Head and Neck CT Segmentation Using Deep Learning

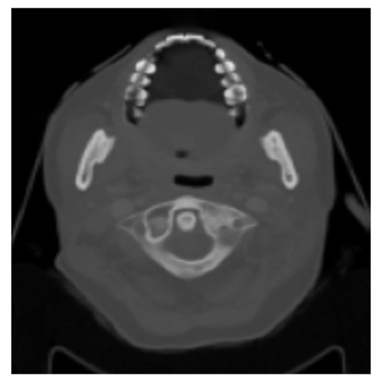
NYU Center for Data Science | NYU Langone Health

Members: Pengyun Ding, Tianyu Zhang

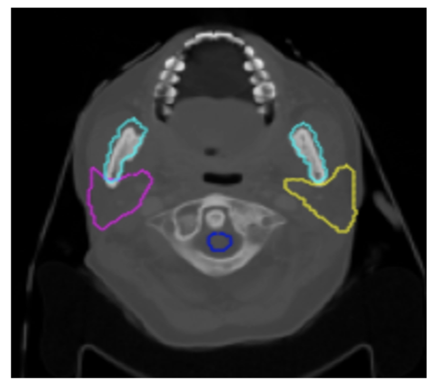
Mentor: Yuan Ye

Background

- **Radiotherapy** is one of the cornerstone therapies for head and neck cancers, and it works by **delivering targeted high dose radiation to patients' tumors to shrink and kill cancer cells.** However, this treatment can be deleterious to the healthy tissues surrounding the tumor and misguided radiation can cause treatment related side effects, which may increase the risk of treatment failure. Therefore, **avoiding the surrounding organs at risk (OARs)** is a critical part in radiation therapy.
- In current clinical practice, OARs are outlined manually on CT scans of the patient's head and neck, which is **very time consuming and can be inaccurate.**
- An improved automatic segmentation algorithm for OAR delineation would **increase the efficiency and standardization of radiation treatment** planning for head and neck cancers.



CT image slice

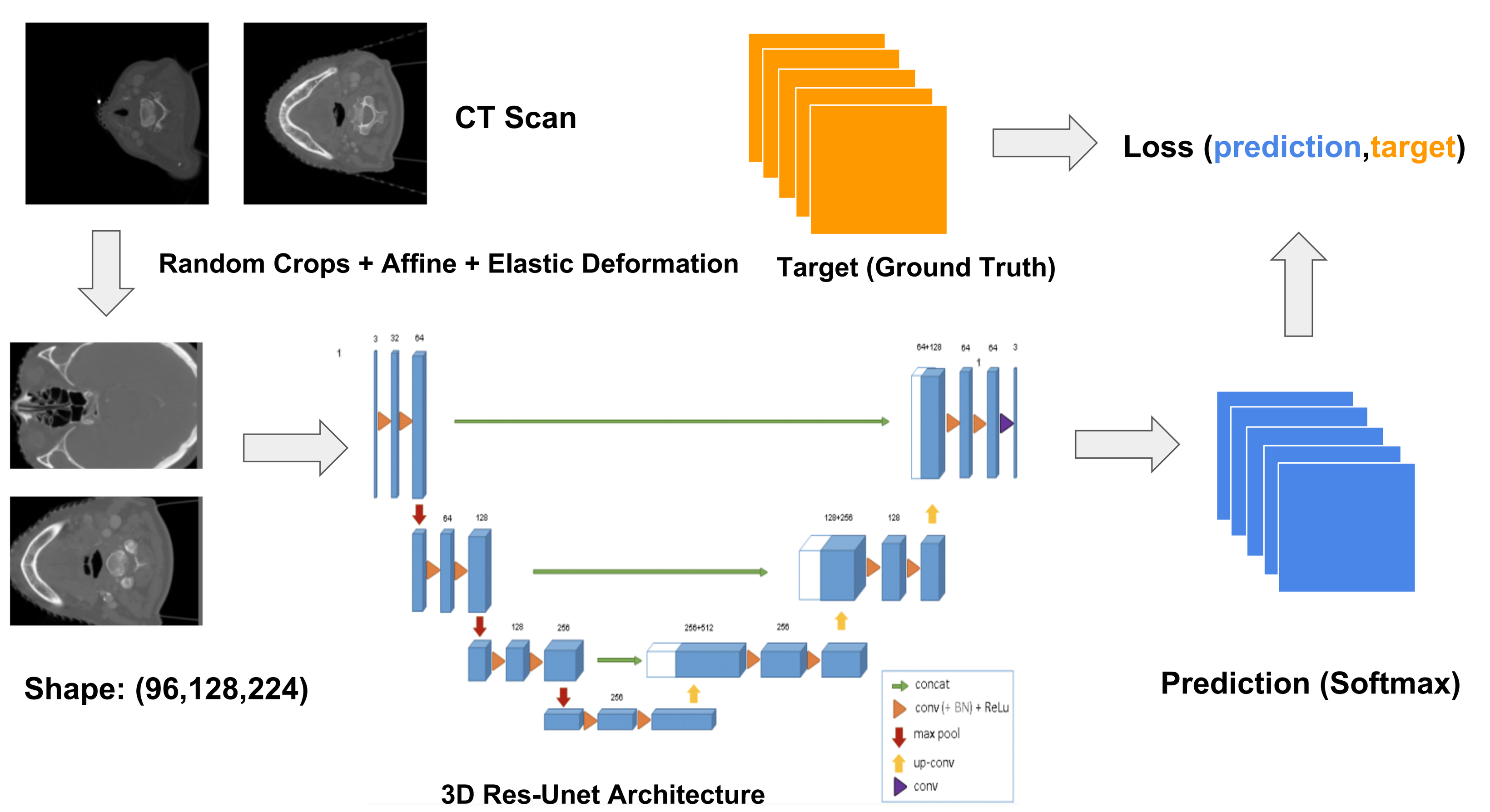


CT Image Slice with OARs outlined manually

Introduction

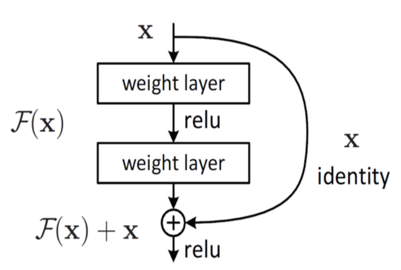
- **Datasets:**
  - **Dataset 1 (MICCAI imaging challenge in 2015):** It includes 3D CT scans from 40 patients with head and neck cancer who were treated on a clinical trial.
  - **Dataset 2 (TCIA HN1 study):** It is comprised of 3D CT scans from 137 head and neck patients who received radiation from two major academic medical centers in the Netherlands.
  - **Dataset 3 (NYU Langone):** It is a de-identified dataset of head and neck patients treated in the NYU Langone Department of Radiation Oncology, which includes 3D CT scans with the same manually segmented OARS from an additional 200 patients
  - Each CT scan contains over a hundred axial slices which is manually annotated at the pixel level for 9 organs at risk (**Brain Stem, Chiasm, Mandible, Optic Nerve Left, Optic Nerve Right, Parotid Left, Parotid Right, Submandibular Left, Submandibular Right**).
  - The percentage of foreground voxels for each of the nine organs ranges from 0.4% to 35%, which causes **data imbalance issue.**
- **Contributions:**
  - Use a deep learning framework by leveraging **3D-Unet** and its refined versions (**Resnet, Transformer**) to segment the organs at risk.
  - Apply novel training methodology by manipulating **various loss functions** and implement **pixel weight** into loss function to settle the **imbalanced data issue**

Model 1 Architecture



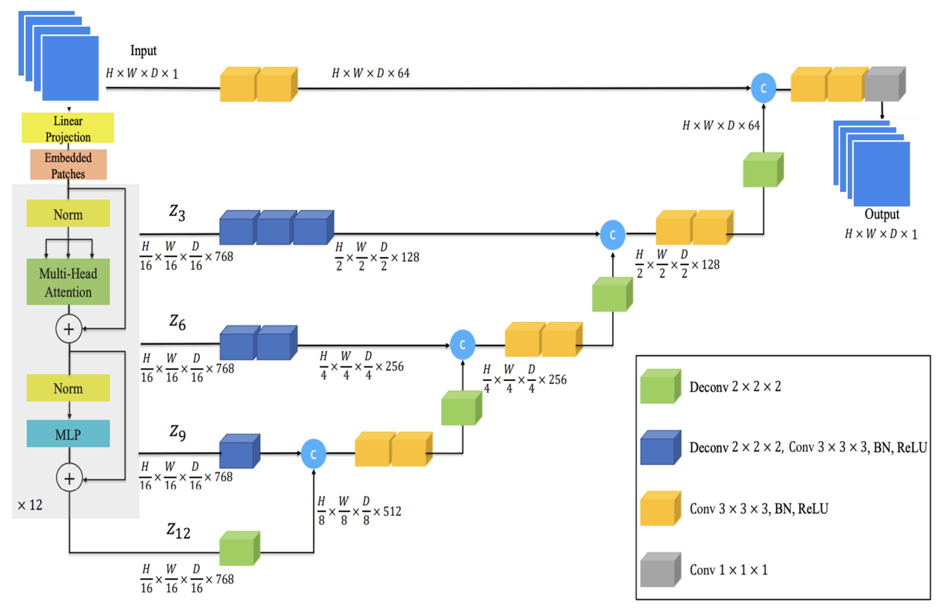
Overview of Our Training Framework

- **3D Res-Unet Model:** We use 3D Res-Unet Model which enhances the traditional 3D-Unet by replacing U-Net blocks by Residual Block in order to improve the ability of generalizing and make network easier to optimize. It takes fixed-size 96 x 128 x 224, 3 dimensional CT images from data preprocessing (Random Crops + Affine + Elastic Deformation) and produces multi-class semantic segmentation.

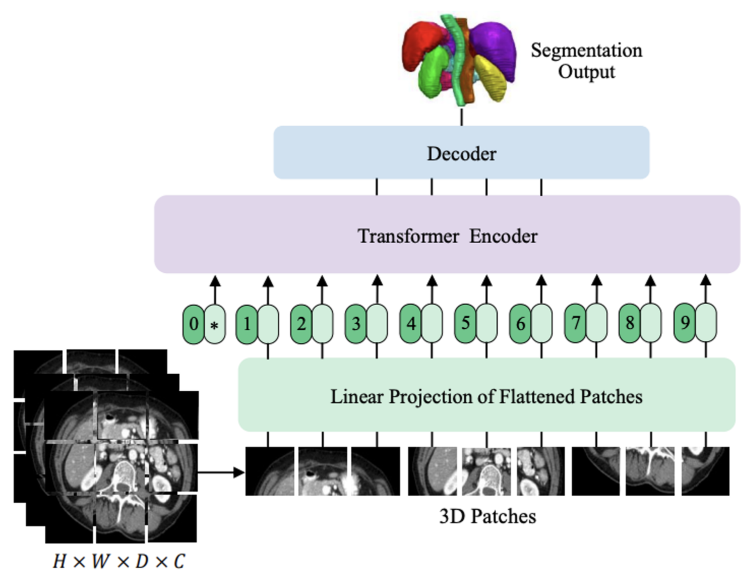


The Residual Block

Model 2 Architecture



UNEt TRansformers (UNETR) Architecture



Overview of UNETR

- **UNEt TRansformers (UNETR) :** We use UNETER to replace 3D Res-Unet in our whole framework to create our second model. It utilizes a transformer as the encoder to learn sequence representations of the input volume and is proven to be able to capture long-range dependencies effectively.

Training Methodology

- **Loss Functions:**

$$\text{Weighted Cross Entropy} = \begin{cases} -w\log p & y=1 \\ -\log(1-p) & \text{otherwise} \end{cases};\qquad \text{Focal Loss} = \begin{cases} -(1-p)^{\gamma}\log p & y=1 \\ -(p)^{\gamma}\log(1-p) & \text{otherwise} \end{cases};\gamma=2$$

$$w = \frac{N-\sum_n p_n}{\sum_n p_n}\text{(then scaled such that w for all classes sum up to 1)}$$

$$\text{Generalized Dice Loss} = 1 - 2\frac{\sum_{labels} w_l \sum_{pixels} y_{li}p_{li}}{\sum_{lables} w_l \sum_{pixels} y_{li}+p_{li}};\ w_l = 1/\left(\sum_{pixels} y_{li}\right)^2 \qquad \text{Dice Loss} = 1 - \frac{2\sum_{\text{pixels}} p_i\cdot y_i}{\sum_{\text{pixels}} p_i + \sum_{\text{pixels}} y_i}$$

- **Dynamically Weighted Loss Function:** We explore a wide range of dynamically weighted loss functions and achieve the following best setting. (Stage switches according to the performance on validation set)
  Stage1: Generalized Dice Loss+ 1* Weighted Cross Entropy + 0.1*Focal Loss
  Stage2: Generalized Dice Loss + 0.1*Weighted Cross Entropy + 0.5*Focal Loss
  Stage3: Generalized Dice Loss + 0.01*Weighted Cross Entropy + 1*Focal Loss
- **Data Augmentation:**
  Random Elastic Deformation: A random displacement is assigned to a coarse grid of control points around and inside the image. The displacement at each voxel is interpolated from the coarse grid using cubic B-splines.
  Random Affine: A geometric transformation that preserves lines and parallelism (but not necessarily distances and angles).

Results

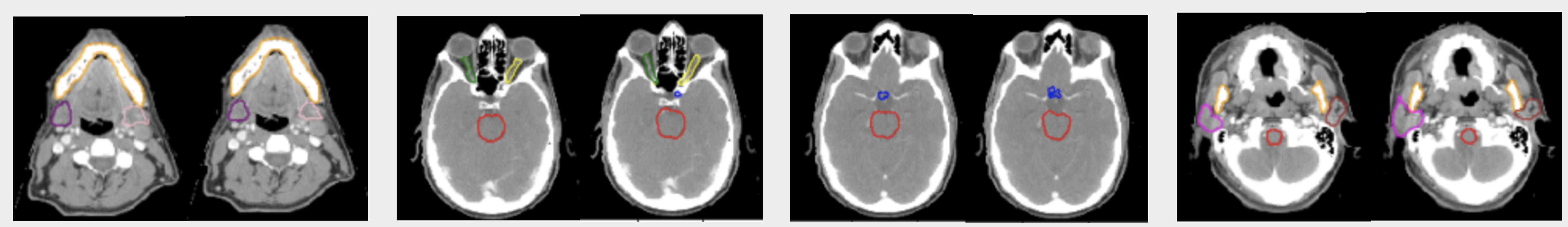
- **Evaluation Metrics:** We use the Dice score as the metric to evaluate the model's performance

Name	Loss Function	Dice Score
Res3DUnet	Dice Loss	0.3180
Res3DUnet	Dice Loss + Focal Loss	0.3963
Res3DUnet	Dice Loss + Weighted Cross-Entropy	0.3560
Res3DUnet	Generalized Dice Loss + Focal Loss+ Weighted Cross-Entropy	0.5744
Res3DUnet	Generalized Dice Loss + Focal Loss+ Weighted Cross-Entropy (Dynamic Weight Loss)	**0.6077**

Name	Data Augmentation	Dice Score
Res3DUnet	RandomAffine(P=0.5)	0.6706
Res3DUnet	RandomAffine(P=1)	0.6928
Res3DUnet	RandomElasticDeformation(P=0.5)	0.6318
Res3DUnet	RandomElasticDeformation(P=1)	0.6412
Res3DUnet	RandomElasticDeformation(P=0.5) +RandomAffine(P=0.5)	0.7155
Res3DUnet	RandomElasticDeformation(P=0.3) +RandomAffine(P=0.7)	**0.7229**

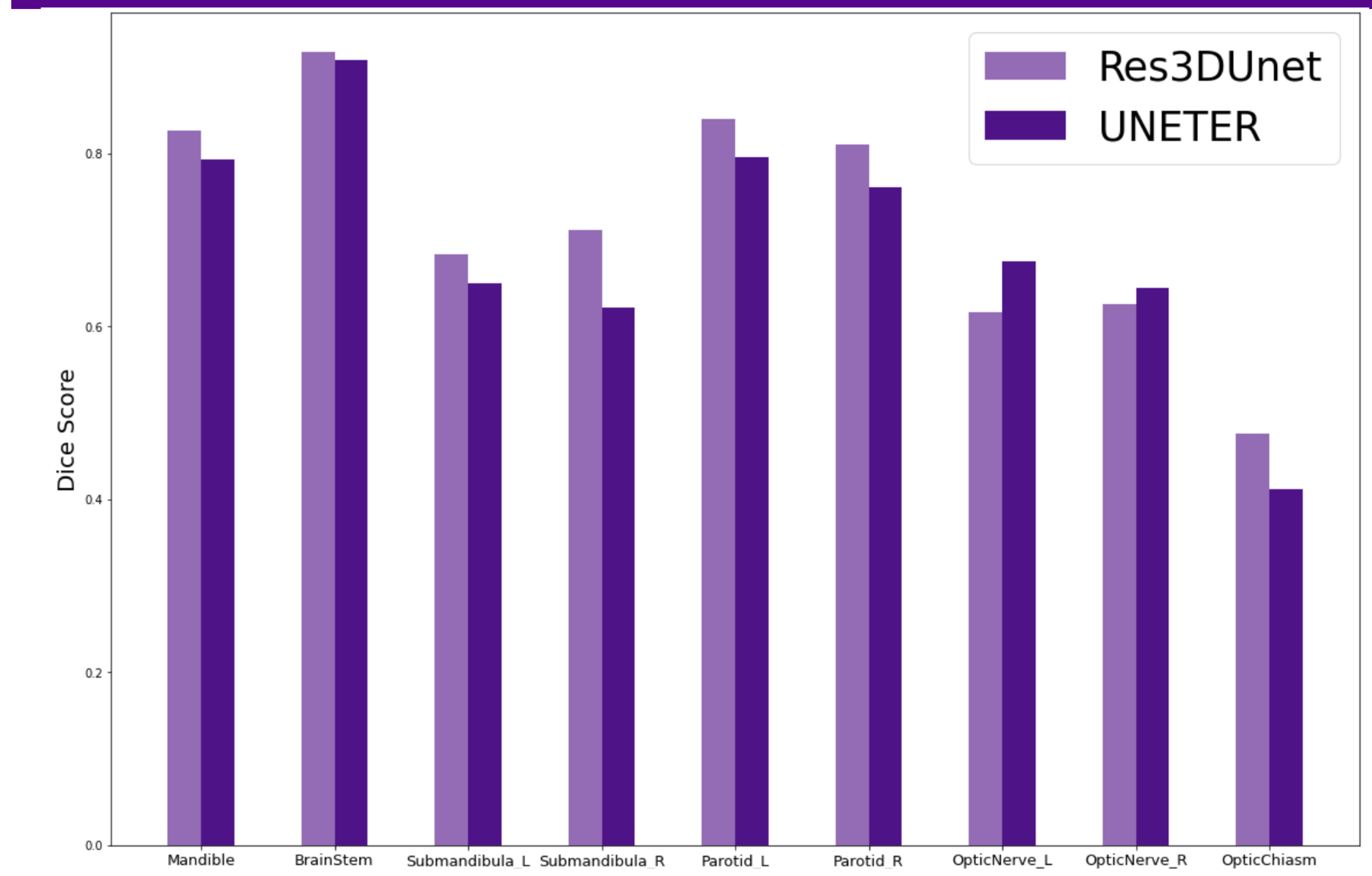
(Training using MICCAI train data, Performance on MICCAI test data, P stands fo the apply probability)

- ResUNet trained with dynamically weighted loss function of Generalized Dice Loss, Focal Loss and Weighted Cross Entropy, together with data augmentation of Random Elastic deformation and Random Affine performs the best.
- Sample segments are visualized below for the best model of ResUNet.

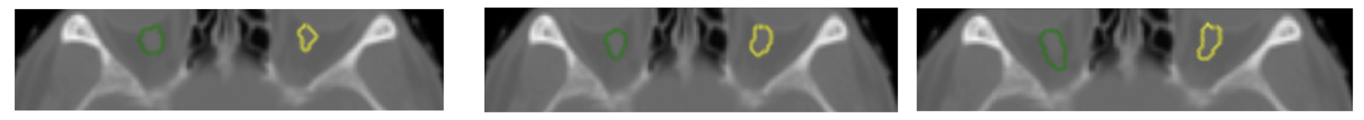


Truth (left)
Prediction (right)
BrainStem (red)
OpticChiasm (blue)
Mandible (orange)
OpticNerve_Left (green)
OpticNerve_Right (yellow)
Parotid_Left (margenta)
Parotid_Right (brown)
Submandibula_Left (purple)
Submandibula_Right (pink)

Transformer vs Resnet



- UNETER is able to achieve 69.03% mean dice score and outperforms Res3DUnet in some tasks (Optic Nerve).
- Though having similar Dice Score performance with Res3DUnet, UNETER contours differently due to its distinct Unet Structure with idea of transformer adopted. It provides us with flexibility to ensemble these two models.
- Sample segments of Optic Nerve are visualized below for the best models of ResUNet and UNETER.



Truth (left), UNETER (middle), Res3DUnet (right)
OpticNerve_Left (green) ,OpticNerve_Right (yellow)

Conclusion and Future Work

- **Conclusion**
  - Choices of loss function are important for alleviating class imbalance issue.
  - Data augmentation is a useful tool to help model generalize better.
  - Our Res3DUnet model with dynamically weighted loss function and data augmentation is the best model with mean dice score of 72.29%.
  - Our UNETER model achieves 69.03% mean dice score with limited time to fine tune and outperforms Res3DUnet in some symmetric tasks.
- **Future Work**
  - Experiment on the entire datasets of collected (Dataset1+Dataset2+Dataset3) with our best models to increase accuracy.
  - Fine tune UNETER model for better performance.
  - Explore the potential ensemble method with UNETER and Res3DUnet to further improve performance.

Reference

- Hatamizadeh, Ali, et al. "Unetr: Transformers for 3d medical image segmentation." *arXiv preprint arXiv:2103.10504* (2021).
- Çiçek, Özgün, et al. "3D U-Net: learning dense volumetric segmentation from sparse annotation." *International conference on medical image computing and computer-assisted intervention*. Springer, Cham, 2016.
- He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- Raudaschl, P. F., Zaffino, P., Sharp, G. C., Spadea, M. F., Chen, A., Dawant, B. M., ... & Jung, F. (2017).Evaluation of segmentation methods on head and neck CT: Auto-segmentation challenge 2015.Medical Physics, 44(5), 2020-2036.