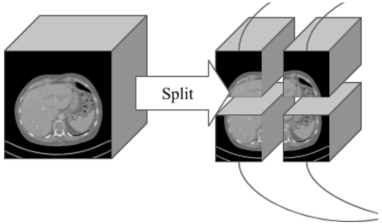


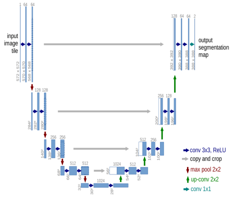




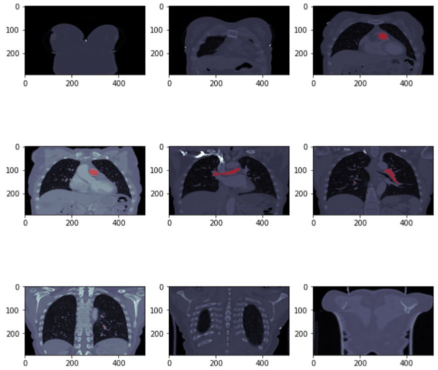
2D Network vs 3D Network



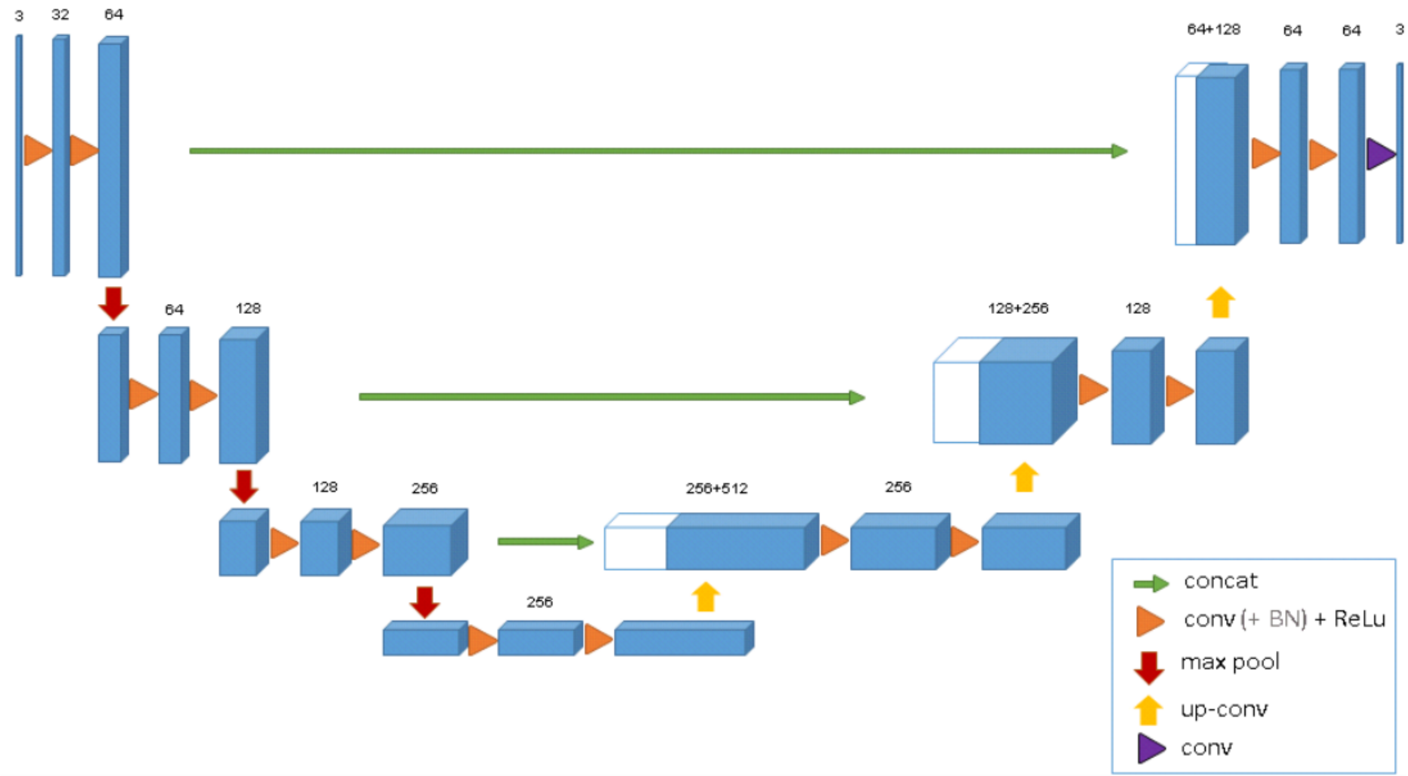




3D model that performs 3D Conv Net



3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation



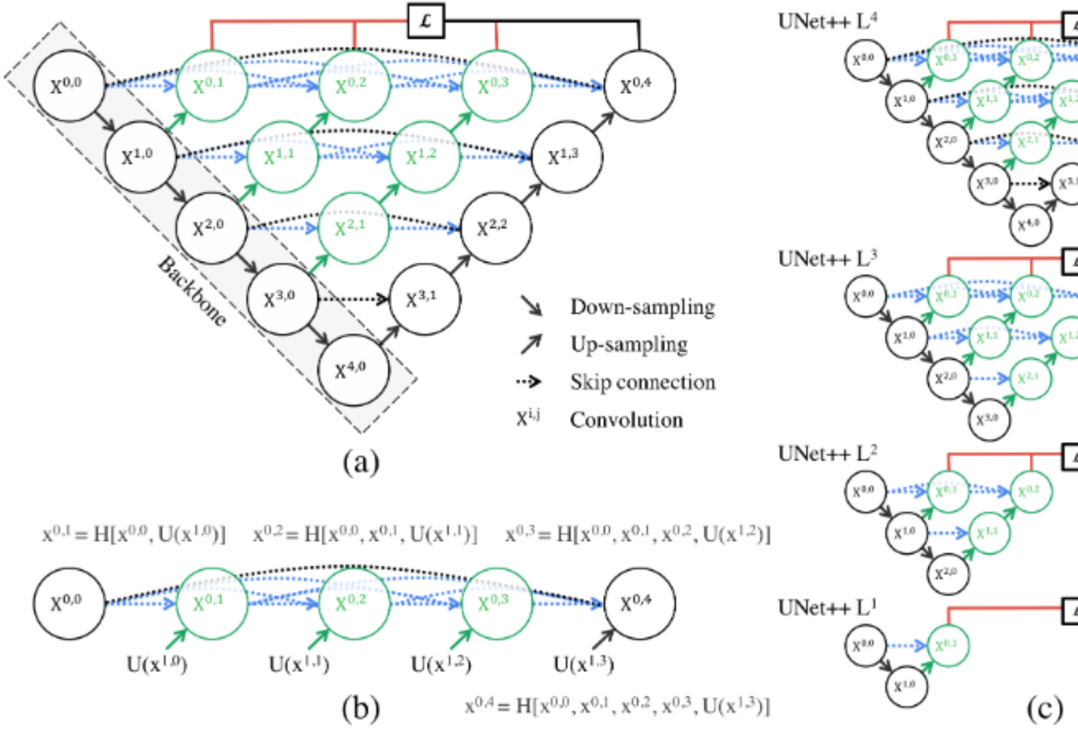
3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, Olaf Ronneberger(MICCAI 2016)

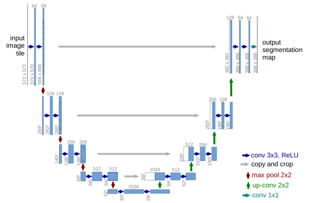
Unet++ : A Nested U-Net Architecture for Medical Image Segmentation

The re-designed skip pathways aim at reducing the **semantic gap** between the feature maps of the encoder and decoder sub-networks.

Dense skip connection inspired from DenseNet were introduced for better transfer of information and flow of gradients.

Unet++ : A Nested U-Net Architecture for Medical Image Segmentation





Unet++ : A Nested U-Net Architecture for Medical Image Segmentation

Deep Supervision in Unet++

- In Unet++, deep supervision is added(shown in red) so that the model can be pruned to adjust the model complexity, to balance between speed and performance.
- For accurate mode, the output from all segmentation is averaged.
- For fast mode the final segmentation map is selected from one of the segmentation branches.

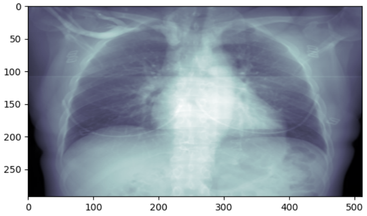
Similar to Unet++, researchers tried to make improvements in the Unet arch:

- ResUnet: Residual connection Unet
- Attention Unet: Where to look for pancreas?
- Siamese Unet

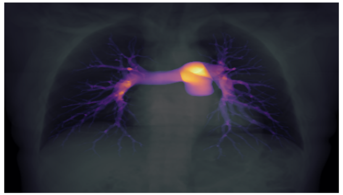
UNet++: A Nested U-Net Architecture for Medical Image Segmentation by Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang

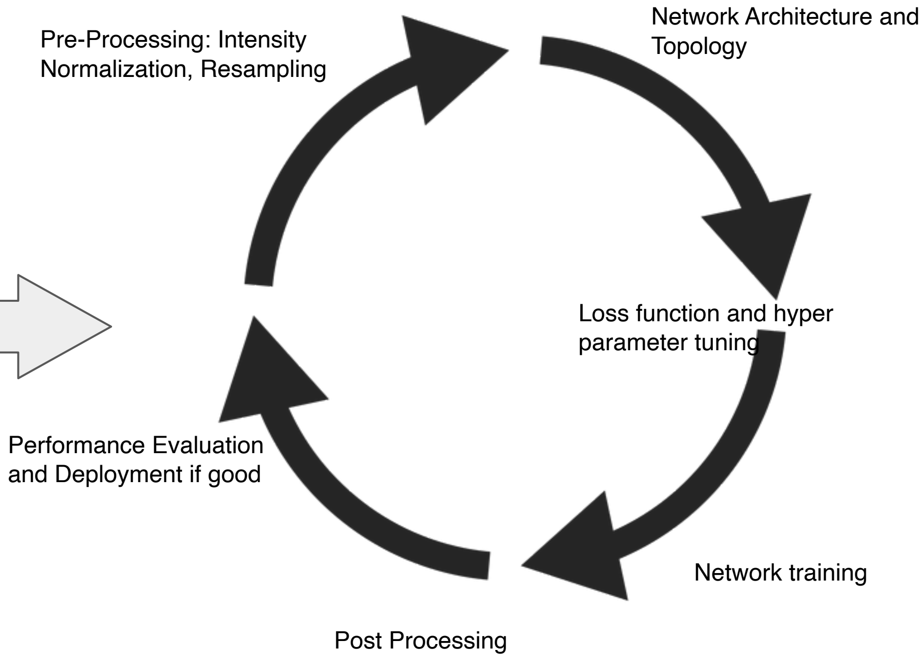
nn-Unet

- Stands for **No-new Unet.**
- Based on more practical and proven approach.



Training Data





Isensee, F., Jaeger, P.F., Kohl, S.A.A. *et al.* nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods* **18,** 203–211 (2021). https://doi.org/10.1038/s41592-020-01008-z

nn-Unet

Manual method configuration:

- Time consuming trial and error process
- Success depends upon the experience of researcher
- Inaccessible to non-expert
- Needs to be repeated for every dataset.

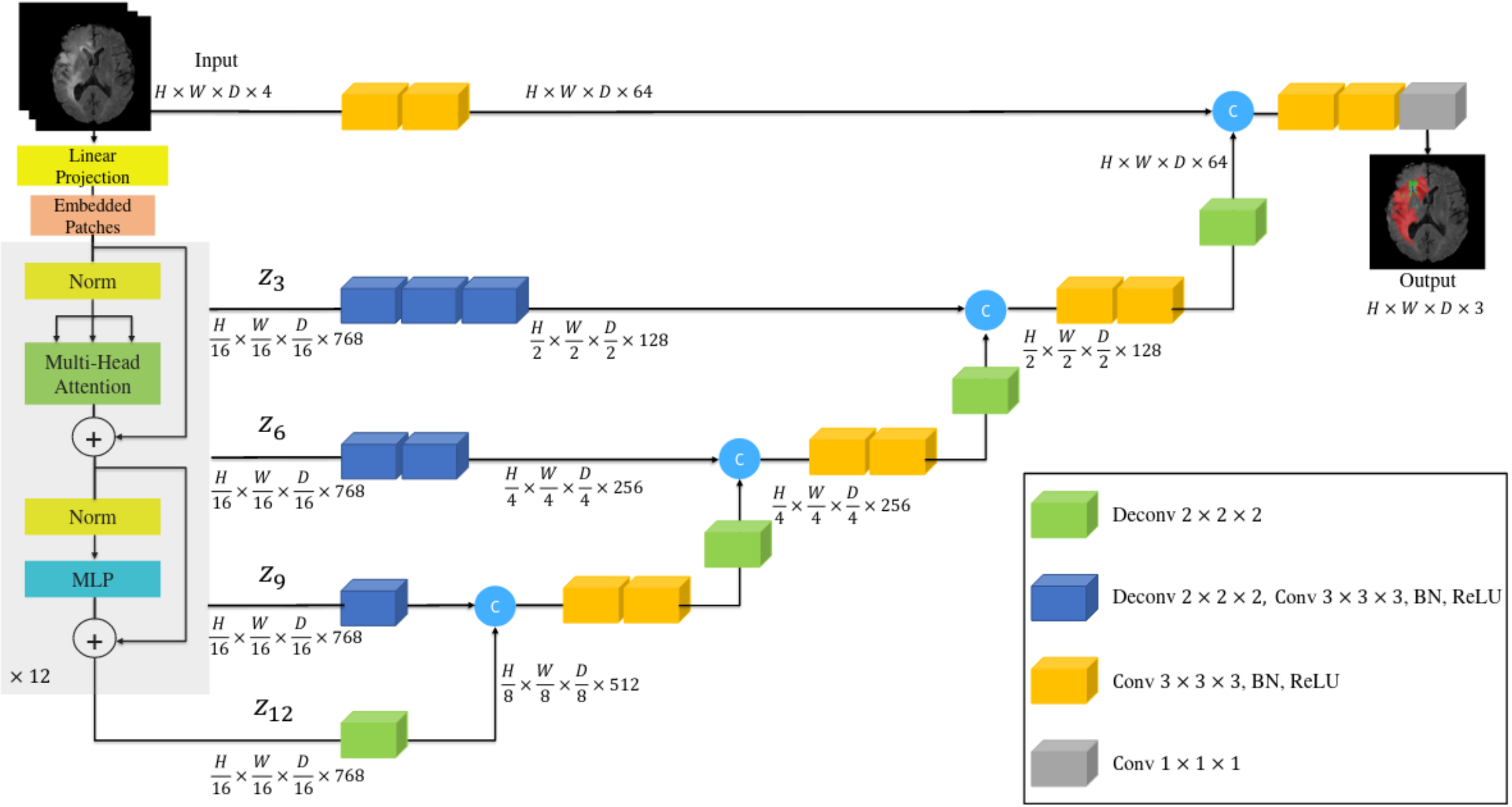
Network Architecture and Topology

- Is Unet++ better than Unet?
- Is Attention Unet better than Unet++?
- Which is the state of the art model for biomedical image segmentation?

Evaluation

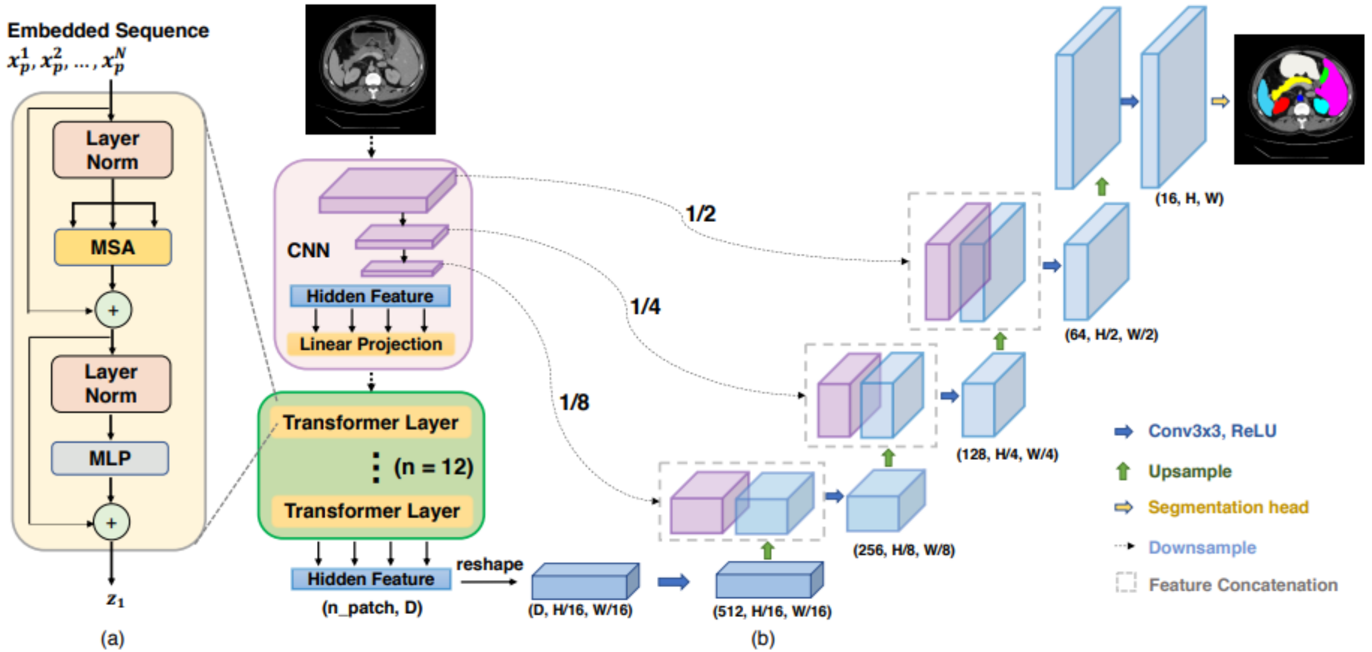
- Evaluation on Variety of dataset.
- Well tuned configuration of proposed method
- Well tuned configuration of baseline method.

UNETR: Transformers for 3D Medical Image Segmentation



UNETR: Transformers for 3D Medical Image Segmentation by Ali Hatamizadeh, Yucheng Tang, Vishwesh Nath, Dong Yang, Andriy Myronenko, Bennett Landman, Holger Roth, Daguang Xu

TransUNet



Overview of the framework.(a) schematic of the Transformer layer (b) architecture of the proposed TransUNet

Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., ... & Zhou, Y. (2021). Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*.

Other Variants Of UNet

1. MAnet
2. Linknet
3. FPN
4. PSPNet
5. PAN
6. DeepLabV3
7. DeepLabV3+
8. Attention UNet

Encoder Related Variants

ResNet
ResNeXt
ResNeSt
Res2Ne(X)t
RegNet(x/y)
GERNet
SE-Net
SK-ResNe(X)t
DenseNet
Inception
EfficientNet
MobileNet
DPN
VGG