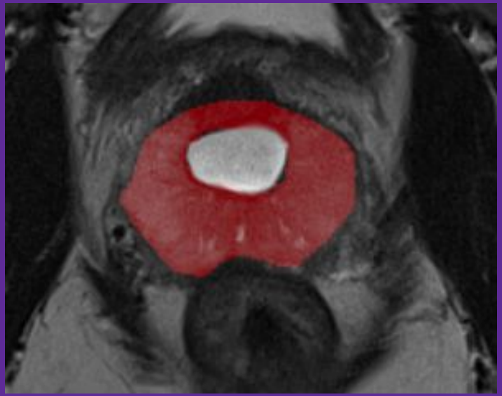
Semantic Segmentation of Prostate Gland from MR Images across Different Loss Functions

Çerağ Oğuztüzün

CSDS600 Project 1 Presentation


Motivation

Clinical Problem for Prostate Gland Segmentation

- Can be used to monitor the growth of diseases
- Can be used to control medication dosages

Image Analysis Problem for Prostate Gland Segmentation

- Can be challenging due to artifacts in the images
- Each pixel of an image is assigned a specific class
- The classification output is a mask image

Methods

Description of Data

- PROSTATEx Gland Segmentations [1]
- Manual segmentations of prostate glands [2]
- Mask images for 204 patients
- Axial T2 weighted MRI scans
- **4166 images in total**
- 80 to 20 ratio for partitioning the dataset into training and validation sets.

Data Augmentations:

- Resizing
- Rotating
- Flipping Horizontally and Vertically

U-Net Model

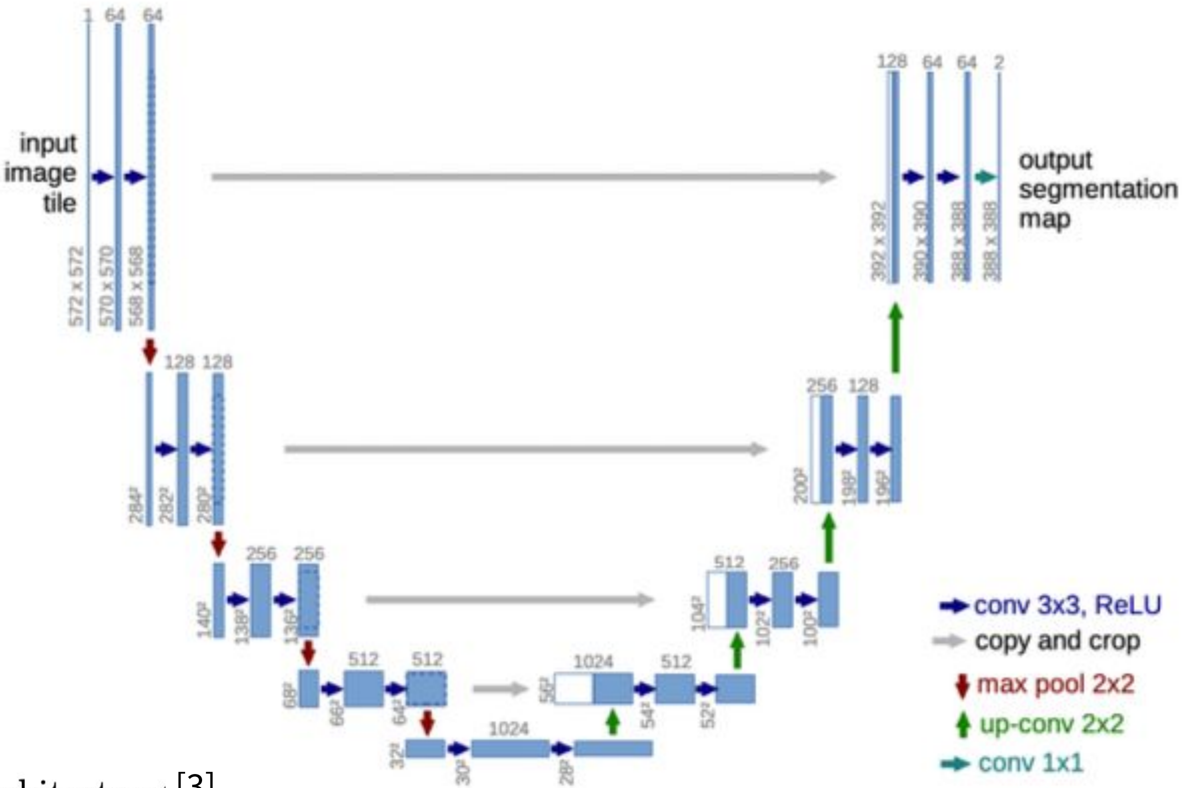


Fig. 1: U-Net Model Architecture [3]

U-Net Model

Skip Connections:

- Lost details are re-injected

Contracting Path:

- Learns what is in the image
- Loses spatial information

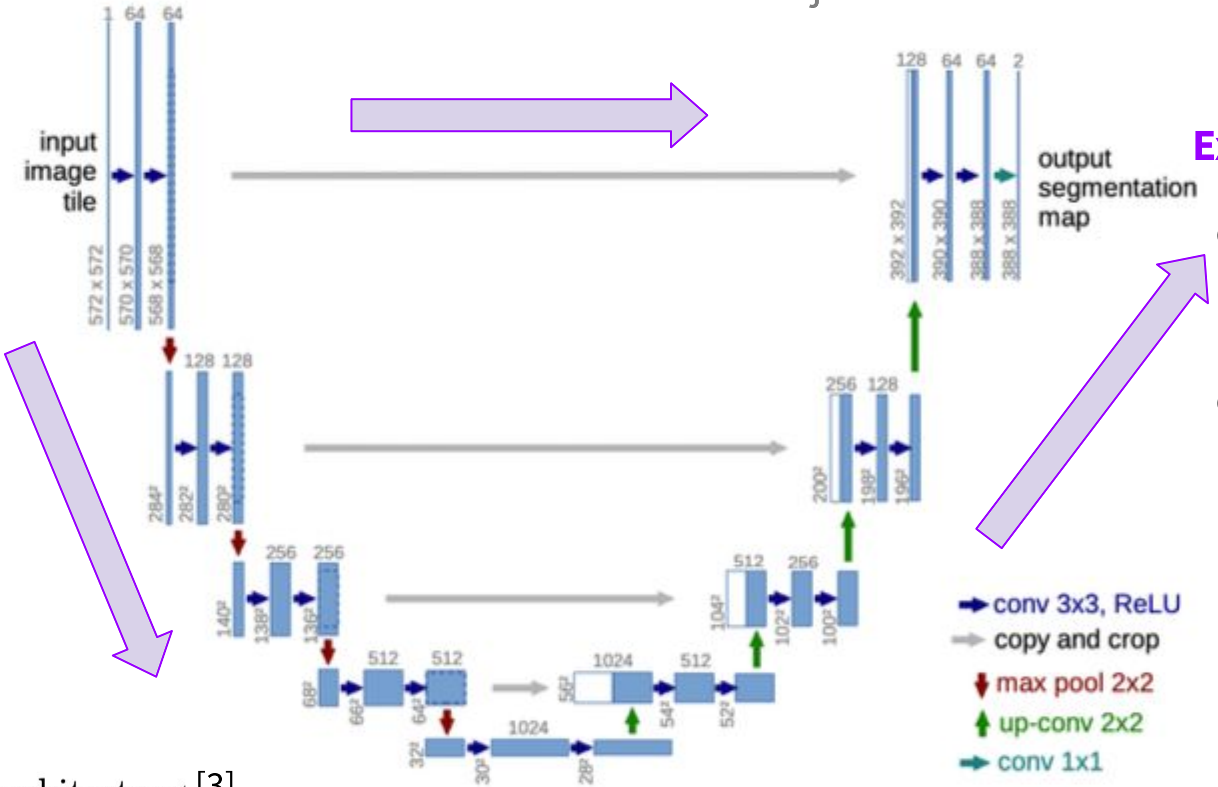


Fig. 1: U-Net Model Architecture [3]

Expansive Path:

- Recovers information lost
- High-level features are propagated into each original pixel

Loss Functions

- **Binary Cross Entropy**

$$BCE = -\sum_{i=1}^{C'=2} t_i log(s_i) = -t_1 log(s_1) - (1 - t_1) log(t - s_1)$$

C’ := number of classes (2 as pixels can be 0 or 1)

ti := target for class i

si := score for class i

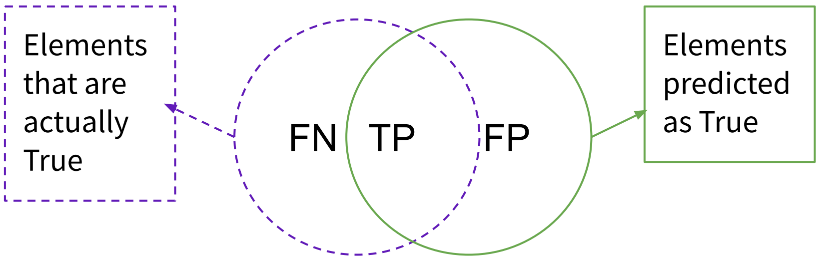
- **Tversky Loss [4]**

$$Tversky = \frac{TP}{(TP + \alpha FP + \beta FN)}$$

α := weight of penalty for FP   β := weight of penalty for FN

α+β=1   high β ~ high recall ~ low precision

For α=β=0.5 Tversky Loss = Dice Loss ≡ F1 Score



Union = FN + TP + FP

Intersection = TP

Training of the 8 Models

Early stopping if the accuracy does not improve by 0.3 for at least 3 epochs.

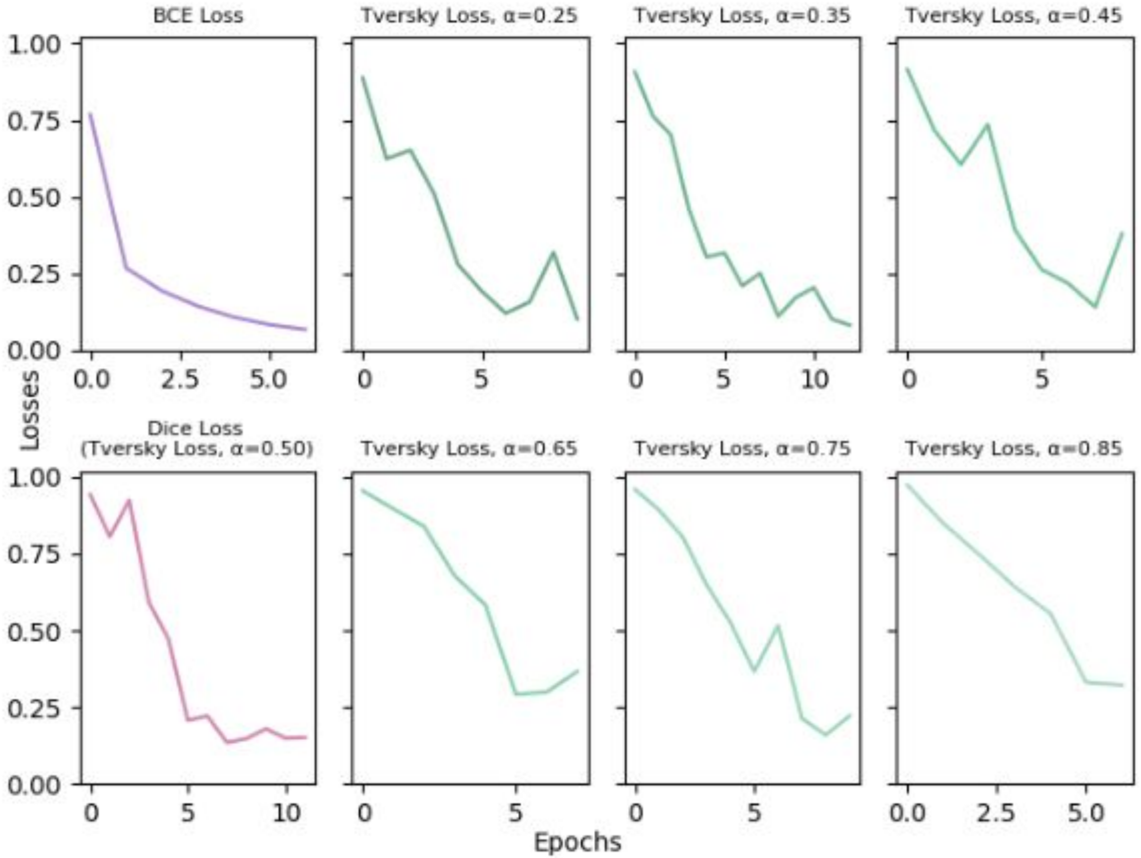


Fig. 2: **Losses** plotted for each model per epoch. Due to early stopping not all models trained for the same time, but all models reached to a convergence.

Hypothesis:

The model with α=0.25 will perform the best segmentation. False Negatives will be penalized more than False Positives.

It is intuitive that False Negatives in medical imaging pose a greater risk since they may cause the reader not to recognize a patient's potential threat.

Segmenting not only the periphery but the contours of the prostate gland is important.

The results of the best model will also have higher recall and a lower precision value.

Evaluation Metrics

- **Dice Similarity Coefficient**

a measure of the overlap between a segmentation result and its ground truth.

$$Dice(A, B) = \frac{2 \times |A \cap B|}{(A + B)}$$

A := pixel values of predicted segmentation result images

B := pixel values of the target images

- **F1 Score**

combines precision and recall into a single measure that captures both characteristics.

$$F1Score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

Results

Table 1: **Performance metrics results** on validation set for models trained using different loss functions are recorded. The best values for each metric have been highlighted in bold.

Loss Function	Dice Score	Precision	Recall	F1 Score
BCE Loss	0.85551	0.81207	0.91116	0.85876
Tversky Loss $\alpha = 0.25$ $\beta = 0.75$	0.85253	0.81798	0.92828	0.86965
Tversky Loss $\alpha = 0.35$ $\beta = 0.65$	0.85776	**0.91000**	0.92189	**0.91590**
Tversky Loss $\alpha = 0.45$ $\beta = 0.55$	0.83326	0.43057	**0.97267**	0.59691
Dice Loss (Tversky Loss $\alpha = 0.50$ $\beta = 0.50$)	0.86958	0.76809	0.94913	0.84907
Tversky Loss $\alpha = 0.65$ $\beta = 0.35$	0.86581	0.70920	0.52906	0.60603
Tversky Loss $\alpha = 0.75$ $\beta = 0.25$	**0.89506**	0.76617	0.81707	0.79081
Tversky Loss $\alpha = 0.85$ $\beta = 0.15$	0.88112	0.64821	0.91519	0.75890

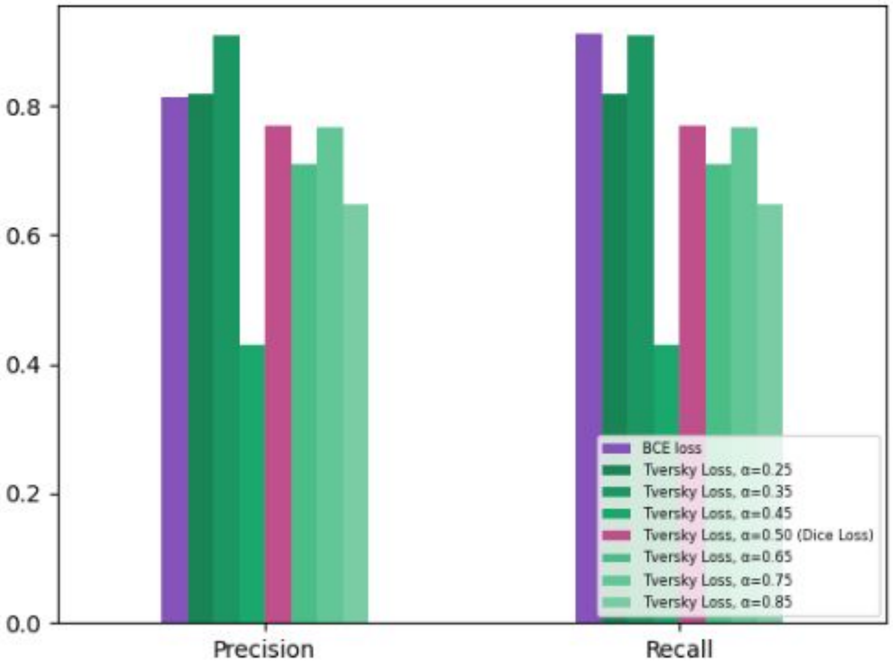


Fig. 3: **Precision and Recall** plotted for each model.

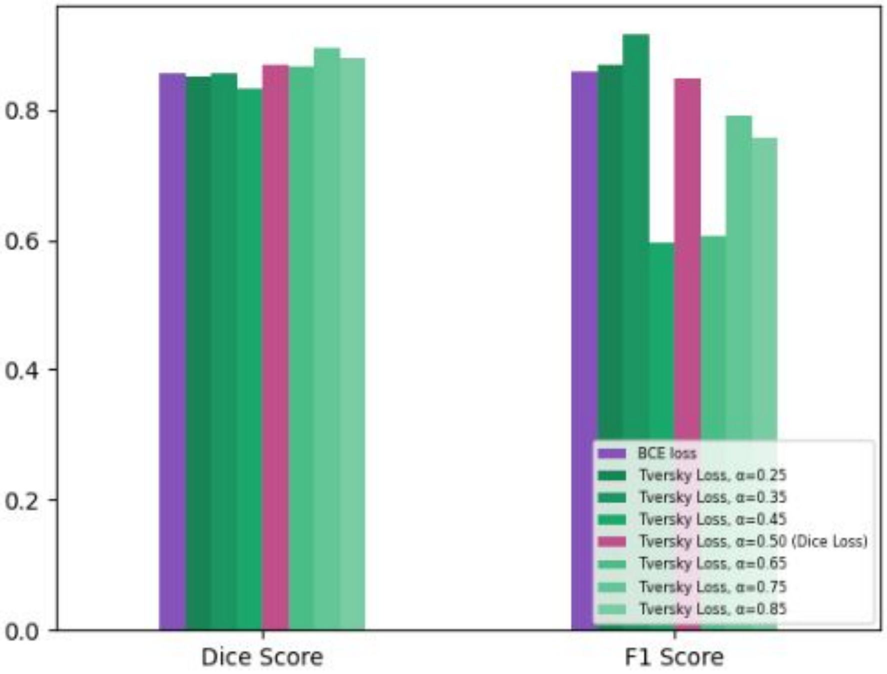


Fig. 4: **Dice and F1 Scores** plotted for each model.

Lighter color is given to models with higher α values, dark red color is used to indicate equal α and β values.

- As α increase ~ FP penalty increases ~ boundaries get more conserved ~ more concave mask.

- As β increase ~ FN penalty increases ~ boundaries get more spread out ~ more convex mask.

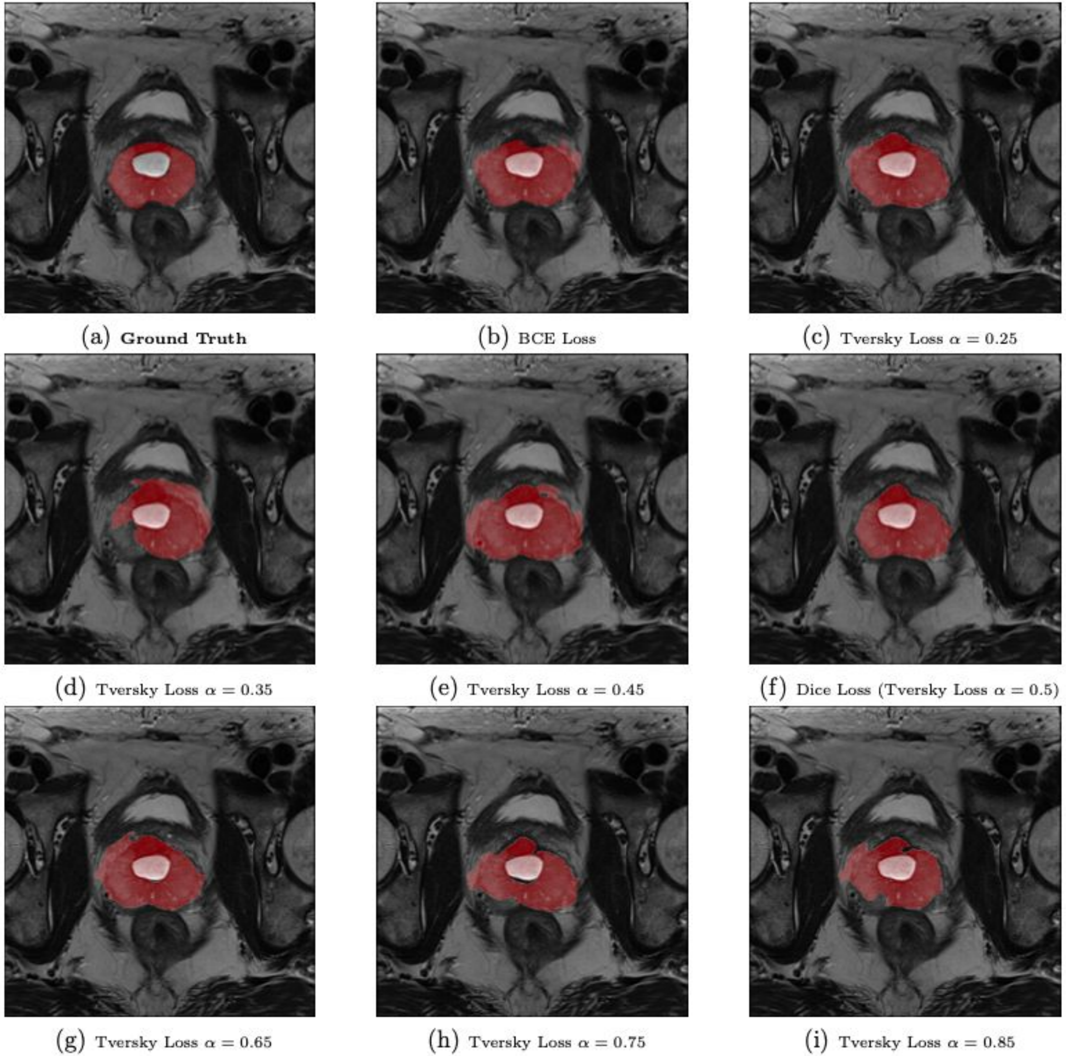


Fig. 5: **Scan Images and Segmentation Overlays** For each prostate gland MR scan image, the segmentation predictions made by the models trained using different loss functions. This particular MR scan has the Patient ID 0200, the sequence name z013.

Reflection on Hypothesis:

Model with Tversky Loss $\alpha = 0.75$ performed better than Tversky Loss $\alpha = 0.25$.

The precision-recall tradeoff can fine tune the model performance depending on the dataset.

The trade-off between penalizing FP and FN pixels can change the contour in a way that can affect the progress control of the patient's disease.

More robust evaluation metrics that use tradeoff of concepts with great detail and attention must be used in practice.

Thanks for Listening,
Q&A

References

[1] Spie-aapm-nci prostatex challenges (prostatex) the cancer imaging archive (tcia) public access -cancer imaging archive wiki.

[2] R. Cuocolo, A. Stanzione, A. Castaldo, D. R. De Lucia and M. Imbriaco, Quality control and whole-gland, zonal and lesion annotations for the prostatex challenge public dataset, European Journal of Radiology 138, p. 109647 (2021).

[3] O. Ronneberger, Invited talk: U-net convolutional networks for biomedical image segmentation, Informatik aktuell , p. 3–3 (2017).

[4] S. S. Salehi, D. Erdogmus and A. Gholipour, Tversky loss function for image segmentation using 3d fully convolutional deep networks, Machine Learning in Medical Imaging , p. 379–387 (2017).