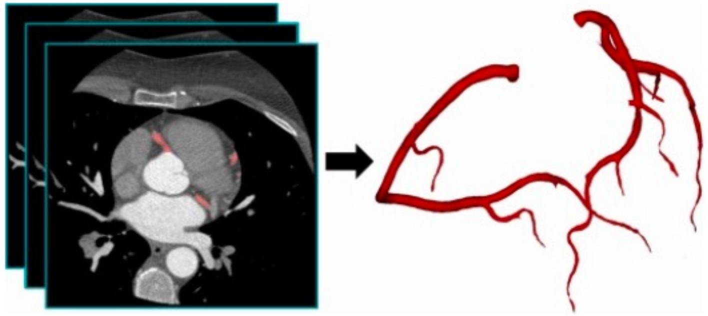
A MODEST TAKE ON AUTOMATED SEGMENTATION OF CORONARY ARTERIES

DIONYSIOS RIGATOS

TDT4265 – COMPUTER VISION & DEEP LEARNING

TASK & APPROACH

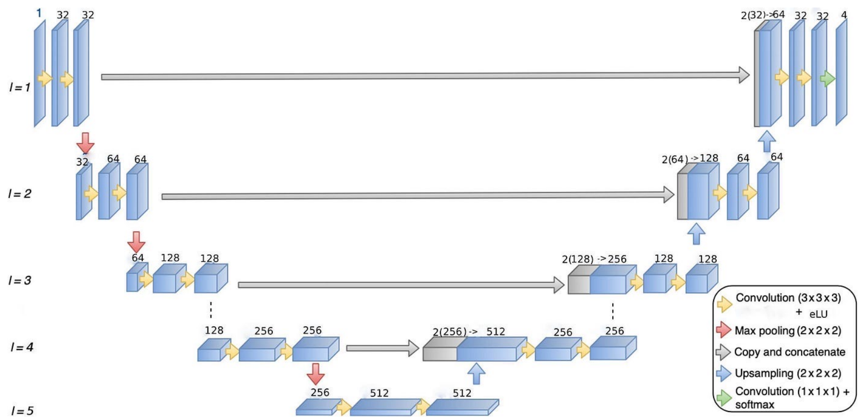
- The ASOCA challenge
  - Online biomedical imaging competition from 2020
  - Semantic 3D segmentation of Computed Tomography Coronary Angiography (CTCA) scans



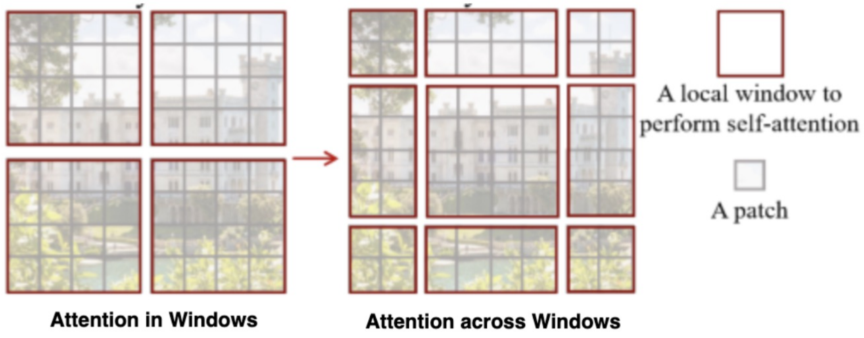
- **Solution**: Swin-UNETRv2
  - MONAI Framework
    - PyTorch wrapper
    - Preprocessing, models, evaluation, optimization and more

SHIFTED WINDOW UNET TRANSFORMER I

- UNET
  - Default choice for semantic segmentation in medical imaging
  - Limited kernel size, long-range information loss
- Swin-UNET Transformer (v2)
  - Transformers excellent at keeping long-range information; problem reformulated as seq2seq
  - **S**hifted **Win**dows: Key to ViTs
    - Compute attention for collection of patches instead of whole image
    - Cyclic shift for cross-window connections
  - **v2**: Reintroduce convolutions before blocks

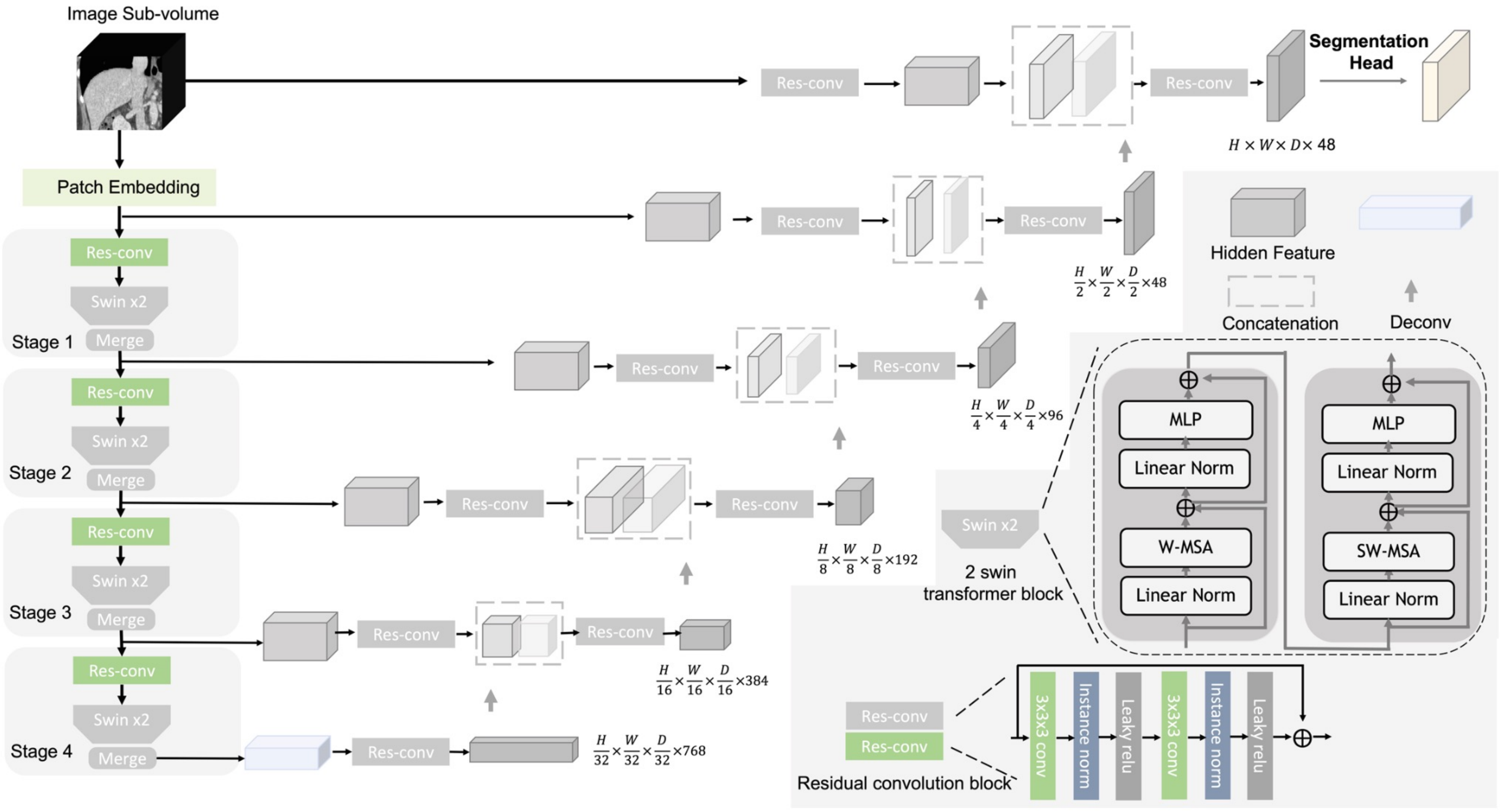


3D UNET Architecture



SWIN Attention

SHIFTED WINDOW UNET TRANSFORMER II



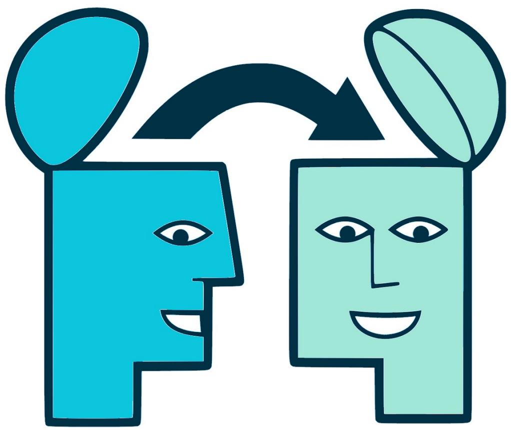
SwinUNETRv2 Architecture

DATASET

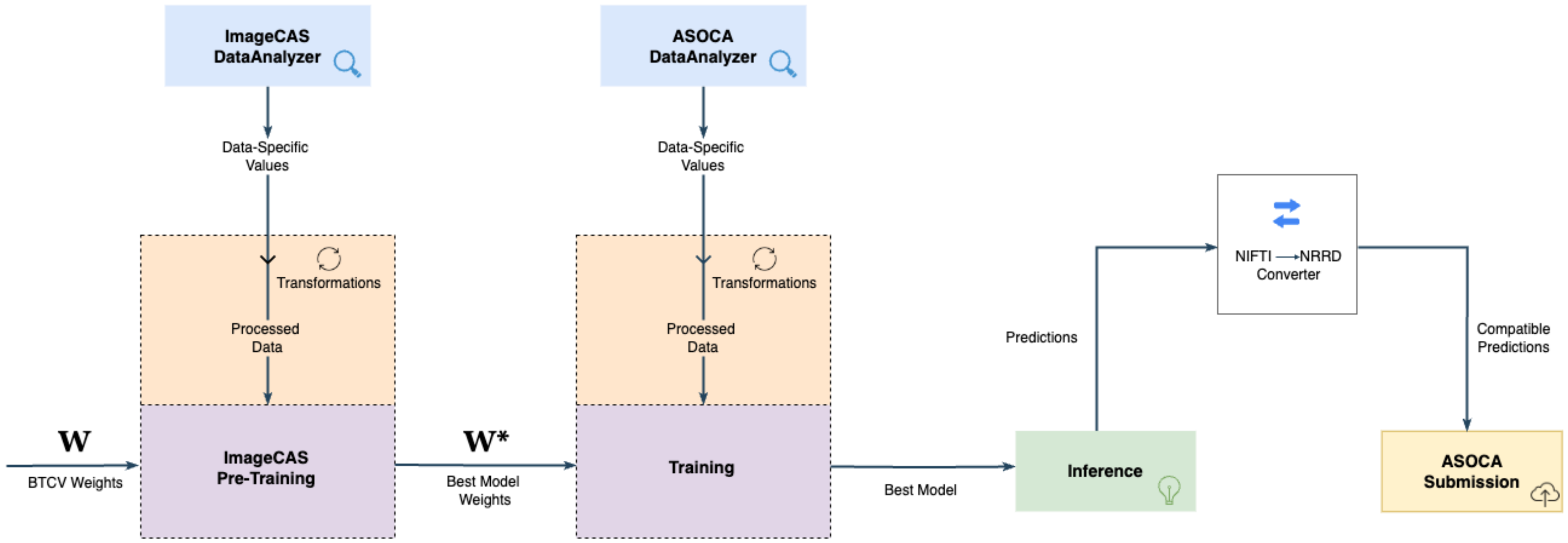
- 60 CTCA scans
  - 30 healthy cases
  - 30 diseased cases
  - Manually labeled by experts
- Splits; all are 50/50 between healthy/diseased
  - **Training**: 36 scans
  - **Validation**: 4 scans
  - **Testing**: 20 scans (hidden labels)
- EDA **only** on training data!

PRETRAINING

- Difficult task, small dataset
  - Pretrained models can give significant initial boost
  - Sadly, lack of models online due to niche problem/model
- Approach **#1**- BTCV Challenge
  - Trained on 30 3D CT volumes
  - Multi-class segmentation, 14 classes
  - Publicly available SWIN-UNETR weights online
- Approach **#2** – BTCV + ImageCAS
  - Contains 1000 3D CTCA scans, we choose **subset of 111**
  - Data highly relevant to our task
  - Use approach #1, then train on ImageCAS and generate weights for fine-tuning on the ASOCA dataset

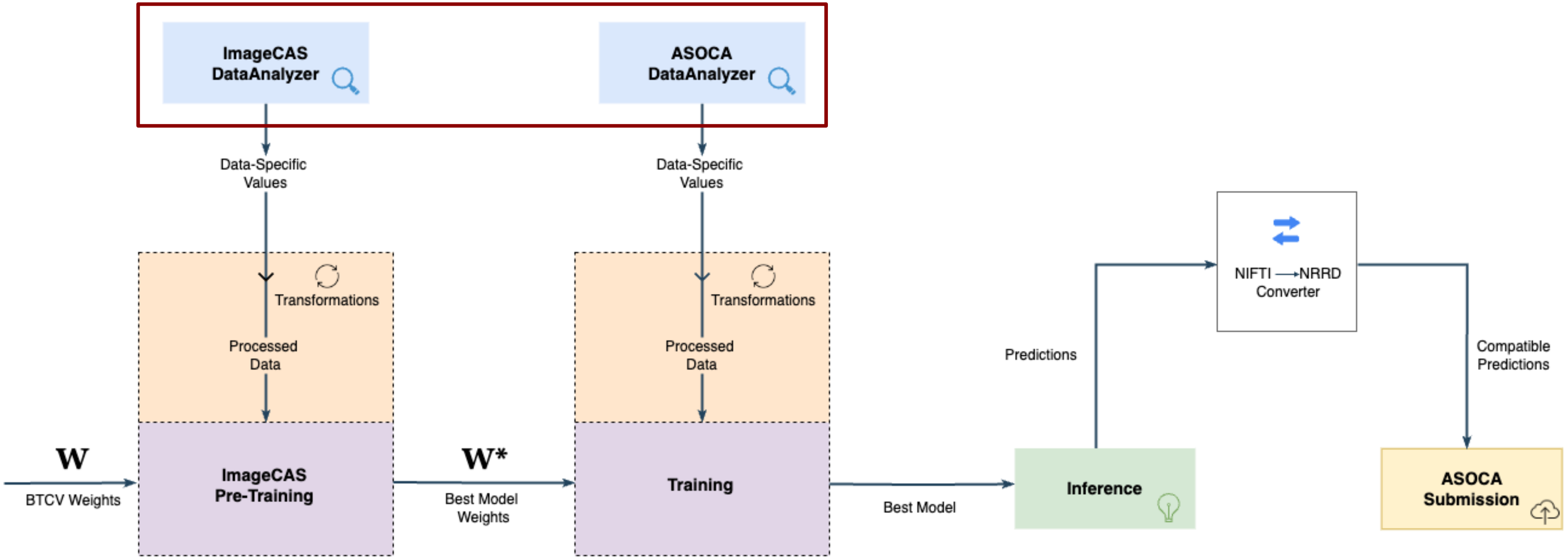


PIPELINE



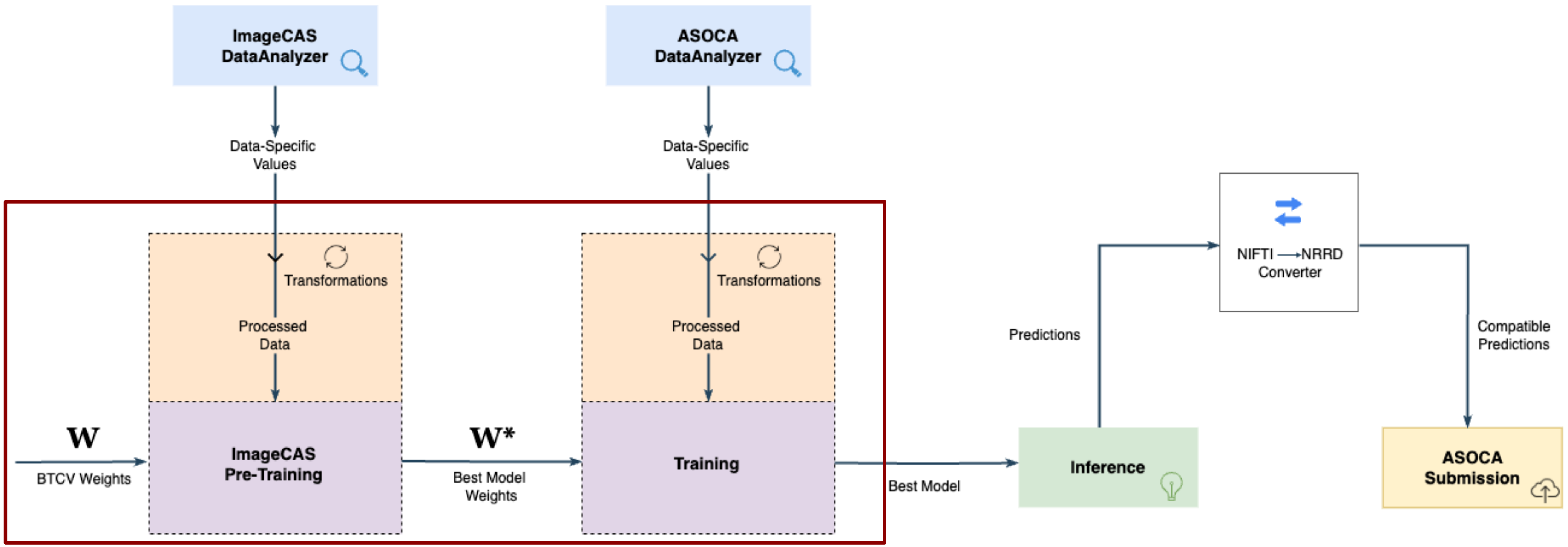
Project Pipeline

PIPELINE



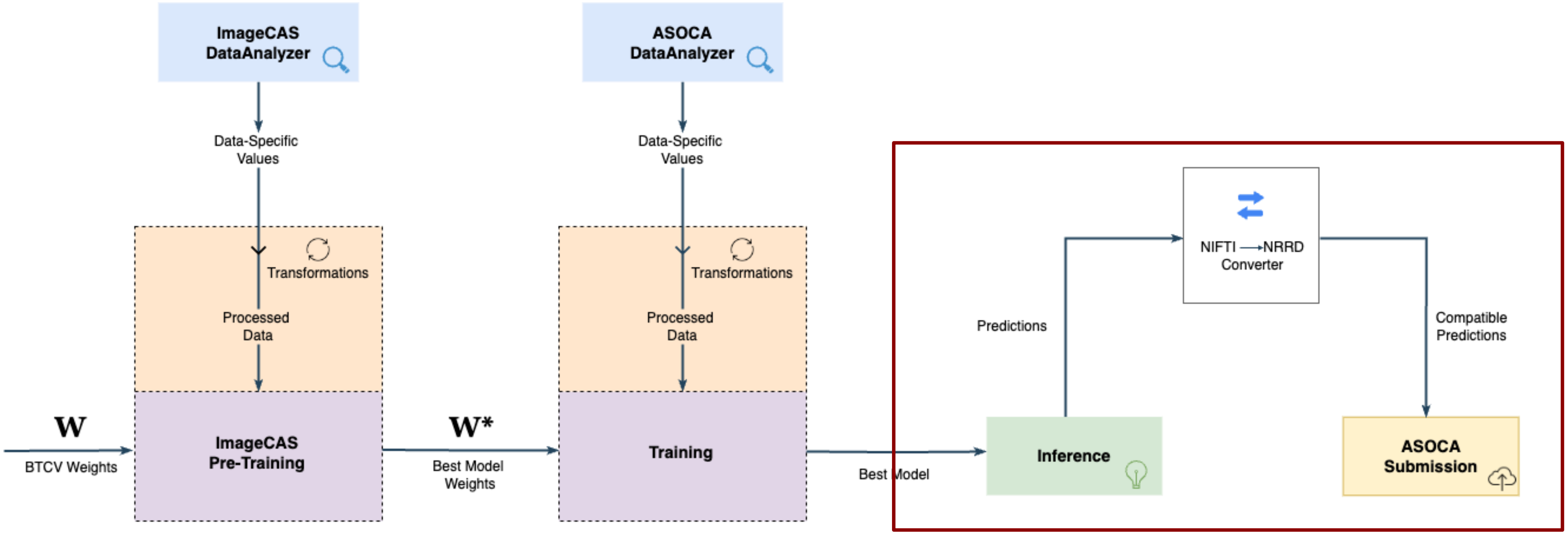
Project Pipeline

PIPELINE



Project Pipeline

PIPELINE

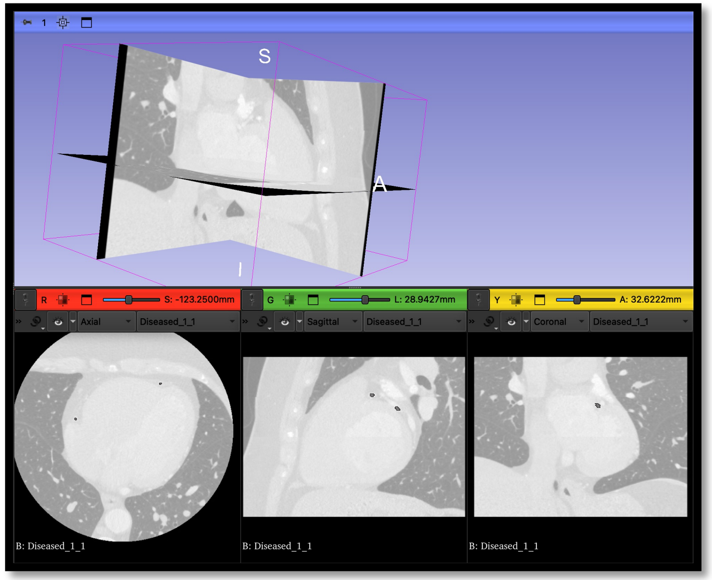


Project Pipeline

EXPLORATORY DATA ANALYSIS - APPROACH

- 3D Slicer
  - Visualize 3D CT images and their segments
  - Useful initially, prefer programmatic approach

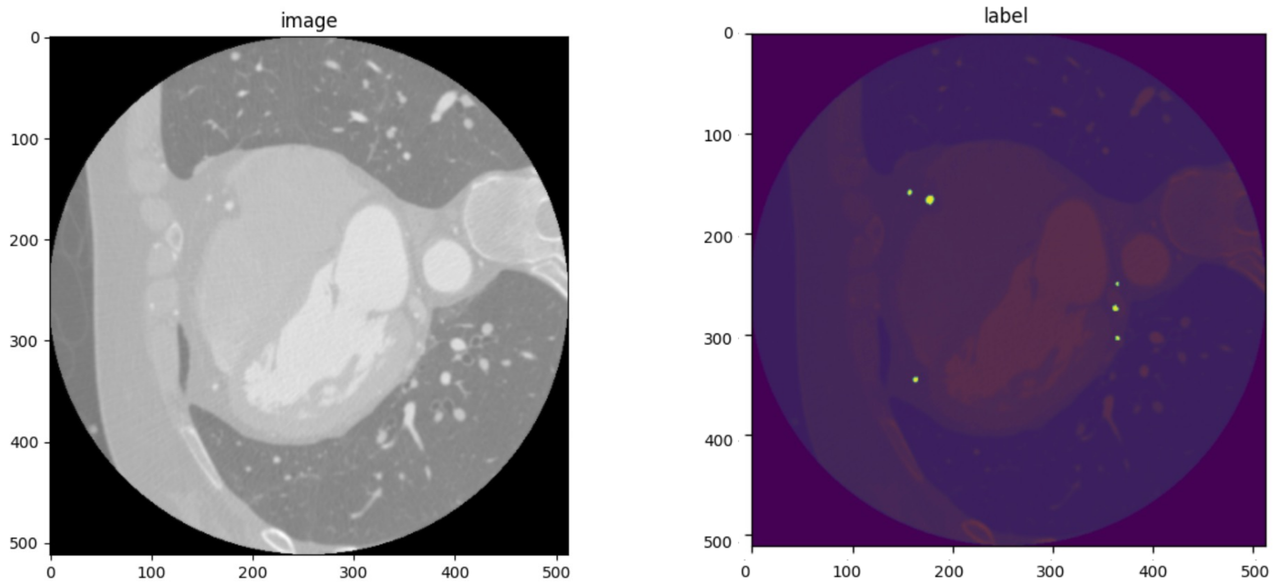
- MONAI's Auto3DSeg DataAnalyzer
  - Provides complete pipeline, we only use the analyzer
  - Exports useful statistics about the input images
  - Key to picking the right transformation values



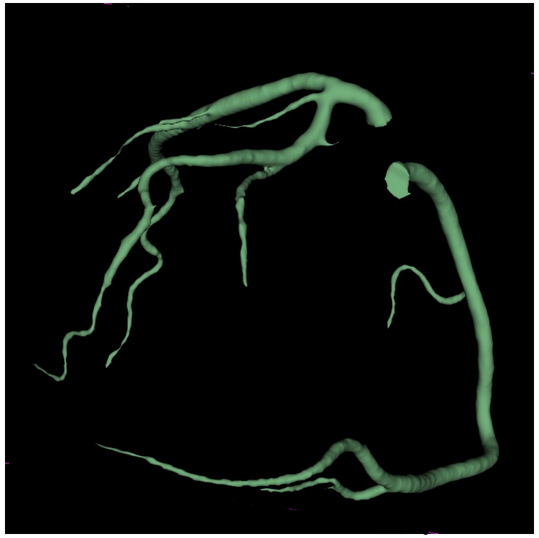
Diseased Sample in 3D Slicer

EXPLORATORY DATA ANALYSIS - KEY STATISTICS

- Mean image shape of [512, 512, 215]
- 1 channel (grayscale)
- Image Intensity (5/99.5 percentiles) of [150, 623]*
- Image Spacing (99.5 percentile) of [0.49, 0.49, 0.63]*



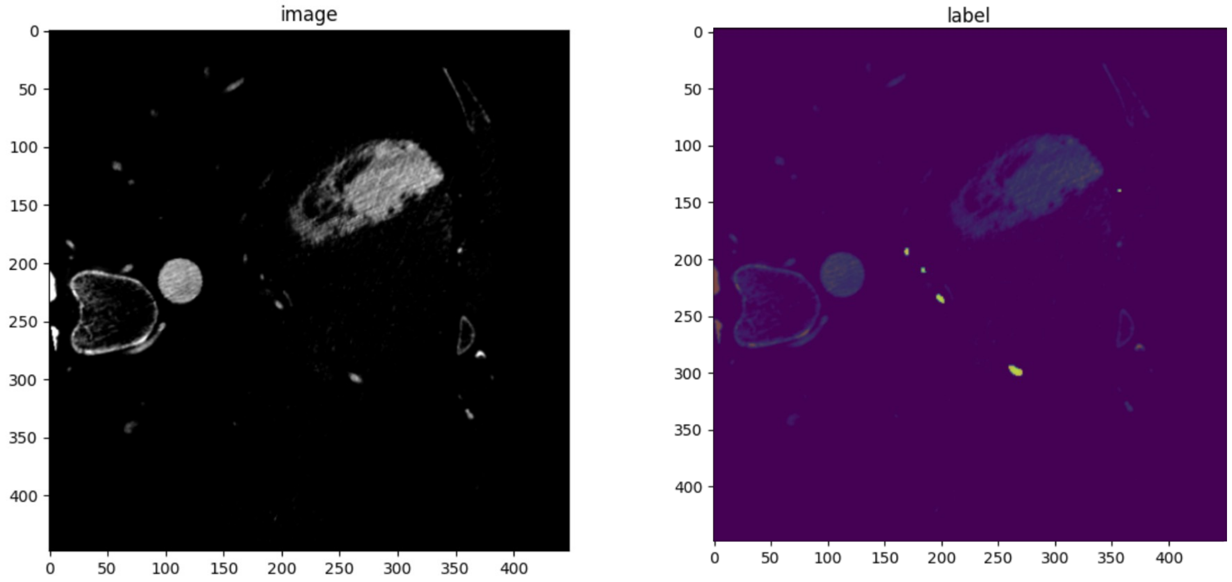
Healthy Sample Slice w/ label



Healthy Sample Segmentation (3D)

PROCESSING – DATA AUGMENTATION

- CT image data is usually complex and machine-dependent
  - Normalization and processing is necessary
- Dataset is small, need to enrich and modify samples
- MONAI provides transforms tailored to biomedical imaging



Post-Transform Healthy Sample Slice w/ label

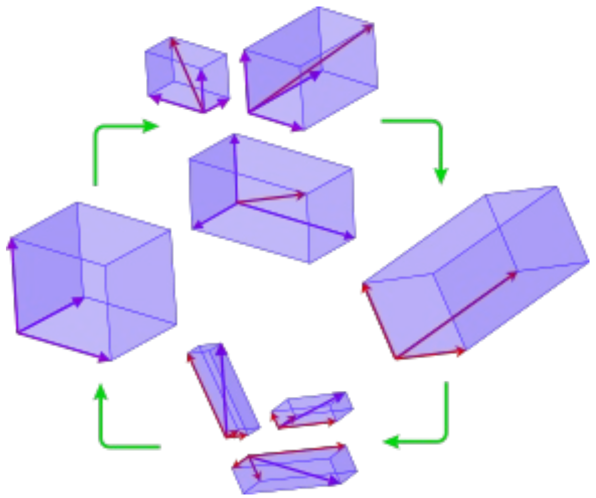
DATA PROCESSING – KEY TRANSFORMS

- Applied to **all** splits and phases!
- **ScaleIntensityRange**
  - CT images have fixed range, highlight important features
  - Squishes the range to [0, 1], values outside that range are omitted
  - *Chosen based on EDA with values [min, max] of [150, 623]
- **Orientation**
  - Sets fixed orientation for all of our inputs based on the affine matrix
  - RAS (Right, Anterior, Superior)
- **Spacing**
  - Adjusts samples so as to ensure that everything is consistently spaced across the dataset
  - *Chosen based on EDA with values of [0.49, 0.49, 0.63]

(trivial transforms such as LoadImage, EnsureChannelFirst are omitted for brevity)

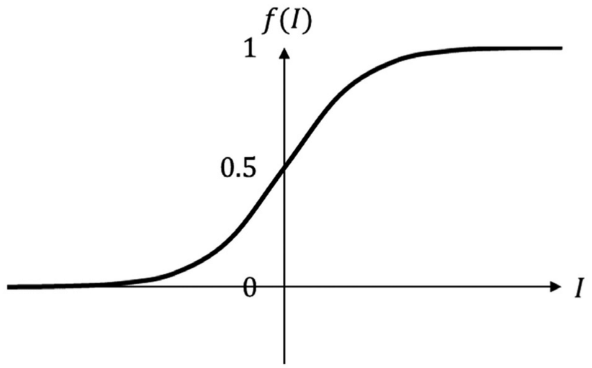
DATA PROCESSING – TRAINING TRANSFORMS

- Only applied during training!
- **RandCropByPosNegLabel**
  - Crops a patch from the image
  - Images have high dimensionality; can't process efficiently without resizing
  - **Solution**: process a random patch of the original image at every iteration
  - Opted for size [128, 128, 128]
  - Oversampling for patches with segments (2:1)
- **RandAffine**
  - Dataset is small, can introduce noise so as to improve generalization capabilities
  - Scale and rotate patch with probability 50%



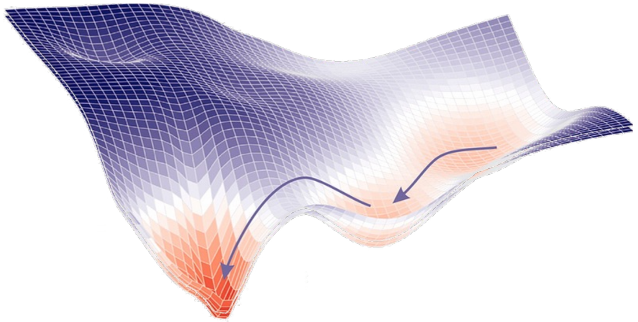
DATA PROCESSING – POST TRANSFORMS

- Applied to labels during validation/testing!
- Some key transforms altered spacing, orientation
  - **Invert** these transforms for final ASOCA submission
- **Activations**
  - Applies the sigmoid activation to the outputs
- **AsDiscrete**
  - Discretizes the outputs (0: Background, 1: Segment)
  - Requires 75% probability for (1)
- **KeepLargestConnectedComponent**
  - In order to omit small noisy segments, we can filter them out if they aren't connected
  - Opt for 3 components



TRAINING - OPTIMIZATION

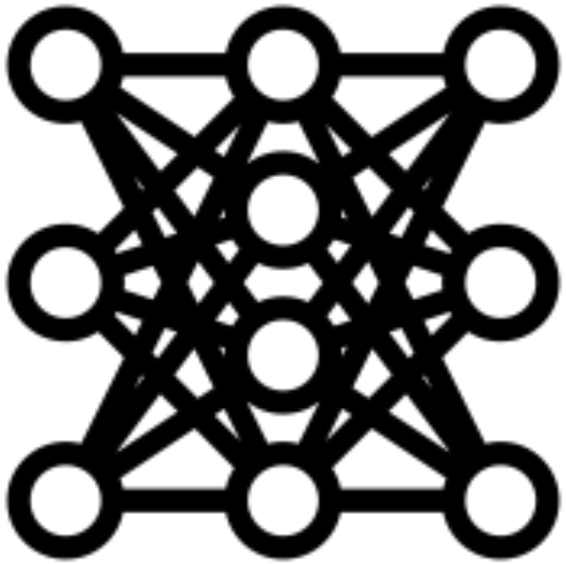
- **Loss Function**: DiceFocalLoss
  - Combination of Dice and Focal loss (weighted sum)
  - **Dice** loss measures dissimilarity between prediction and ground truth segments
  - **Focal** loss regulates learning by weighing high confidence prediction down, focusing on missclassified examples
  - Better than DiceCELoss or standalone DiceLoss in my experiments
- **Optimizer**: AdamW
  - Improvement over Adaptive Momentum (ADAM), weight decay directly on parameters instead of gradient
  - Learning rate of 2e-4
  - Weight decay of 1e-6
  - **Linear warmup** of 6 epochs
  - Significantly better than SGD, Novograd in my experiments



FINAL MODEL

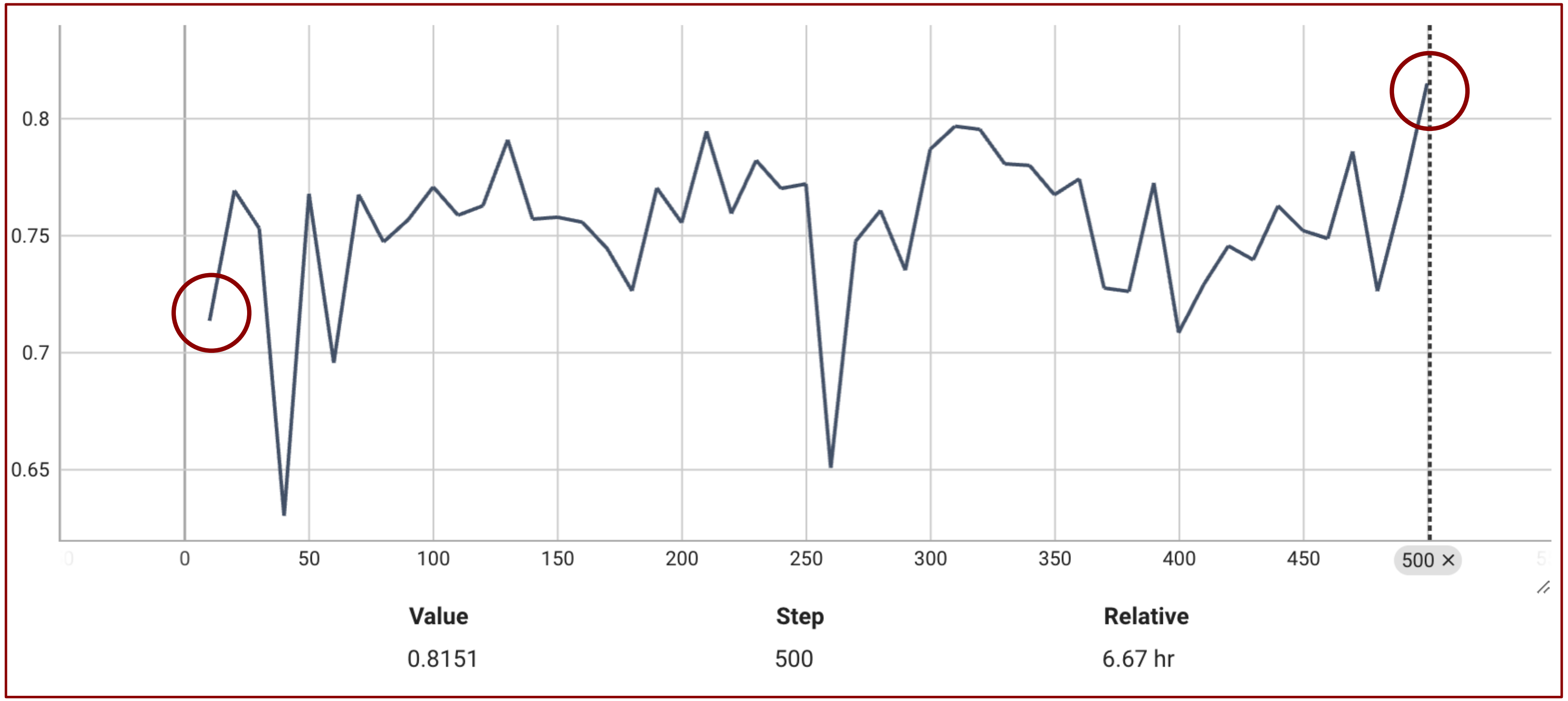
- Pre-trained weights from subset of ImageCAS
  - 111 examples (101 train, 10 val)
  - Identical EDA & transforms, initialized with BTCV weights
  - Trained for 160 epochs
  - Approximately 6.7 hours on one RTX 4090

- Final model - SwinUNETRv2
  - Trained for 500 epochs, validation every 10
  - Validation and testing using sliding window inferer with window of size equal to the patches (128)
  - Approximately 6.7 hours on one RTX 4090
  - Inference time of approximately 10 minutes – good enough
    - Lots of transforms, inversions, saving the outputs etc…



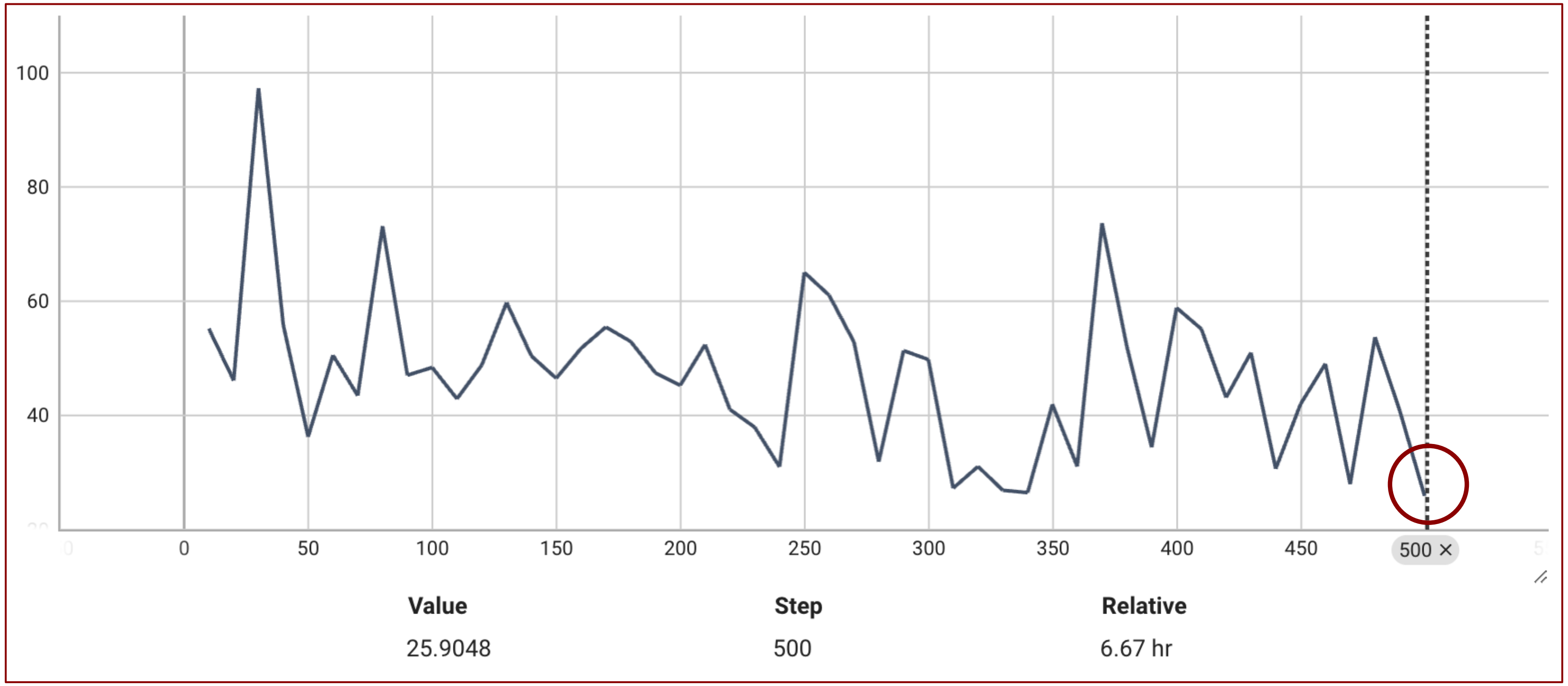
EVALUATION – QUANTITATIVE METRICS (DICE)

Validation Mean Dice Score



EVALUATION – QUANTITATIVE METRICS (HD95)

Hausdorff Distance 95% Score



EVALUATION – ASOCA LEADERBOARD

- After weeks of waiting (and a couple emails), access to the ASOCA challenge submissions!
- Hidden test labels; scores are undisputable and concrete
  - No doubt about data leakage or "beneficial" choice of test set
- Satisfying results given constraints
  - 83% Dice, 17th on the leaderboard
  - BTCV-only pretraining also worthy of mention, 79% Dice

Dice Coefficient

HD95

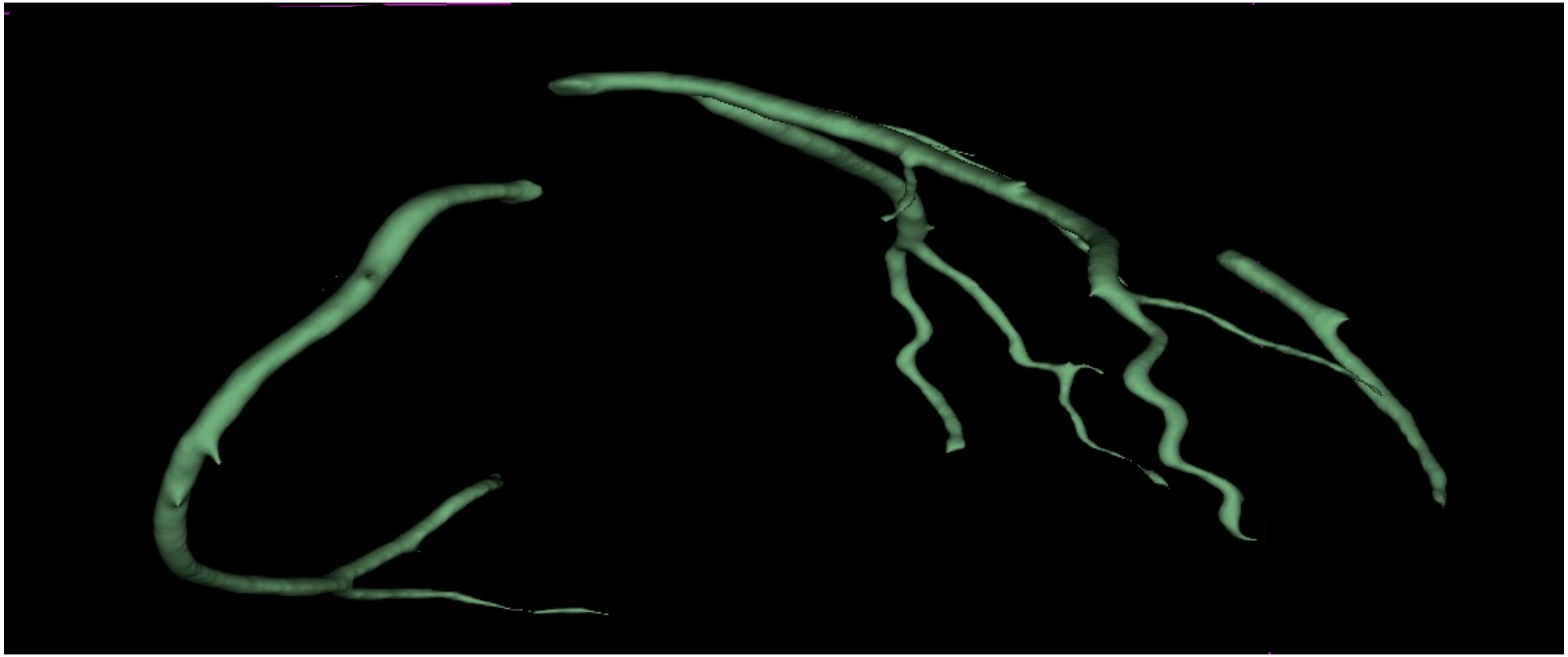
17th	DionGR	1 May 2024	$\boxed{0.8305}$	$\boxed{11.9732}$

Best model - BTCV + ImageCAS Pretrain

22nd	DionGR	28 April 2024	0.7985

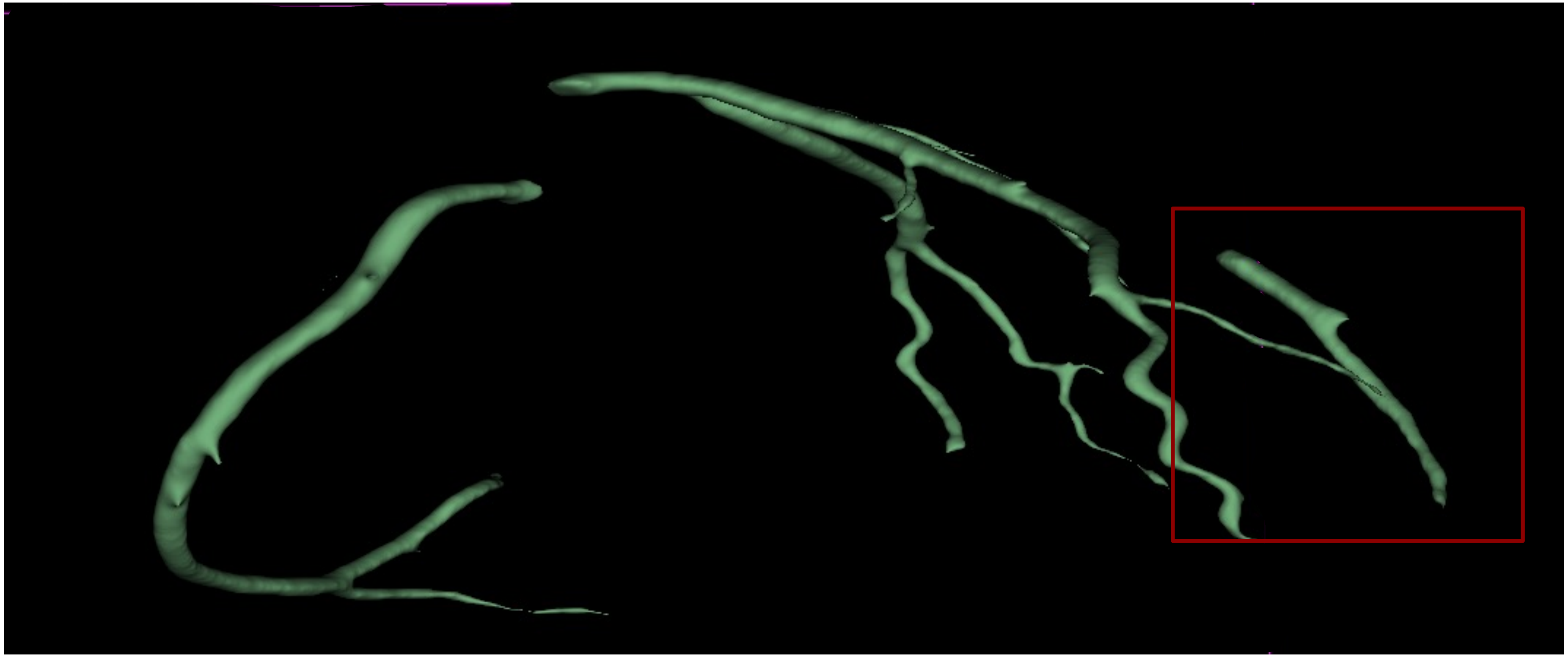
Honorable Mention – BTCV only

EVALUATION – QUALITATIVE ANALYSIS (3D)



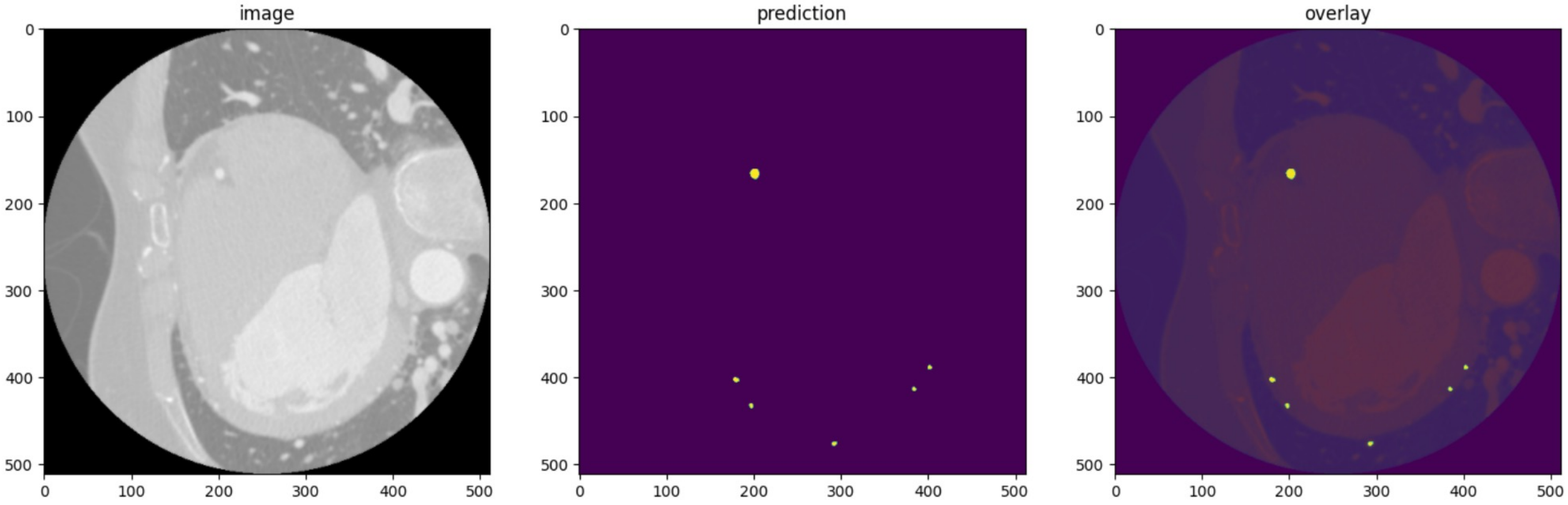
Healthy Test Sample 0 – Prediction (3D Slicer)

EVALUATION – QUALITATIVE ANALYSIS (3D)



Healthy Test Sample 0 – Prediction (3D Slicer)

EVALUATION – QUALITATIVE ANALYSIS (2D)



Healthy Test Sample 0 – Prediction Slice

CARBON FOOTPRINT

- Model underwent training plenty of times
  - Both for parameter tuning as well as trial-and-error
- Computationally demanding
  - Cybele computers are the minimum; wont run on less VRAM than a RTX4090
  - Also RAM heavy due to dataloders; ImageCAS especially

- A total training time of approximately:
  - 4635.98 minutes or
  - 77.26 hours or
  - 3.2 days
- Power consumption for RTX 4090 (450W TDP)
  - Average of 90% usage, approximately 405W or 0.405kW
  - 0.405 kW for 77.26 hours = 31.30 kWh
- Tesla Model X Plaid (18.1 kWh/100 km)
  - Could approximately drive **172.93km**
  - **or 98817.14 left coronary arteries**

DISCUSSION

- Half the task was figuring out the transforms
- BTCV-only pretraining was satisfactory
- If there was more time:
  - nnUNET
  - Tailored pre-training on ImageCAS
  - More random transformations, longer pretraining
  - Maybe steal the 9th position on the leadearboard? :D

- Overall more than satisfied with the results
- Very fun task
  - Sometimes lack of documentation in MONAI
  - Lab assistants were very helpful

THANK YOU FOR YOUR ATTENTION[1]

DIONYSIOS RIGATOS

[1] Bahdanau et al. (2016)