Semantic Segmentation of Brain Tumor on multi-band 2D stacked-up slices using Attention-Based U-Net

Kulendu K Chakraborty (CSE/190310007025)[1]
Bishal Roy(CSE/190310007008)[1]
Abhimanyu Kumar(CSE/190310007002)[1]

[1]Dept. of Computer Science & Engineering, GIMT

Guided by Dr. Minakshi Gogoi

HoD and Associate Prof., Dept. of Computer Science & Engineering, GIMT

27^{th} June, 2023

Guess?

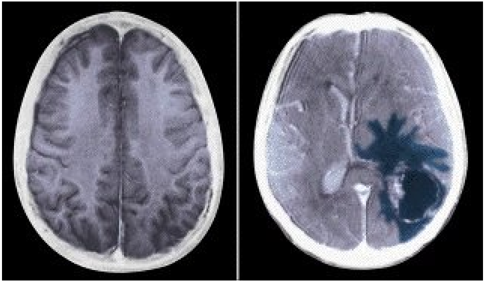


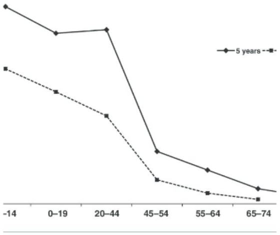
Figure 1: Taken from Research Gate

Guess?

Image Description

Healthy Brain (left) compared to Brain tumor (right) [marked in red]

Figure 2: Taken from Research Gate



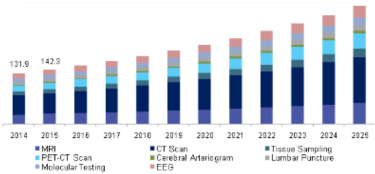


Figure 3: Statistics of Survival Rates by Age, 2006-2010

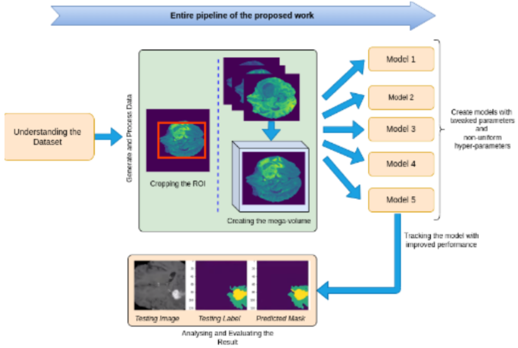
- Brain Tumor formed is dangerous, but out of which **36%** of them are cancerous tumors.
- The **10-year survival rate** is almost 31%.
- The **5-year** survival rate for people younger than age 15 is about **75%**.

Follow-up Baseline

Prior work done →

Aim

Semantic Segmentation of Brain Tumor on multi-band 2D stacked-up slices using non-uniform 3D U-Nets



Follow-up Baseline

Models	Training		Validation		Mean IoU	IoU score
	Accuracy	Loss	Accuracy	Loss		
$\phi_{default}$	0.9767	0.8325	0.9467	0.8900	0.4282	0.5948
ϕ_{elu}	0.9821	0.8144	0.9456	0.8827	0.4448	0.6602
ϕ_{selu}	0.9816	0.8184	0.9229	0.8955	0.4381	0.6448
$\phi_{five-conv}$	0.9767	0.8361	0.9560	0.8899	0.4429	0.5766
$\phi_{uniform-dropout}$	0.9784	0.8285	0.9536	0.9038	0.3806	0.6070

Note: The best comparative metric values are shown in red

Models (Tweaked Architectures, ϕ)	Input block size (C)	Trainable Params. (α_{train})	Activation functions ($f(A)$)
$\phi_{default}$	128, 128, 128	5,645,828	ReLU
ϕ_{elu}	128, 128, 128	5,645,828	ELU
ϕ_{selu}	128, 128, 128	5,645,828	SELU
$\phi_{five-conv}$	128, 128, 128	14,481,572	ReLU
$\phi_{uniform-dropout}$	128, 128, 128	5,645,828	ReLU

- **Step 1:** Understand the Dataset (*BraTS'20 Challenge Data*), with all the medical arrangements of the terminologies.
- **Step 2:** Generate, scale, and process the Data.
- **Step 3:** Define variants of the 3D U-Net architecture.
- **Step 4:** Train the Segmentation Models (*U-Net*).
- **Step 5:** Track the performance of the other models while training, with hand-tweaked parameters.
- **Step 6:** Performance analysis of the models trained and generated outputs.
- **Step 7:** Evaluate with the Benchmark parameters and compare the Mask.
- **Step 8:** Select the model with the best performance.

ϕ_{elu} shows best performance with an accuracy of 0.9821 and loss of 0.8144, showing better comparative results on all the other models (ϕ)

Given a multi-band 3D scan of the brain (*preferably MRI scans*), we aim to extract the slices from the scan (*3D scan model*), and intend to segment the Tumor from the volumized layers semantically using Attention-based U-Net.

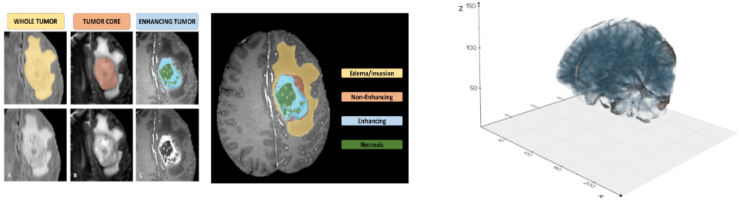
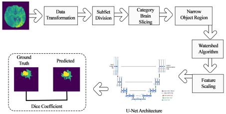
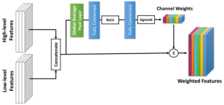


Figure 4: From Left, Analogy of components of a Brain Tumor, output of a processed MRI-scan

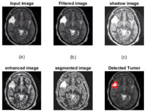
Motivation



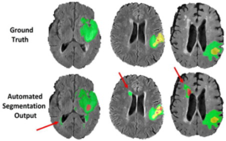
Anuja Arora et. al (2021)



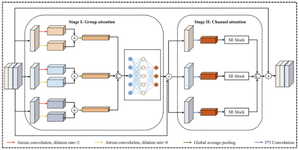
Mehrdad Noori et. al (2020) - Baseline



Md Khairul Islam et. al (2021)



Ujjwal Baid et. al (2021)



Zheng Huang et. al (2021) - Main Motivation

More Literature reviews

Table 1: Comparison with previous work

Authors	**Methodology**	**Benchmark Dataset used**	**Accuracy Metric (dice score)**
Sasha Yousef et al.	VM,original U-net and Modified U-net architecture by adding VGG16 and VGG19 structure	MICCAI BraTS 2020	0.65
Shaoguo Cui et.al	A Deep cascaded convolution neural network,consisting of two sub network TLN and ITCN	BraTS 2015	DSC (0.89, 0.77, and 0.80)
Leon Weninger et.a	Unsupervised deep learning methods such as VAEs and GANs	BraTs 2018	0.75-0.90
Andriy myronenko et.al	Encoder-decoder based CNN architecture ResNet	BraTs 2019	(0.826, 0.882, 0.882) for EC, WT and TC
MD Abdullah Al Nasim et.al	2D U-Net Architecture	BraTs 2017, 2018, 2019 and 2020	0.84, 0.81, 0.84 and 0.83
Runwei Zhou et.al	3D Fully convolutional neural network with a three layer encoder-decoder	BraTs 2019	DSC for training dataset (0.88, 0.83, 0.73) and validation dataset (0.87, 0.75, 0.76)
Xue Feng et.a	Ensemble of multiple models trained with different hyperparameters based on 3D U-Net. U-Net vs Dense-Net	BraTs 2018	(0.7946, 0.9114, 0.8304) for ET, WT and TC
Mehrdad Noori et.al	2D U-Net architecture used with two approaches, Attention mechanism and Multi-View Fusion.	BraTS 2017 and 2018	(0.813, 0.895 and 0.823) with BraTS2018 and (0.791 0.885 0.783) with BraTS 2017 for ET, WT and TC
Navchetan Awasthi et.al	Segmentation of different tumor region using Multi-Threshold model based on 2D Attention U-Net (MTAU)	BraTS 2020	Dice coefficient (0.59, 0.72 and 0.61) for training dataset, with validation dataset (0.57, 0.73, and 0.61) and with test dataset (0.59, 0.72 and 0.57)

Thorough **Literature** Review: Link

Our world is **3D**

Why 3D scans?

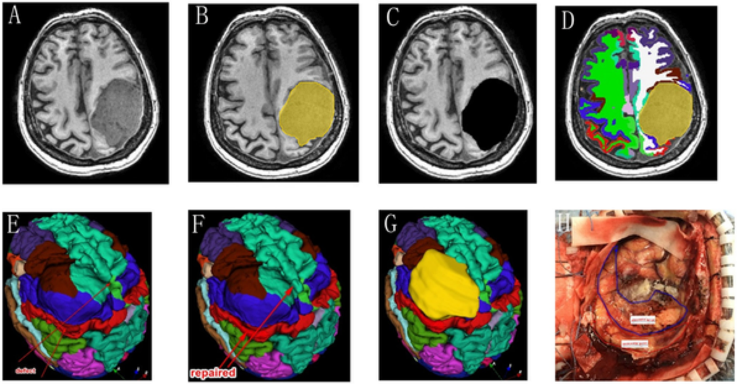


Figure 5: Xiaolin Hou et. al, "A new simple brain segmentation method for extracerebral intracranial tumors"

Why 3D scans?

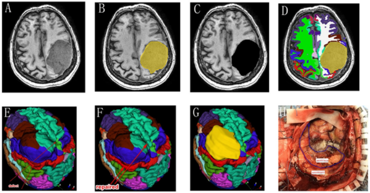
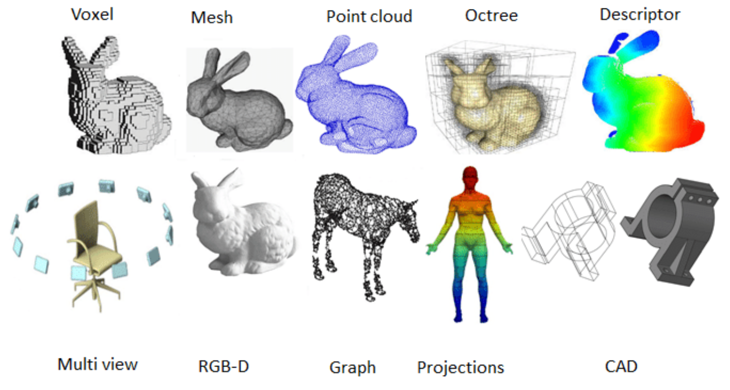


Figure 6: Xiaolin Hou et. al, "A new simple brain segmentation method for extracerebral intracranial tumors"

- More **dimensional-information**,
- Better parameterized **arguments** to localize the **area of abnormality**,
- Better **voxel neighborhood**,
- Registration of **High frequencies**

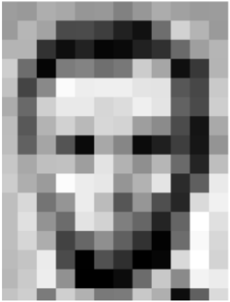
Representation of 3D data

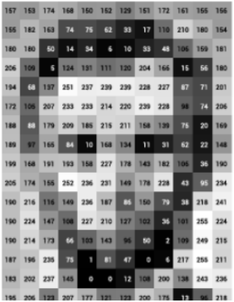


Source: Prof. A Sharma's slides

Pixels (*2D images*)

Pixelization





157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

What are Voxels?

Vox-els - Volumetric Pixels

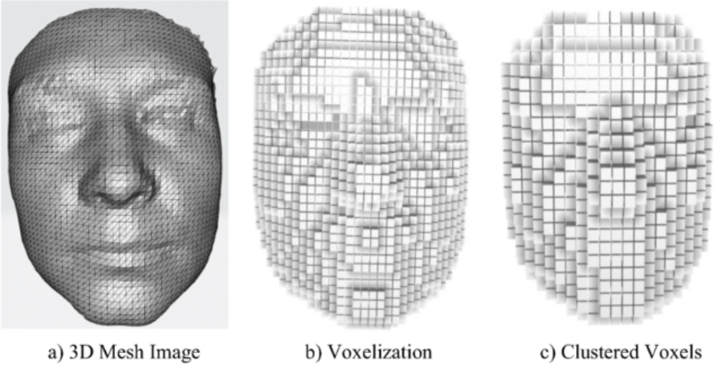


Figure 7: Volumized clusters created from a 3D face scan

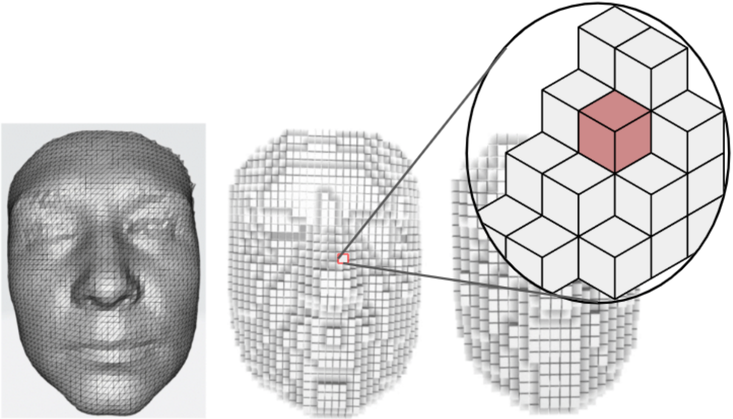


Figure 8: Enlarged view of one voxel from the voxelization block

Dataset

BraTS 2020 Challenge:
Multimodal Brain Tumor Segmentation Challenge Data

Imaging data description

All *MRI modality* scans [a] are available in the following descriptions:

- T1 (*native*)
- T1Gd (*post-contrast T1 weighted*)
- T2 (*T2-weighted*)
- FLAIR (*Fluid-Attenuated Inversion Recovery*)

[a] Provided data is co-registered in the same anatomical template, interpolated to the same resolution (1mm3) and skull stripped.

Visualizing the multimodal scans

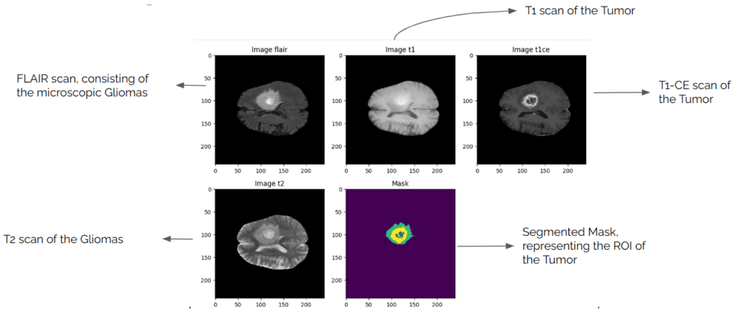


Figure 9: Scans of all the modalities of a MRI image, where the Mask is the segmented tumor

Overview of the Proposed method

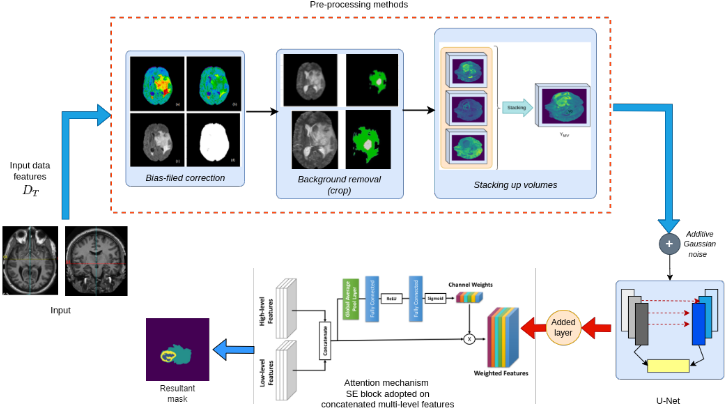


Figure 10: Proposed pipeline of the work

Pre-processing methods

Bias field correction

Background removal ($top_{margin} : bottom_{margin} \approx fixed$)

Stacking combined volumes

Bias field correction

A technique to improve an *MRI/fMRI* image, by removing some *noise signals* in the image generated by the scanner.

$\hat{f}_e(p) \leftarrow \mathbf{0}$ {initialize control point lattice}

$n \leftarrow 1$

for each resolution level **do**

for each iteration at current resolution level **do**

$\widehat{f}_r^n \leftarrow S^*\{\widehat{u}^{n-1} - E[\widehat{u}^n|\widehat{u}^{n-1}]\}$ {calculate residual field}

$\widehat{u}^n \leftarrow \widehat{u}^{n-1} - \widehat{f}_r^n$ {update the estimate of $\widehat{u}$}

$\widehat{f}_e(p) \leftarrow \widehat{f}_e(p)+\widehat{f}_r^n(p)$ {add residual control point lattice to the total estimate}

$n \leftarrow n+1$

end for

$\hat{f}_e(p) \leftarrow \hat{f}_e(p)'$ {refine control point lattice}

end for

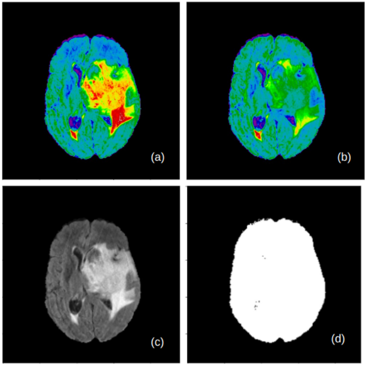


Figure 11: (a) Orignal image, (b) bias field corrected, (c) greyscale image, (d) Transformed Threshold Mask

Background removal (*cropping the data*)

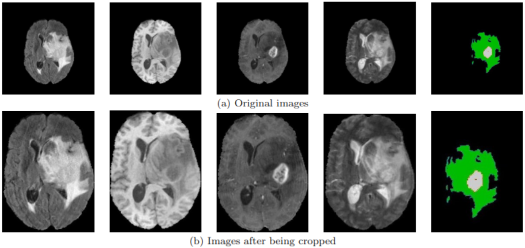


Figure 12: The columns from left to right are FLAIR sequences, T1 sequences, T1-CE sequences, T2 sequences and their corresponding masks

- The brain tumor regions compose only a small fraction of brain MRI images, which can lead to class imbalance problem.
- The background in *multimodality brain* MRI images is completely removed, which can alleviate *class imbalance* problems.

Stacking 2D slices (*creating a combined volume*)

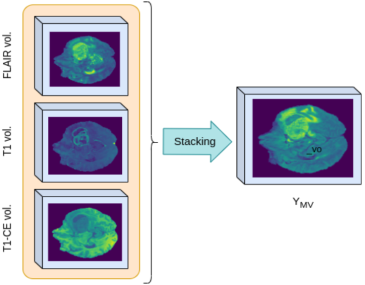


Figure 13: Stacking FLAIR, T1, and T1-CE volumes for generating a stacked mega volume. The resultant Y_{MV} is the 3D volume of stacked multi-volumes, generated for retaining maximum information

- Different information about one's surroundings and core values, resulting in distinct pixel-neighborhoods.
- A mega-volume is generated by stacking and combining the FLAIR, T1, and T1-CE volumes w/ the heterogeneous properties in a multi-volume single scan.

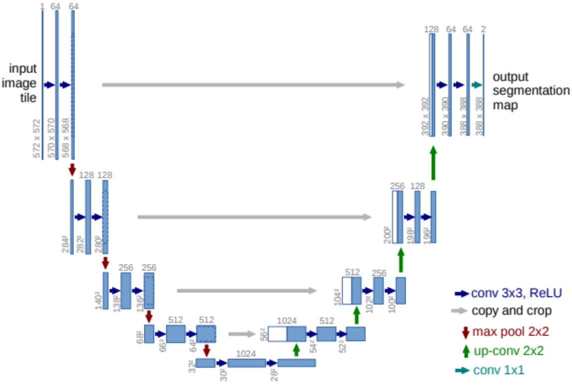
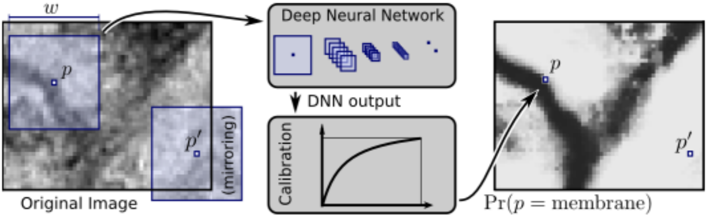


Figure 14: U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.

Olaf Ronneberger et. al (2015)

Motivation Paper

"*Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images*", Ciresan et al, 2012 : paper

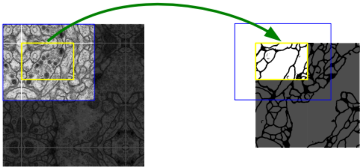


NEURAL INFORMATION PROCESSING SYSTEMS

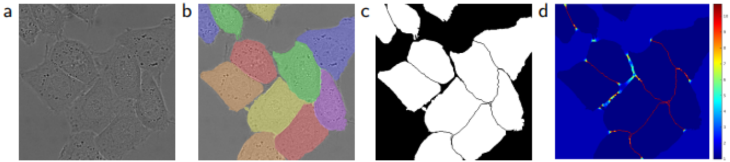
Architecture (*Olaf Ronneberger et. al*)

key design choices

1. Network & training strategy relying on the strong use of data augmentation key Design choices to use the available annotated samples more efficiently.
2. Segmentation + localization
3. Replaces pooling operators with upsampling layers, for performing well w/ fewer training samples.



Olaf Ronneberger et. al (2015)



Results from paper (*Olaf Ronneberger et. al*)

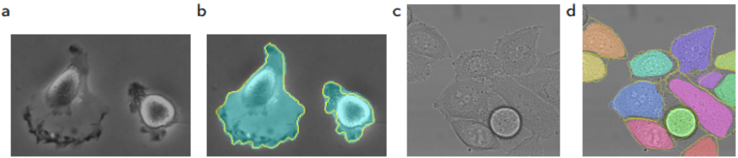


Figure 15: Results on the ISBI cell tracking challenge.

Name	PhC-U373	DIC-HeLa
IMCB-SG (2014)	0.2669	0.2935
KTH-SE (2014)	0.7953	0.4607
HOUS-US (2014)	0.5323	-
second-best 2015	0.83	0.46
u-net (2015)	**0.9203**	**0.7756**

Figure 16: Segmentation results (IOU) on the ISBI cell tracking challenge 2015

Rank	Group name	Warping Error	Rand Error	Pixel Error
	** human values **	0.000005	0.0021	0.0010
1.	u-net	**0.000353**	0.0382	0.0611
2.	DIVE-SCI	0.000355	0.0305	0.0584
3.	IDSIA [1]	0.000420	0.0504	0.0613
4.	DIVE	0.000430	0.0545	**0.0582**
⋮				
10.	IDSIA-SCI	0.000653	**0.0189**	0.1027

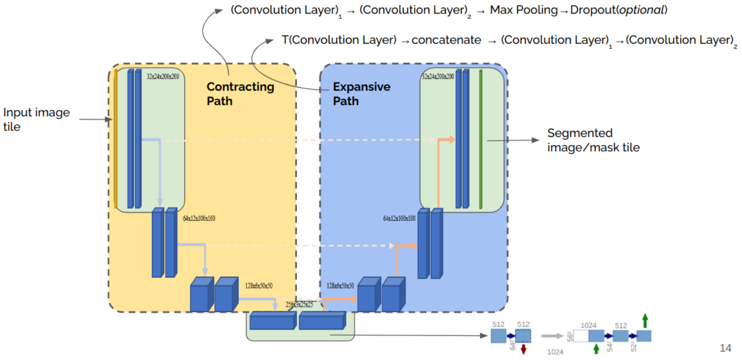
Figure 17: Ranking on the EM segmentation challenge (march 6th, 2015), sorted by warping error

Benchmark validation dataset used:

- PhC-U373
- DIC-Hela

Network Analogy

How UNet works?



Contracting Path

Downsampling (Encoder)

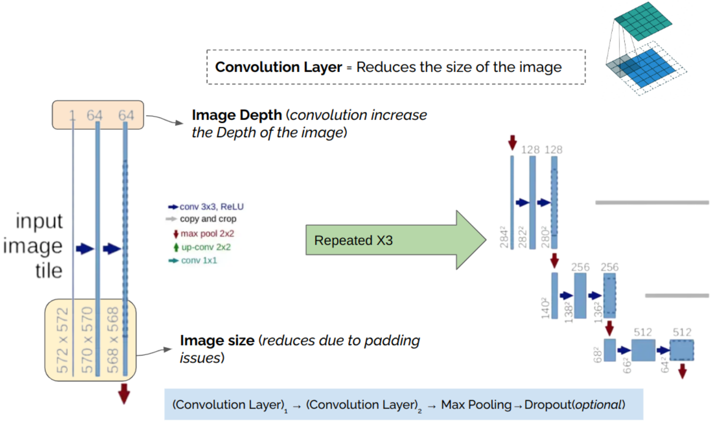


Figure 18: Contracting path

Expanding Path

Upsampling (Decoder)

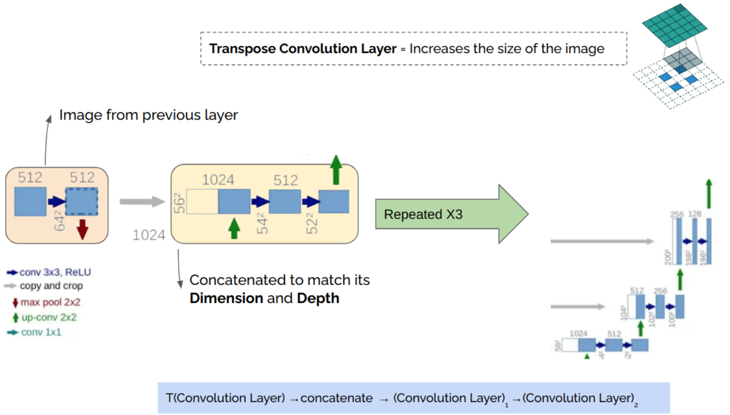


Figure 19: Expansive path

Gaussian Noise – *as additive input noise*

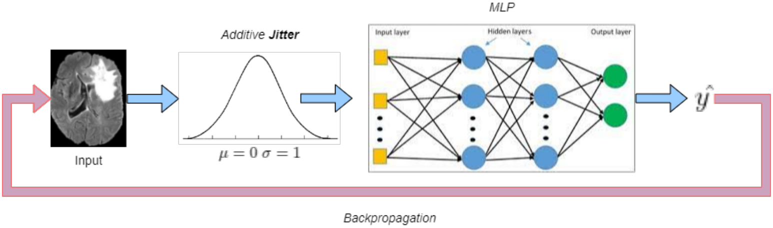
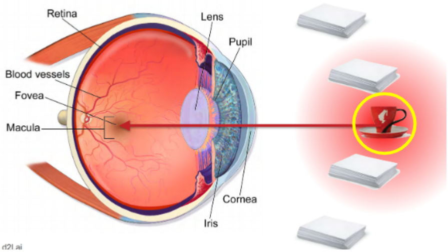


Figure 20: Additive gaussian noise to the proposed network

Reason

- Sometime dataset can trick the model in learning complex *functions/patterns*
- Adding *noise* during training can make the training process more robust and reduce *generalization error*
- In other way, adding noise to input samples is a simple form of *data augmentation*.

Attentions!!



“Attention” is a psychological phenomenon, where the human mind focuses on an object while de-focusing it from the other things present in the environment.

Attention in computer vision

Attention was presented by Dzmitry Bahdanau, et al. in their 2014 paper "Neural Machine Translation by Jointly Learning to Align and Translate", which reads as a natural extension of their previous work on the *Encoder-Decoder* model.

This very paper laid the foundation of the famous paper "Attention is All You Need" by Vaswani et al., on **transformers** that revolutionized the deep learning arena with the concept of **parallel processing** of words instead of processing them **sequentially**.

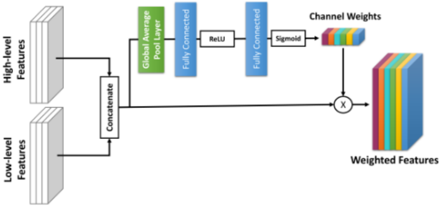


Figure 21: A Generalised architecture of Attention mechanism

History and Evolution of "Attention"

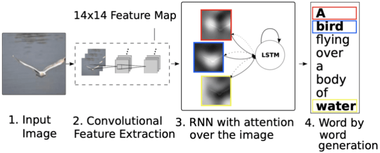
Evolution of Attentions

Attention in NLP

- Initially emerged to solve problems in NLP (*in handling sequence of words*)
- Deals w/ the drawbacks of **seq2seq** model of dealing to retain information from long sequences.
- Often it forgets the earlier elements of the input sequence once it has processed the complete sequence. The attention mechanism was created to resolve this problem of long dependencies.

Attention in computer vision

- Attention has found its way into computer vision, including some applications such as image classification, image segmentation, and image captioning.
- in some cases, *CNNs* are used to capture the salient features and then *RNNs* are used to interpret them.



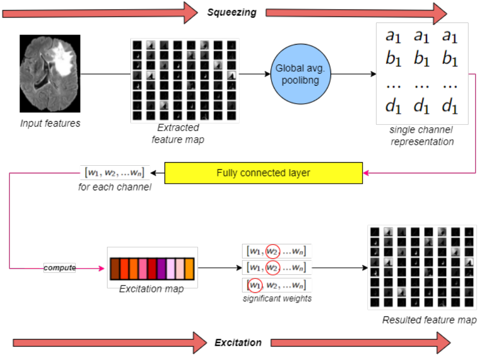


Figure 22: SE block architecture

The **Squeeze-and-Excitation Block**is an architectural unit designed to improve the representational power of a network by enabling it to perform dynamic channel-wise feature re-calibration.

** Representation power is related to the ability of a neural network to assign proper labels to a particular instance and create well-defined accurate decision boundaries for that class.

$$L_{dice} = \frac{2 * \sum p_{true} * p_{pred}}{\sum p_{true}^2 + \sum p_{pred}^2 + \epsilon}$$

$$L_{focal} = -\sum_{i=1}^{n} \alpha_i (i - p_i)^{\gamma} log_b(p_i)$$

Dice Loss

1. Dice Loss is widely used in medical image segmentation tasks to address the **data imbalance problem**
2. It only addresses the imbalance problem between **foreground** and **background**

Focal Loss

1. Addresses the issue of the **class imbalance** problem.
2. A **modulation term** applied to the **Cross-Entropy loss** function, making it easy to learn on harder problems

```
# defining the weighted losses
import segmentation_models_3D as sm

def loss_fn():
    wt0, wt1, wt2, wt3 = 0.25,0.25,0.25,0.25

    dice_loss = sm.losses.DiceLoss(class_weights=np.array([
    wt0, wt1, wt2, wt3
    ]))
    focal_loss = sm.losses.CategoricalFocalLoss()
    total_loss = dice_loss + (1 * focal_loss)
```

Evaluation metric

IoU (Intersection over Union) is chosen as choice of metric for evaluating the result.

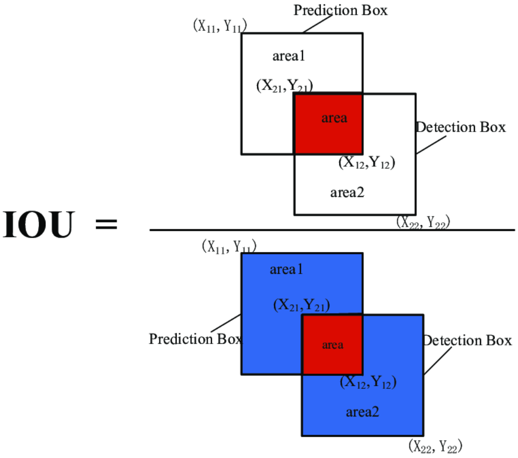
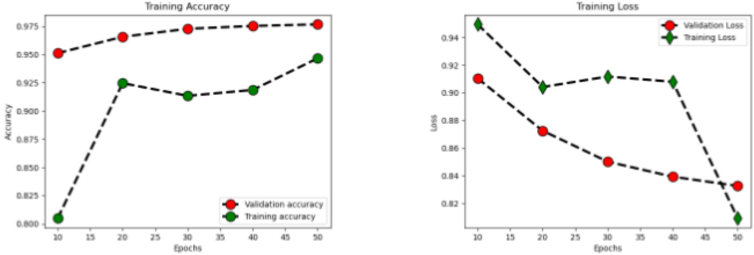


Figure 23: Intersection over Union - IoU

Results



(a) Training and validation accuracy

(b) Training and validation loss

Figure 24: Graph for model's performance

Table 2: Performance statistics for every 10 Epochs

Epochs	Loss	Accuracy	Val. Loss	Val. Acc.	IoU Score	Val. IoU Score
10	91.05	95.12	94.96	80.48	33.98	28.56
20	87.25	96.57	90.42	92.46	46.06	37.84
30	85.01	97.28	91.18	91.34	52.89	36.81
40	83.92	97.53	90.80	91.86	56.78	37.46
50	83.25	97.67	80.90	94.67	59.48	42.43

... accuracy of *97.67%* w/ a loss of *83.25%*

References

Md Khairul Islam et. al.
Brain tumor detection in MR image using superpixels, principal component analysis and template-based K-means clustering algorithm
Machine Learning with Applications..

Anuja Arora et. al.
Brain Tumor Segmentation of MRI Images Using Processed Image Driven U-Net Architecture
MDPI.

Mehrdad Noori et. al.
Attention-Guided Version of 2D UNet for Automatic Brain Tumor Segmentation
arXiv.

Navchetan Awasthi et. al.
Multi-Thresold Attention U-Net(MTAU) based Model for Multimodal Brain Tumor Segmentation in MRI scans
arXiv.

Jie Hu et. al.
Squeeze-and-Excitation Networks
arXiv:1709.01507.

Olaf Ronneberger et. al,
U-Net: Convolutional Networks for Biomedical Image Segmentation
arXiv:1505.04597.

Venkatesan Rajinikanth et. al
A Customized VGG19 Network with Concatenation of Deep and Handcrafted Features for Brain Tumor Detection
MDPI.

..... and more

Thank You

Any Questions!!