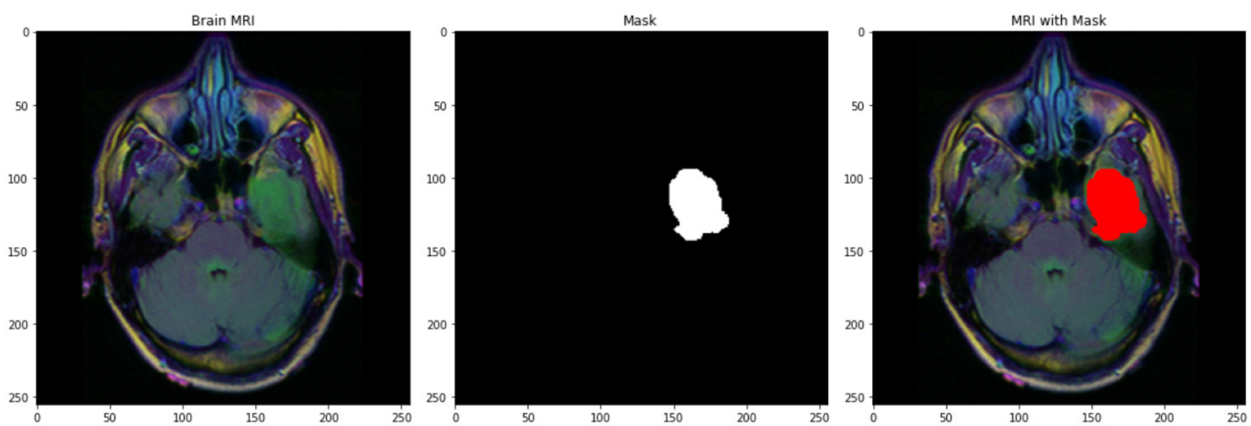
- Artificial Intelligence is revolutionizing Healthcare in many areas such as:
  - Disease Diagnosis with medical imaging
  - Surgical Robots
  - Maximizing Hospital Efficiency

- AI healthcare market is expected to reach $45.2 billion USD by 2026 from the current valuation of $4.9 billion USD.

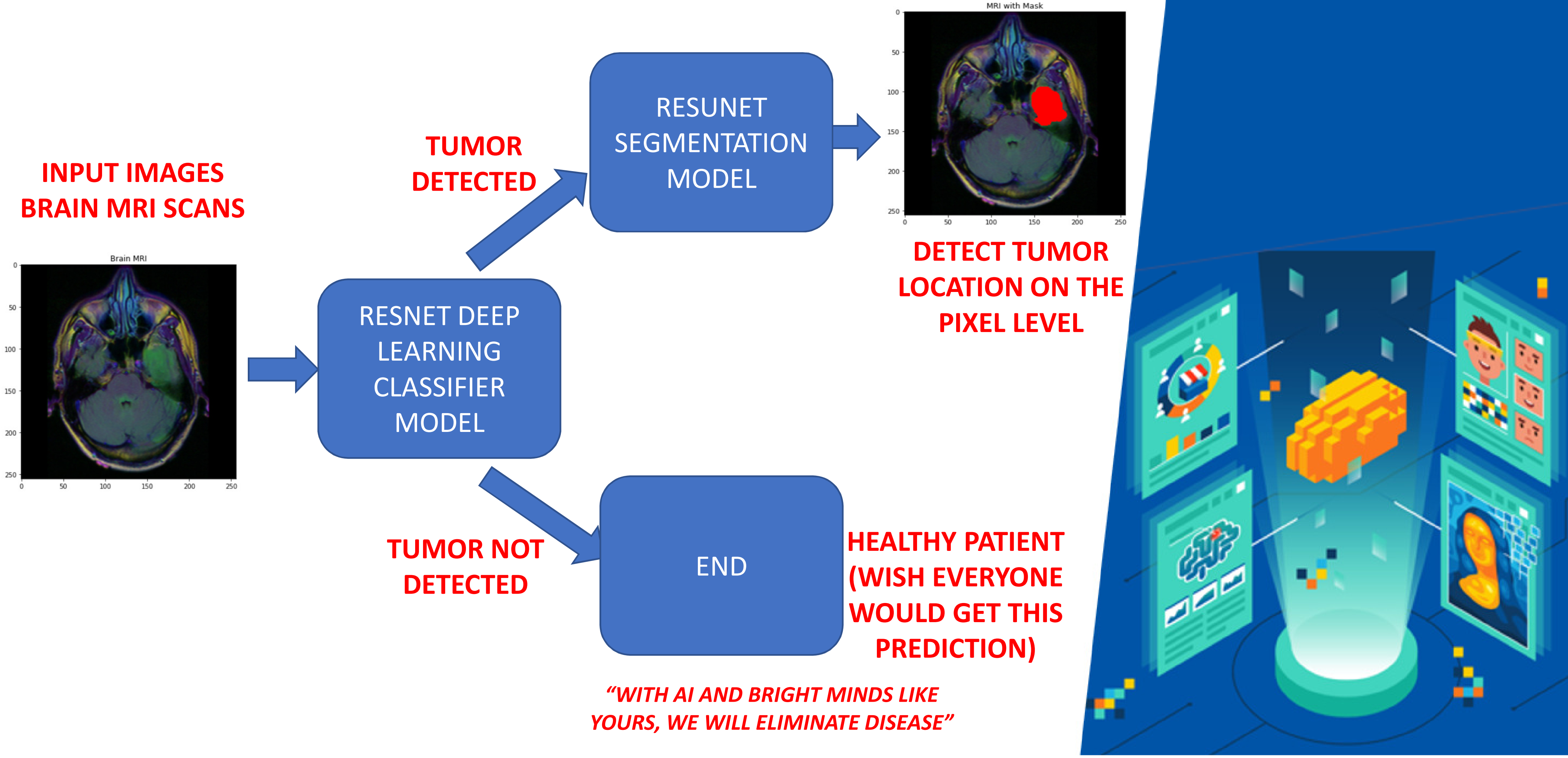
- Deep learning has been proven to be superior in detecting diseases from X-rays, MRI scans and CT scans which could significantly improve the speed and accuracy of diagnosis.

- Full list of startups: https://research.aimultiple.com/looking-for-better-medical-imaging-for-early-diagnostic-and-monitoring-contact-the-leading-vendors-here/

- In this case study, we will assume that you work as an AI/ML consultant and you have been hired by a medical diagnosis company in NYC.
- You have been tasked to improve the speed and accuracy of detecting and localizing brain tumors based on MRI scans.
- This would drastically reduce the cost of cancer diagnosis & help in early diagnosis of tumors which would essentially be a life saver.
- The team has collected brain MRI scans and have approached you to develop a model that could detect and localize tumors.
- You have been provided with 3929 Brain MRI scans along with their brain tumour location.

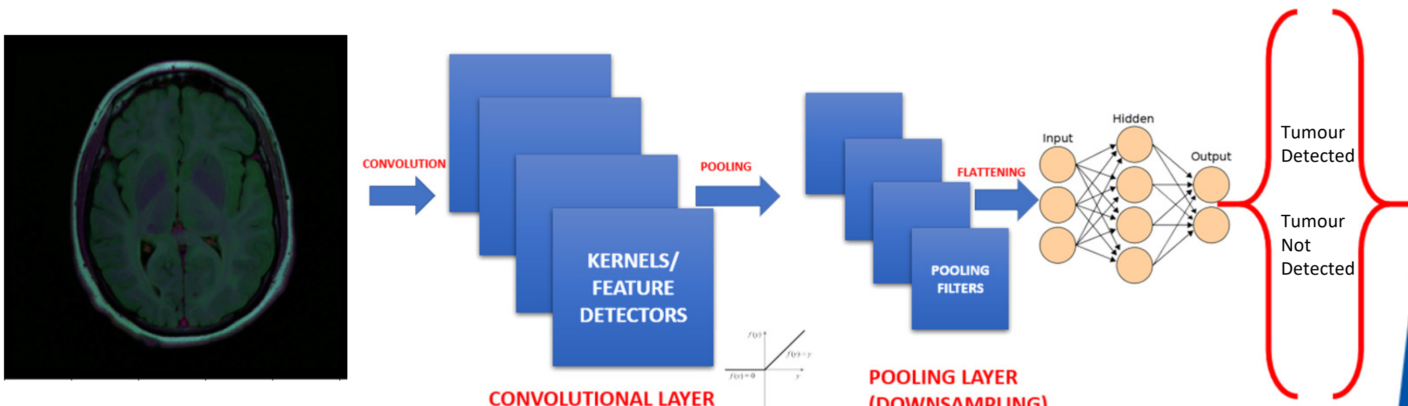


LAYERED DEEP LEARNING PIPELINE TO PERFORM CLASSIFICATION & SEGMENTATION



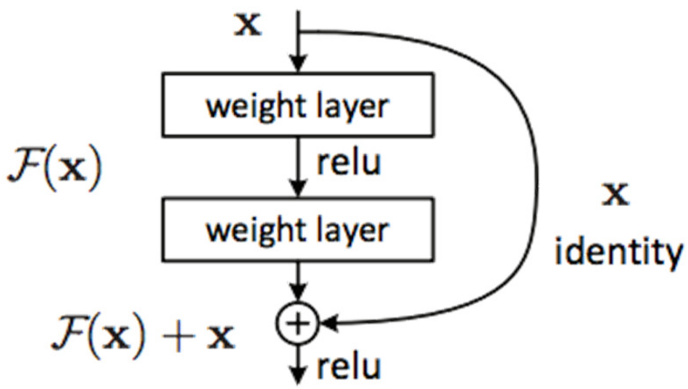
CONVOLUTIONAL NEURAL NETWORKS (REVIEW)

- The first CNN layers are used to extract high level general features.
- The last couple of layers are used to perform classification (on a specific task).
- Local respective fields scan the image first searching for simple shapes such as edges/lines
- These edges are then picked up by the subsequent layer to form more complex features.



RESNET (RESIDUAL NETWORK) (REVIEW)

- As CNNs grow deeper, vanishing gradient tend to occur which negatively impact network performance.
- Vanishing gradient problem occurs when the gradient is back-propagated to earlier layers which results in a very small gradient.
- Residual Neural Network includes "skip connection" feature which enables training of 152 layers without vanishing gradient issues.
- Resnet works by adding "identity mappings" on top of the CNN.
- ImageNet contains 11 million images and 11,000 categories.
- ImageNet is used to train ResNet deep network.



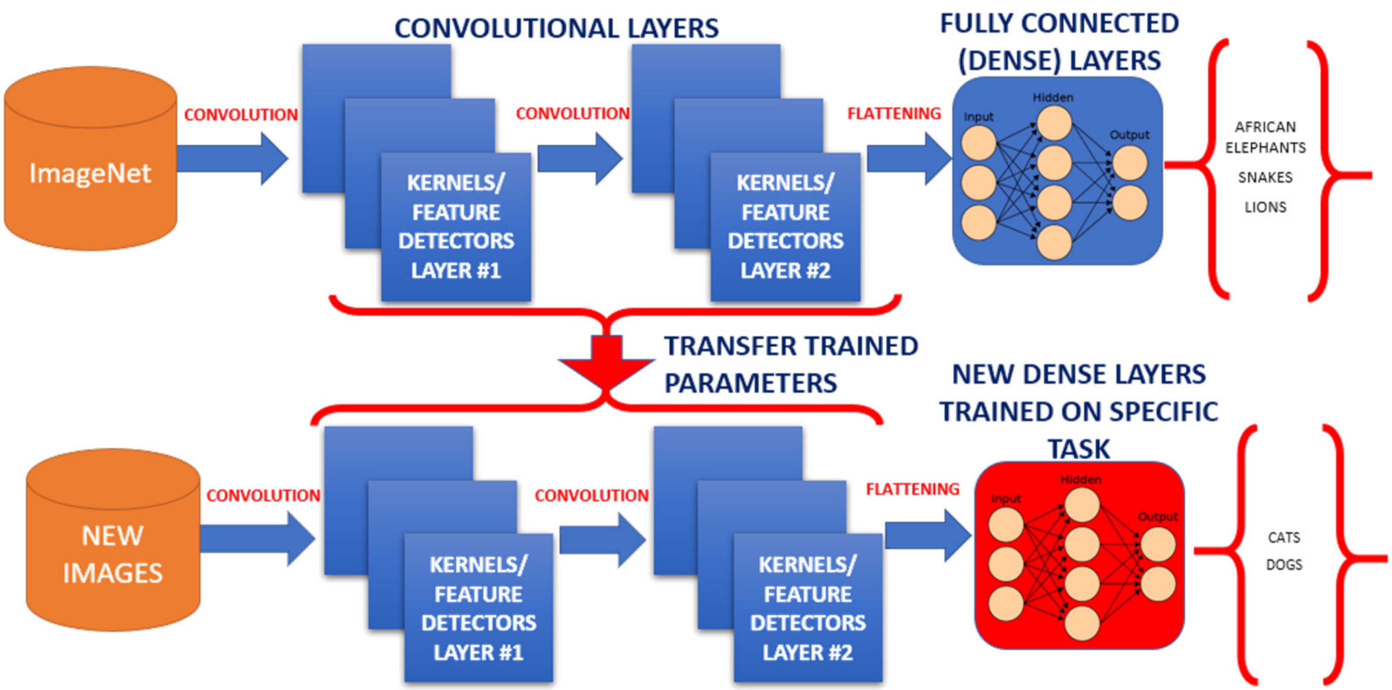
TRANSFER LEARNING?

- Transfer learning is a machine learning technique in which a network that has been trained to perform a specific task is being reused (repurposed) as a starting point for another similar task.
- Transfer learning is widely used since starting from a pre-trained models can dramatically reduce the computational time required if training is performed from scratch.

KNOWLEDGE TRANSFER

- Photo Credit: https://commons.wikimedia.org/wiki/File:Lillehammer_2016_-_Figure_Skating_Men_Short_Program_-_Camden_Pulkinen_2.jpg
- Photo Credit: https://commons.wikimedia.org/wiki/Alpine_skiing#/media/File:Andrej_%C5%A0porn_at_the_2010_Winter_Olympic_downhill.jpg
- Citations: Olga Russakovsky, *Jia Deng*, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg and Li Fei-Fei.
- ImageNet Large Scale Visual Recognition Challenge. arXiv:1409.0575, 2014.

TRANSFER LEARNING PROCESS

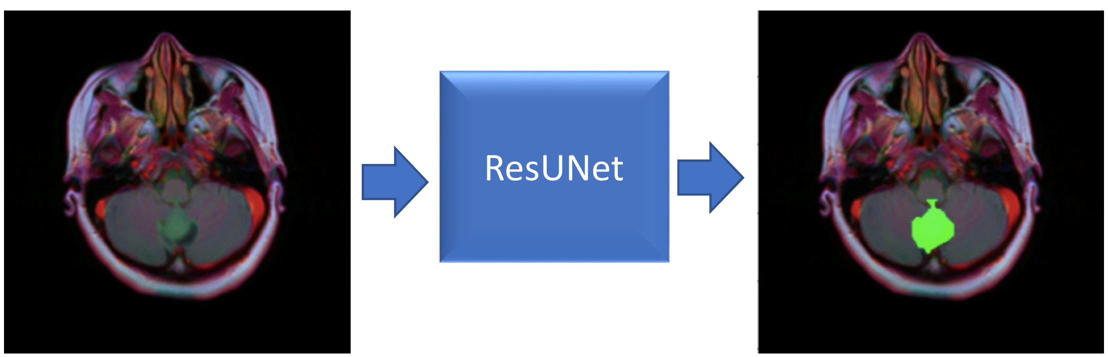


TRANSFER LEARNING TRAINING STRATEGIES

- Strategy #1 Steps:
  - Freeze the trained CNN network weights from the first layers.
  - Only train the newly added dense layers (with randomly initialized weights).
- Strategy #2 Steps:
  - Initialize the CNN network with the pre-trained weights
  - Retrain the entire CNN network while setting the learning rate to be very small, this is critical to ensure that you do not aggressively change the trained weights.
- Transfer learning advantages are:
  - Provides fast training progress, you don't have to start from scratch using randomly initialized weights
  - You can use small training dataset to achieve incredible results

WHAT IS IMAGE SEGMENTATION?

- The goal of image segmentation is to understand and extract information from images at the pixel-level.
- Image Segmentation can be used for object recognition and localization which offers tremendous value in many applications such as medical imaging and self-driving cars etc.
- The goal of image segmentation is to train a neural network to produce pixel-wise mask of the image.
- Modern image segmentation techniques are based on deep learning approach which makes use of common architectures such as CNN, FCNs (Fully Convolution Networks) and Deep Encoders-Decoders.
- You will use ResUNet architecture to solve the current task.



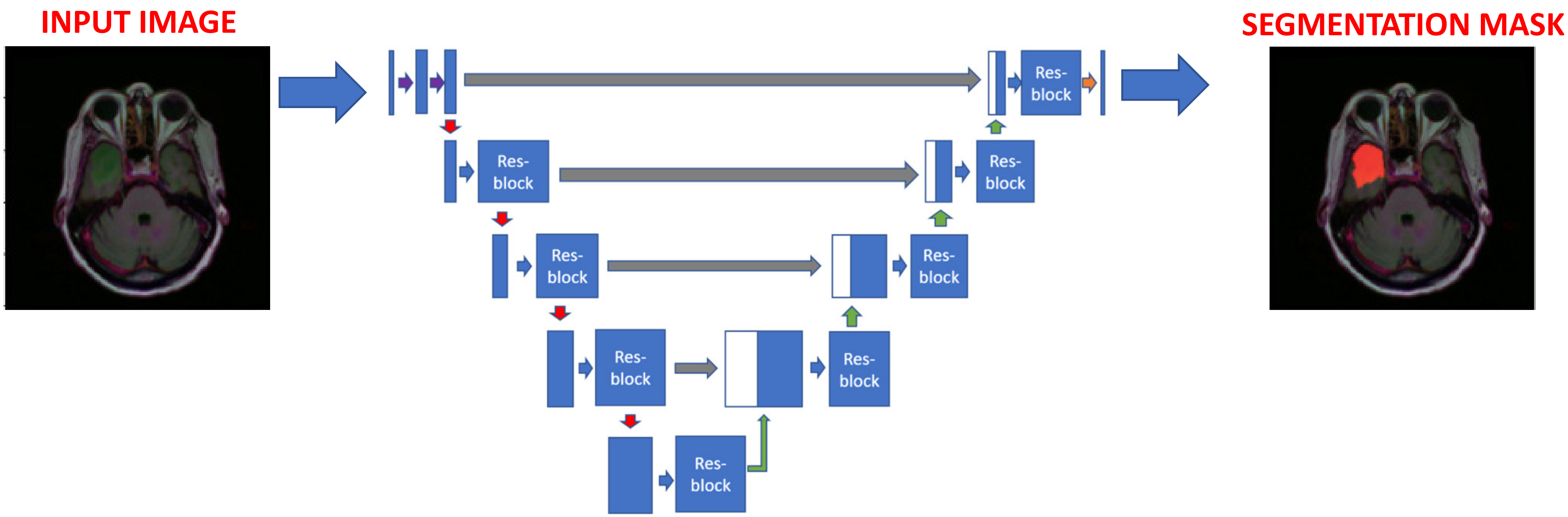
WHAT IS IMAGE SEGMENTATION?

- Recall when we applied CNN for image classification problems? We had to convert the image into a vector and possibly add a classification head at the end.
- However, in case of Unet, we convert (encode) the image into a vector followed by up sampling (decode) it back again into an image.
- In case of Unet, the input and output have the same size so the size of the image is preserved.
- <u>For classical CNNs:</u> they are generally used when the entire image is needed to be classified as a class label.
- <u>For Unet:</u> pixel level classification is performed.
- U-net formulates a loss function for every pixel in the input image.
- Softmax function is applied to every pixel which makes the segmentation problem works as a classification problem where classification is performed on every pixel of the image.

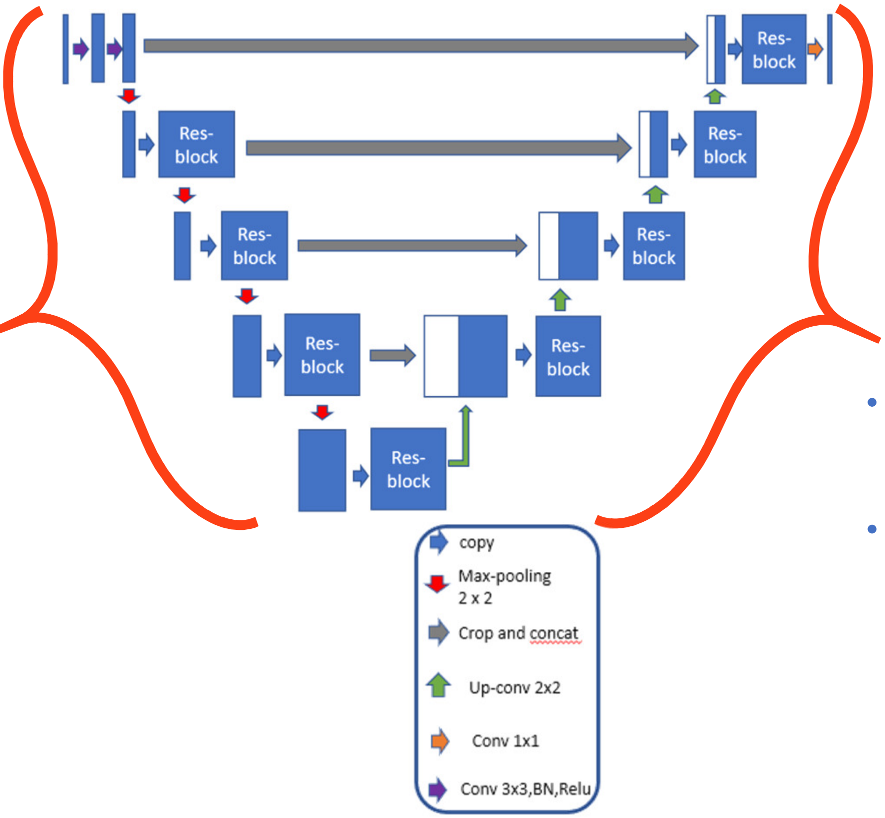
Great article by Aditi Mittal: https://towardsdatascience.com/introduction-to-u-net-and-res-net-for-image-segmentation-9afcb432ee2f

RESUNET

- ResUNet architecture combines UNet backbone architecture with residual blocks to overcome the vanishing gradients problems present in deep architectures.
- Unet architecture is based on Fully Convolutional Networks and modified in a way that it performs well on segmentation tasks.
- Resunet consists of three parts:
  - (1) Encoder or contracting path
  - (2) Bottleneck
  - (3) Decoder or expansive path



RESUNET ARCHITECTURE:



CONTRACTION PATH (ENCODER)

- The contraction path consist of several contraction blocks, each block takes an input that passes through res-blocks followed by 2x2 max pooling. Feature maps after each block doubles, which helps the model learn complex features effectively.

EXPANSION PATH (DECODER)

- Significant advantage of this architecture lies in expansion or decoder section. Each block takes in the up-sampled input from the previous layer and concatenates with the corresponding output features from the res-blocks in the contraction path. This is then again passed through the res-block followed by 2x2 up-sampling convolution layers.
- This helps to ensure that features learned while contracting are used while reconstructing the image.
- Finally in the last layer of expansion path, the output from the res-block is passed through 1x1 convolution layer to produce the desired output with the same size as the input.

BOTTLENECK

- The bottleneck block, serves as a connection between contraction path and expansion path.
- The block takes the input and then passes through a res-block followed by 2x2 up-sampling convolution layers.

RESUNET ARCHITECTURE:

<u>1. Encoder or contracting path consist of 4 blocks:</u>

- First block consists of 3x3 convolution layer + Relu + Batch-Normalization
- Remaining three blocks consist of Res-blocks followed by Max-pooling 2x2.

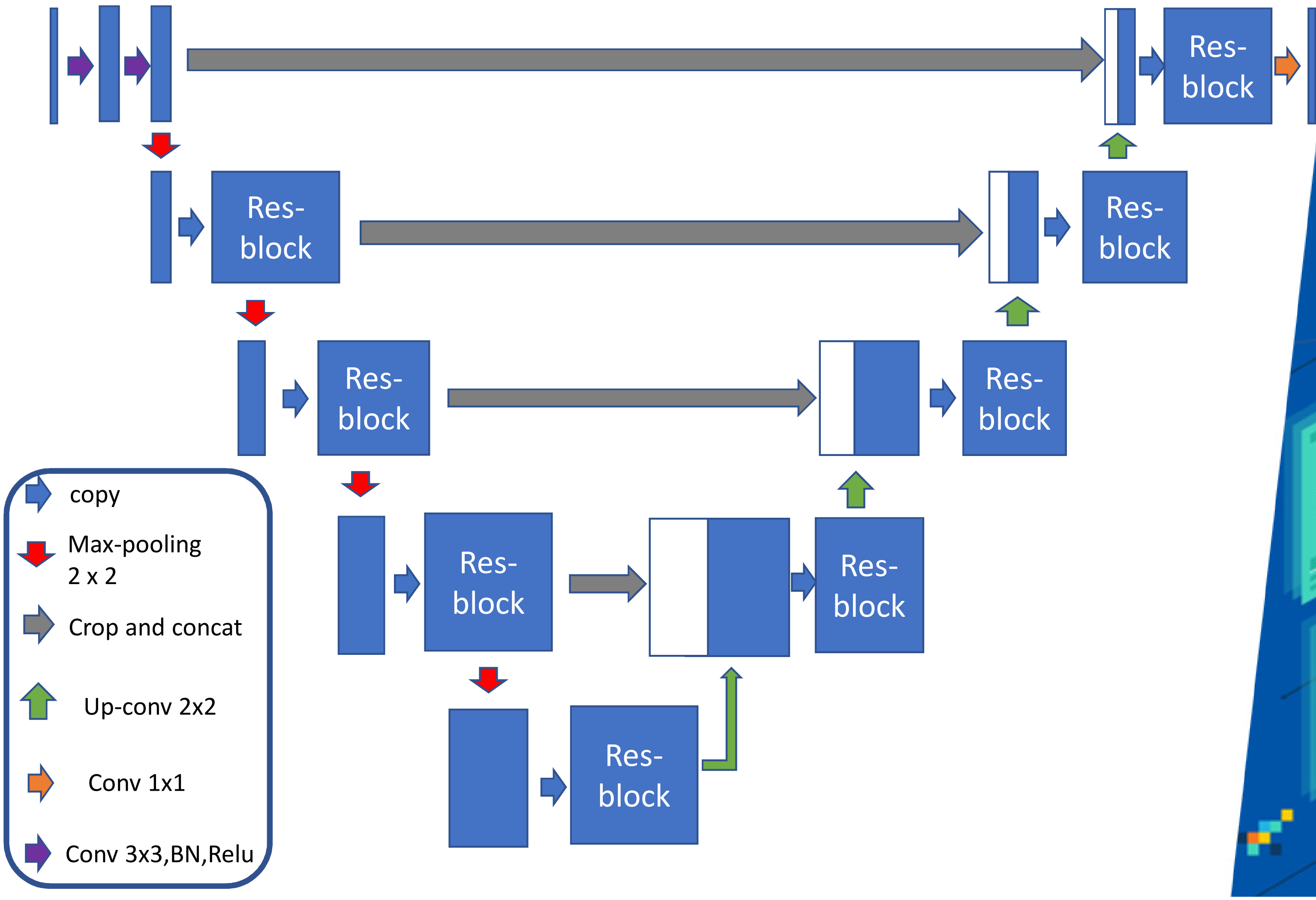
<u>2. Bottleneck:</u>

- It is in-between the contracting and expanding path.
- It consist of Res-block followed by up sampling conv layer 2x2.

<u>3. Expanding or Decoder path consist of 4 blocks:</u>

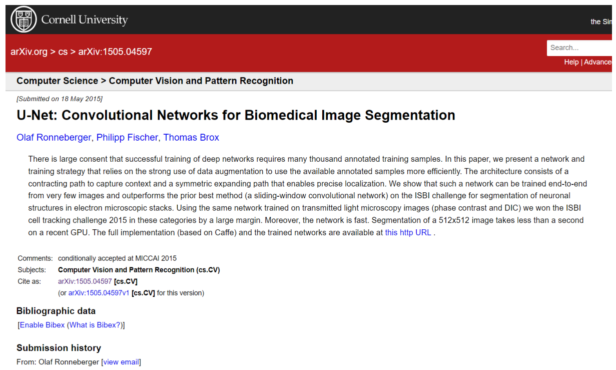
- 3 blocks following bottleneck consist of Res-blocks followed by up-sampling conv layer 2 x 2
- Final block consist of Res-block followed by 1x1 conv layer.

RESUNET ARCHITECTURE:



RESUNET ADDITIONAL RESOURCES:

Paper #1: https://arxiv.org/abs/1505.04597



Cornell University

arXiv.org > cs > arXiv:1505.04597

Computer Science > Computer Vision and Pattern Recognition

[Submitted on 18 May 2015]

U-Net: Convolutional Networks for Biomedical Image Segmentation

Olaf Ronneberger, Philipp Fischer, Thomas Brox

There is large consent that successful training of deep networks requires many thousand annotated training samples. In this paper, we present a network and training strategy that relies on the strong use of data augmentation to use the available annotated samples more efficiently. The architecture consists of a contracting path to capture context and a symmetric expanding path that enables precise localization. We show that such a network can be trained end-to-end from very few images and outperforms the prior best method (a sliding-window convolutional network) on the ISBI challenge for segmentation of neuronal structures in electron microscopic stacks. Using the same network trained on transmitted light microscopy images (phase contrast and DIC) we won the ISBI cell tracking challenge 2015 in these categories by a large margin. Moreover, the network is fast. Segmentation of a 512x512 image takes less than a second on a recent GPU. The full implementation (based on Caffe) and the trained networks are available at this http URL .

Comments: conditionally accepted at MICCAI 2015
Subjects: **Computer Vision and Pattern Recognition (cs.CV)**
Cite as: arXiv:1505.04597 **[cs.CV]**
(or arXiv:1505.04597v1 **[cs.CV]** for this version)

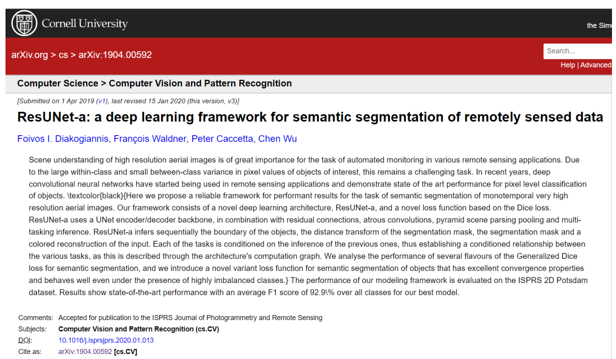
Bibliographic data

[Enable Bibex (What is Bibex?)]

Submission history

From: Olaf Ronneberger [view email]

Paper #2: https://arxiv.org/abs/1904.00592



Cornell University

arXiv.org > cs > arXiv:1904.00592

Computer Science > Computer Vision and Pattern Recognition

[Submitted on 1 Apr 2019 (v1), last revised 15 Jan 2020 (this version, v3)]

ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data

Foivos I. Diakogiannis, François Waldner, Peter Caccetta, Chen Wu

Scene understanding of high resolution aerial images is of great importance for the task of automated monitoring in various remote sensing applications. Due to the large within-class and small between-class variance in pixel values of objects of interest, this remains a challenging task. In recent years, deep convolutional neural networks have started being used in remote sensing applications and demonstrate state of the art performance for pixel level classification of objects. \textcolor{black}{Here we propose a reliable framework for performant results for the task of semantic segmentation of monotemporal very high resolution aerial images. Our framework consists of a novel deep learning architecture, ResUNet-a, and a novel loss function based on the Dice loss. ResUNet-a uses a UNet encoder/decoder backbone, in combination with residual connections, atrous convolutions, pyramid scene parsing pooling and multi-tasking inference. ResUNet-a infers sequentially the boundary of the objects, the distance transform of the segmentation mask, the segmentation mask and a colored reconstruction of the input. Each of the tasks is conditioned on the inference of the previous ones, thus establishing a conditioned relationship between the various tasks, as this is described through the architecture's computation graph. We analyse the performance of several flavours of the Generalized Dice loss for semantic segmentation, and we introduce a novel variant loss function for semantic segmentation of objects that has excellent convergence properties and behaves well even under the presence of highly imbalanced classes.} The performance of our modeling framework is evaluated on the ISPRS 2D Potsdam dataset. Results show state-of-the-art performance with an average F1 score of 92.9\% over all classes for our best model.

Comments: Accepted for publication to the ISPRS Journal of Photogrammetry and Remote Sensing
Subjects: **Computer Vision and Pattern Recognition (cs.CV)**
DOI: 10.1016/j.isprsjprs.2020.01.013
Cite as: arXiv:1904.00592 **[cs.CV]**
(or arXiv:1904.00592v3 **[cs.CV]** for this version)

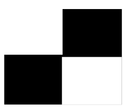
Great article: https://towardsdatascience.com/introduction-to-u-net-and-res-net-for-image-segmentation-9afcb432ee2f

MASK:

- The goal of image segmentation is to understand the image at the pixel level. It associates each pixel with a certain class. The output produce by image segmentation model is called a “mask” of the image.
- Masks can be represented by associating pixel values with their coordinates. For example if we have a black image of shape (2,2), this can be represented as:

[[0, 0],
[0, 0]]

If our output mask is as follows:



[[255, 0],
[0,255]]

- To represent this mask we have to first flatten the image into a 1-D array. This would result in something like [255,0,0,255] for mask. Then, we can use the index to create the mask. Finally we would have something like [1,0,0,1] as our mask.