Eye Vessel Prediction

Pixel-wise Segmentation with U-Net Architectures

Agenda

- Introduction
- The DRIVE Dataset
- The U-Net
- The Fully Convolutional Transformer
- Training & Results
- Summary
- Appendix

Introduction

A Brief History of Deep Learning

- The perceptron was first invented in 1943 and backpropagation in 1986.
- Deep learning gained massive popularity in the 2010s with the advent of big data, big compute, and the success of large computer vision models and large language models.
- There are many different sub-families of neural network architectures, each specializing in a specific task or modality (i.e. image data, sequence/text data, or data generation).
- Convolutional neural nets are still the gold standard in computer vision tasks. However, many recent applications of transformers to these tasks have proven successful.

Deep Learning in Healthcare

- With the recent success of neural networks, many researches have attempted to apply them to various areas of healthcare, with varying levels of success.
- There are many challenges in the healthcare domain which are problematic for deep learning
  - Datasets are usually very small
  - The costs of wrong predictions are very high
  - Model interpretability is highly desired
- Image segmentation is a common task of deep learning in the healthcare domain

*COVID-NET detecting infected areas within a chest radiography image

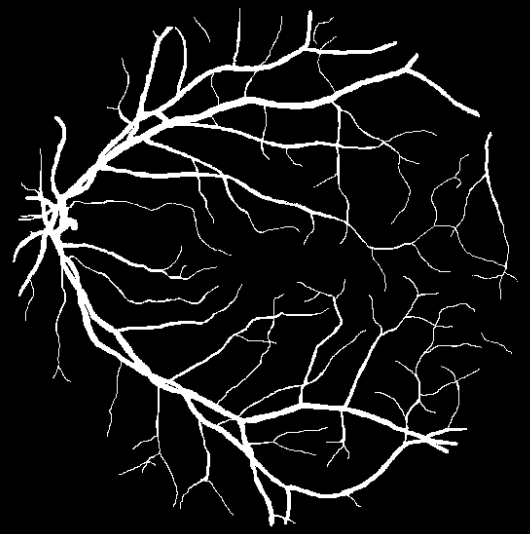
The Task

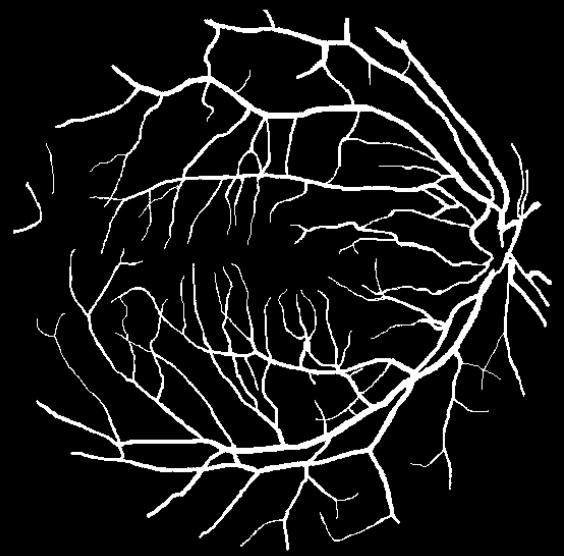
- Identify the blood vessels in the images of retina
- Train a model that takes the RBG retina image as input and produces a binary blood vessel map as output
- In machine learning, this is known as image segmentation
- Measure model performance on a set of held-out images that were not used in training

The DRIVE Dataset

Digital Retina Images for Vessel Extraction (DRIVE)

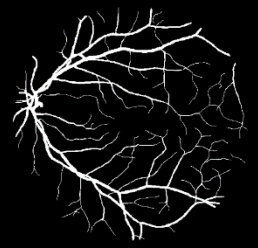
- 40 RGB images of the retina obtained from a diabetic retinopathy screening
- Training and testing sets have been divided into 20 images each
- Each image is paired with a manual segmentation map of the vasculature serving as the ground truth label
  - These human annotations were overseen by medical professionals

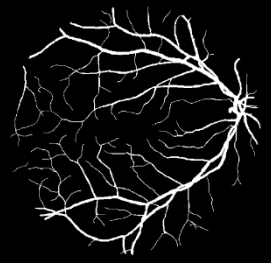


Data Augmentation

Original Image:



Horizontal Flip:

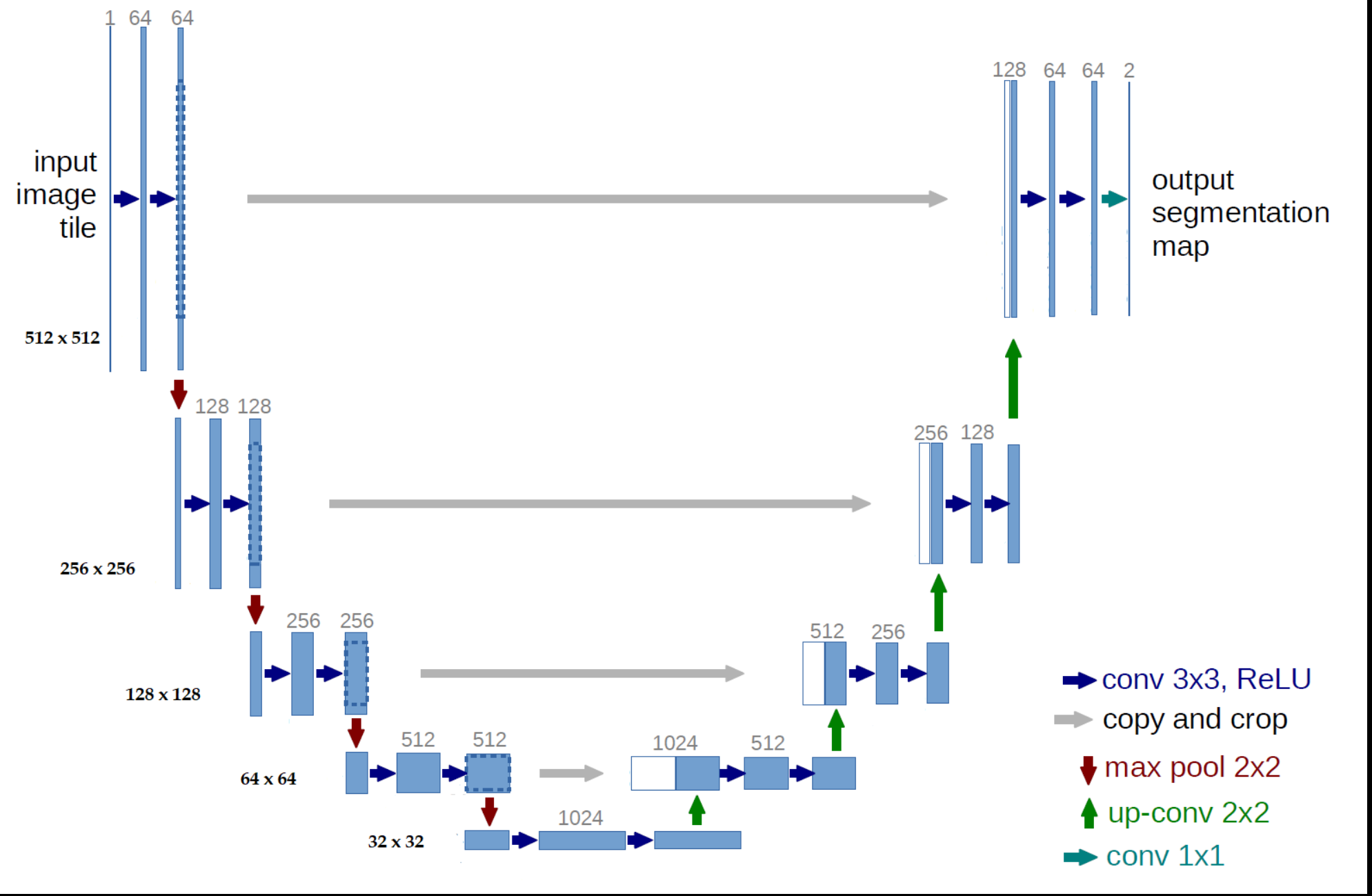


Vertical Flip:

45 deg Rotation:

- Increases training set size by applying neutral transformations to images
- Applied a horizontal flip, vertical flip, and a 45-degree rotation.
- Training set size boosted from 20 images to 80 images

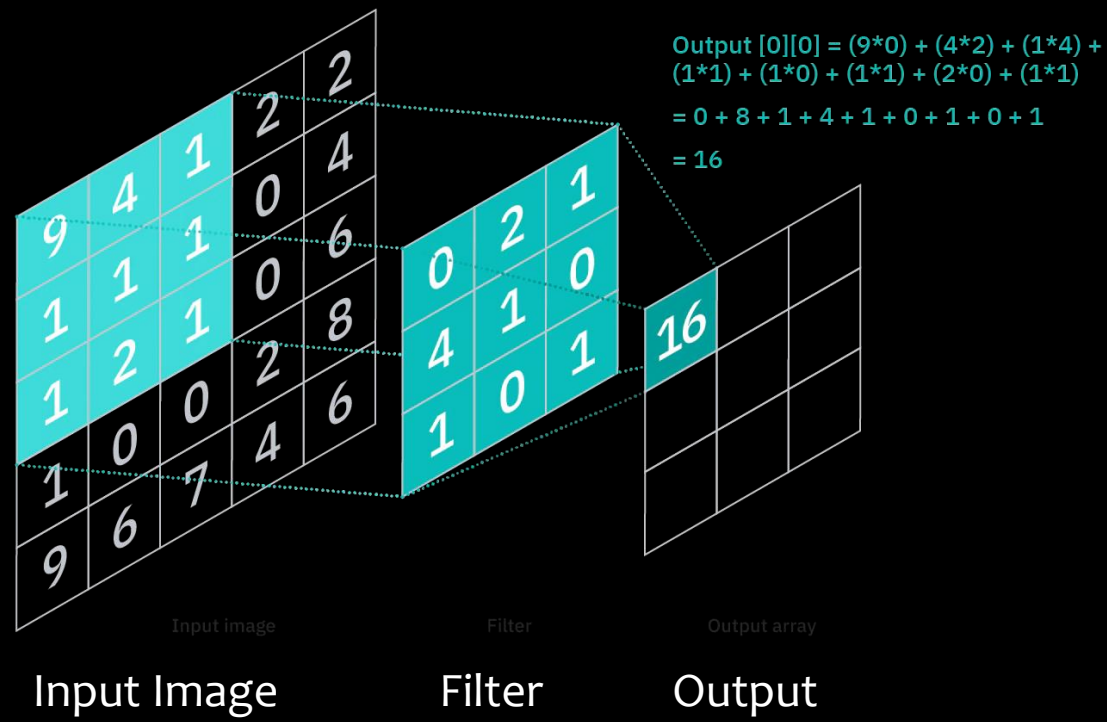
The U-Net



O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," Arxiv.org, 18 May 2015. [Online]. Available: https://arxiv.org/pdf/1505.04597v1.pdf.

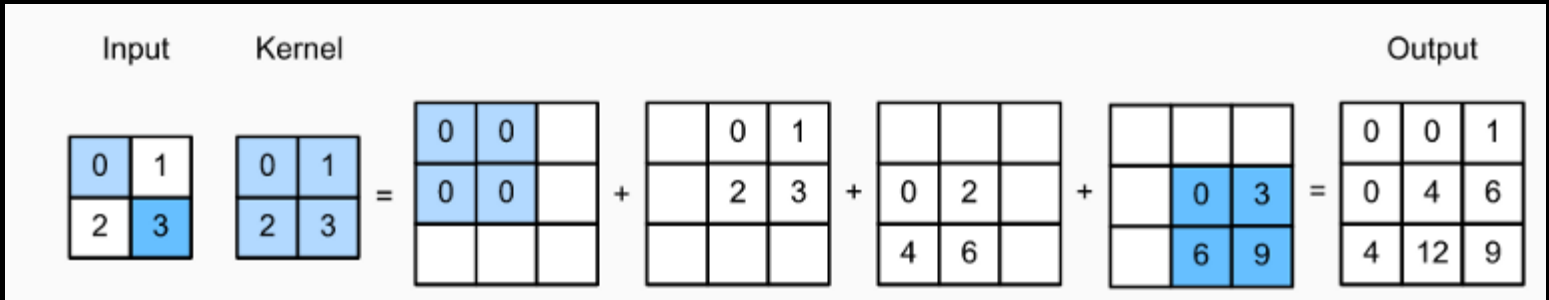
The Convolutional Layer

- An image is effectively a matrix of integers between 0-255.
- A ‘kernel’ or ‘filter’ is a smaller matrix that slides over the input image (i.e. convolves), and the dot product of is taken at each step, which forms a smaller output matrix.
- Many such filters are often applied at each layer.
- The weights of the kernel matrix are learned during training to extract salient features from the images.

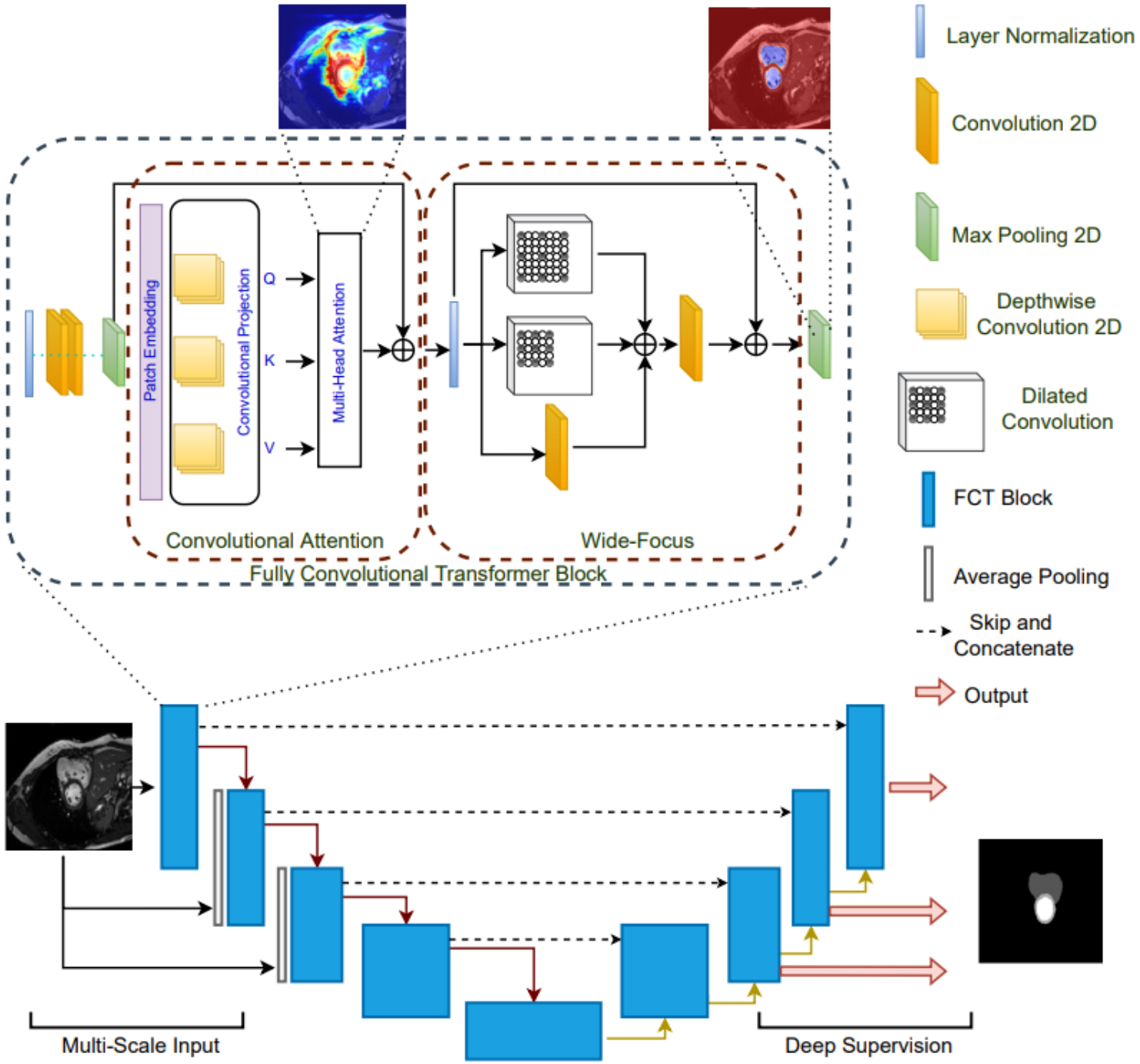


The Transposed Convolution

- How does the decoder up-sample the convoluted outputs?
- A kernel is learned such that each entry in the input matrix is multiplied by the entire kernel.
- The resulting matrices are then summed element-wise to return an output matrix with a larger size.



The Fully Convolutional Transformer



- Same basic U-Net shape
- A novel FCT block has replaced each convolutional block from U-Net
- Each FCT block is composed of a "Convolutional Attention" block and a "Wide-Focus" block
- "Patch Encoding" will convert each input image into a series of flattened token vectors to be passed through a multi-head self-attention layer
- "Wide-Focus" applies basic convolution and two layers of dilated convolution independently
- Up-sampling the decoder uses the more basic nearest neighbors

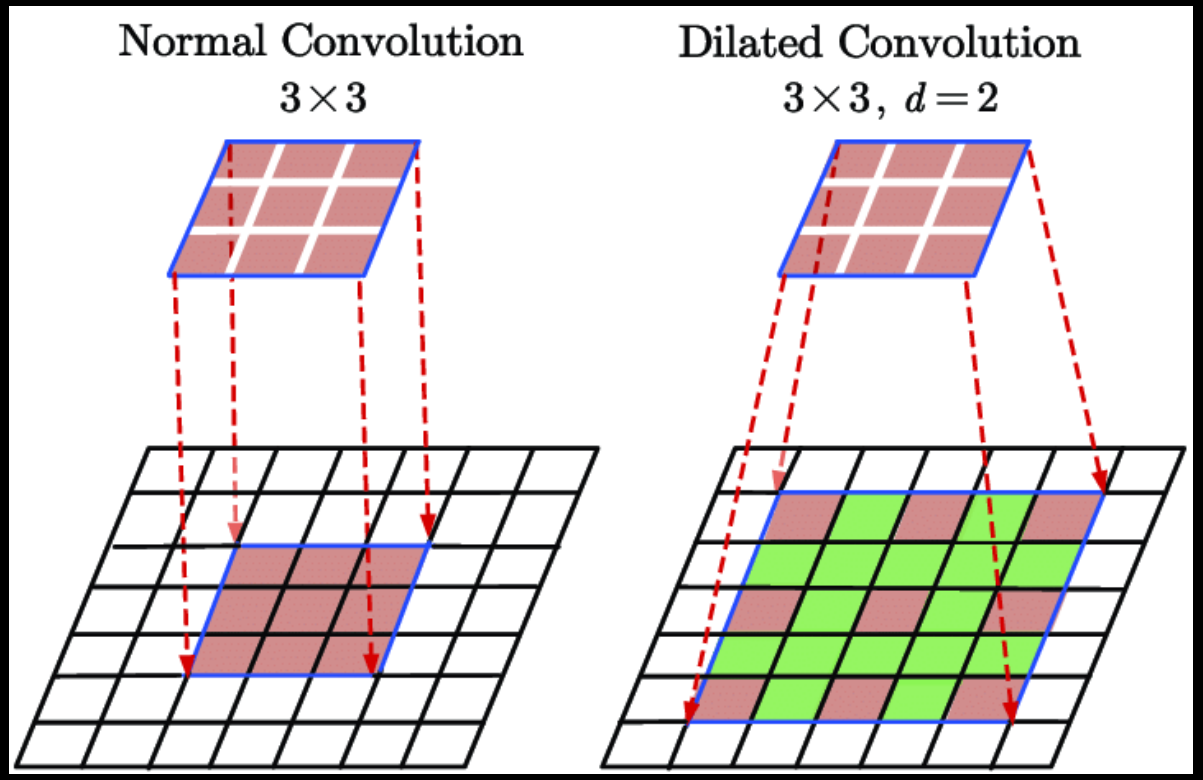
A. Tragakis, C. Kaul, R. Murray-Smith and D. Husmeier, "The Fully Convolutional Transformer for Medical Image Segmentation," Arxiv.org, 1 June 2022. [Online]. Available: https://arxiv.org/pdf/2206.00566.pdf.

Convolutional Attention

- Tokens are created by convolving over the input with a 3x3 kernel, with ‘same’ padding to preserve the output shape. After layer normalization, the output is flattened into a token vector
- The tokens are then passed through a multi-head self-attention (MHSA) layer that uses depth-wise convolution rather than linear projections, which removes the need for positional encoding
- MHSA forms the foundation for the innovation of the Transformer architecture. The details are out-of-scope for this presentation but suffice it to say it should allow the network to learn long range spatial context from the images.

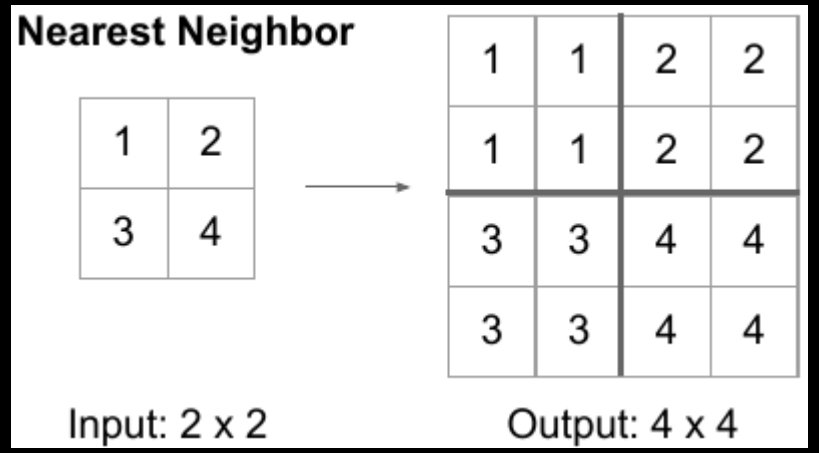
Dilated Convolution

- A generalization of the convolution operation
- Standard convolution is dilated convolution with d=1
- Expands the kernel outward, skipping over entries in the input matrix when performing the dot product
- Potentially learns salient features over a larger resolution



Nearest Neighbor Up-sampling

- The up-sampling method in FCT is simpler than the transposed convolution seen in U-Net
- Each entry in the input matrix is directly projected without any learnable weights or filters



Training & Results

Training Schemas

	U-Net	FCT
Python Library	PyTorch	PyTorch
GPU	NVIDIA 3090 (24 GB)	NVIDIA 3090 (24 GB)
Epochs*	63	46
Learning Rate	0.001	0.0001
Optimizer	Adam	Adam
Loss Function	Dice + BCE	BCE
Batch Size	2	1
Image Scale	100%	50%

*100 epochs, but with early stopping patience of 5

Results

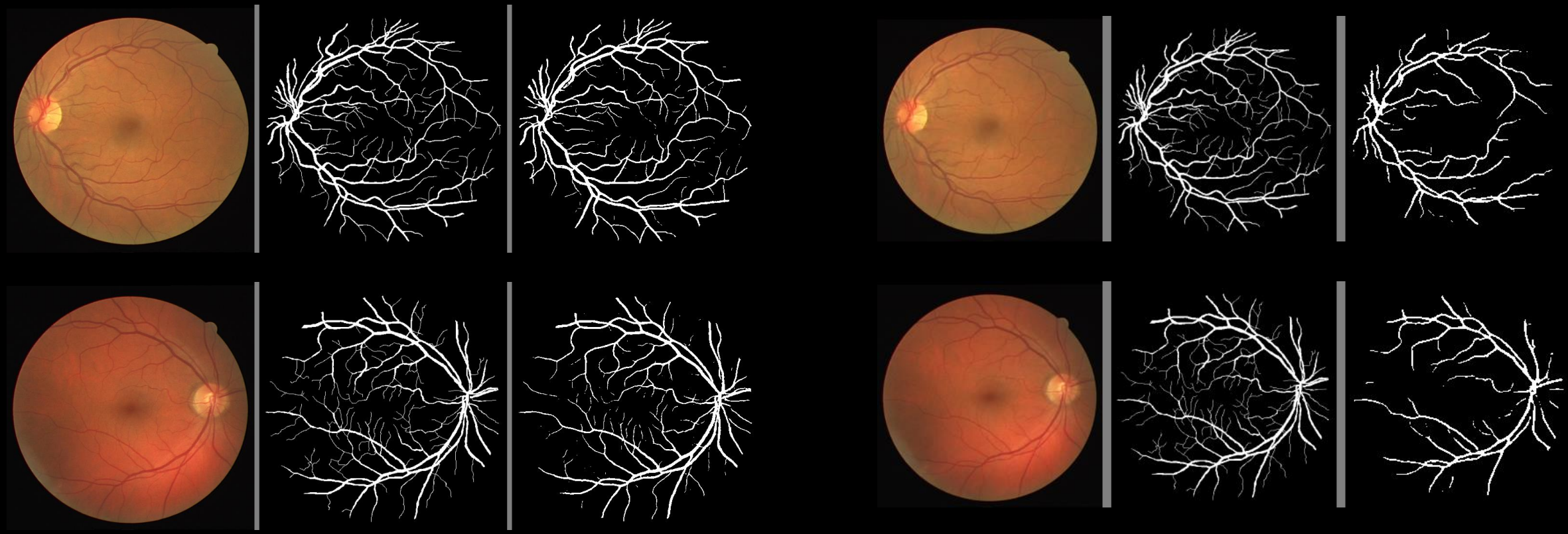
- Neither of my implementations achieve SOTA
- Likely due to a combination of more complex data augmentation and algorithms (the top SOTA model was a GAN)
- U-Net performed better than FCT, which I didn't expect
- Likely due to the rescaled images necessary for training

Metric	U-Net	FCT	SOTA
F-1 Score	0.7994	0.7154	0.8690
Precision	0.8163	0.8180	N/A
Recall	0.7878	0.6404	0.7927
Accuracy	0.9657	0.9581	0.9790
AUC ROC	0.8854	0.8137	0.9887

Test Set Prediction Examples

U-Net

FCT



Summary

Capstone Summary

- Implemented two deep learning architectures designed for image segmentation, one classic and one new: U-Net and the Fully Convolutional Transformer
- Modeled the DRIVE dataset, a small set of retina blood vessel images
- Data augmentation was employed to expand the size of the training set from 20 to 80 images
- Neither model achieved SOTA performance, but both did well
- U-Net surprisingly performed better than FCT, but likely an artifact do to training limitations
- Models of this nature offer innovative technologies to the healthcare profession

Next Steps: If I had more time

- Analyze misclassified pixels. Any patterns?
- Additional data augmentation to increase training set
- Experiment with tuning hyperparameters
- If I had a better workstation, train FCT with out rescaling

Appendix

GitHub repository containing all python code
(www.github.com/ddixonAI/SLU_Capstone)

Performance Metric Definitions:

- Jaccard Score
- F1-Score
- Precision
- Recall
- Accuracy
- AUC ROC

Loss Functions:

- BCE - Dice Loss
- BCE Loss (Binary Cross Entropy)

O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," Arxiv.org, 18 May 2015. [Online]. Available: https://arxiv.org/pdf/1505.04597v1.pdf.

C. Kaul, S. Manandhar and N. Pears, "FocusNet: An Attention-based Fully Convolutional Network for Medical Image Segmentation," arxiv.org, 8 February 2019. [Online]. Available: https://arxiv.org/pdf/1902.03091.pdf.

C. Kaul, N. Pears and H. e. a. Dai, "FocusNet++: Attentive Aggregated Transformations For Effecient and Accurate Medical Image Segmentation," arxiv.org, 7 April 2021. [Online]. Available: https://arxiv.org/pdf/1912.02079.pdf.

Y. N. S. L. N. O. GM Venkatesh, "A Deep Residual Architecture for Skin Lesion Segmentation," 2018. [Online]. Available: https://doras.dcu.ie/22685/1/Deep-Residual-Architecture_Camera_Ready.pdf.

C. Wang, Z. Zhao, Q. Ren, Y. Xu and Y. Yu, "Dense U-Net Based on Patch-Based Learning for Retinal Vessel Segmentation," 21 February 2019. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7514650/.

Z. Zhou, M. Siddiquee, N. Tajbakhsh and J. Liang, "U-Net++: A Nested U-Net Architecture for Medical Image Segmentation," 18 July 2018. [Online]. Available: https://arxiv.org/pdf/1807.10165.pdf.

A. Vaswani, N. Shazeer and e. all, "Attention Is All You Need," arxiv.org, 12 June 2017. [Online]. Available: https://arxiv.org/pdf/1706.03762.pdf.

A. Dosovitskiy, L. Beyer and e. al., "An Image is Worth 16x16 Words: Transformers for Image Recognition At Scale," 3 June 2021. [Online]. Available: https://arxiv.org/pdf/2010.11929.pdf.

H. Wu, B. Xiao, N. Codella and e. al., "CvT: Introducing Convolutions to Vision Transformers," 29 March 2021. [Online]. Available: https://arxiv.org/pdf/2103.15808.pdf.

"Swin Transformer: Hierarchical Vision Transformer using Shifted Windows," arxiv.org, 17 August 2021. [Online]. Available: https://arxiv.org/pdf/2103.14030.pdf.

J. Chen, Y. Lu, Q. Yu and e. al., "TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation," arxiv.org, 8 February 2021. [Online]. Available: https://arxiv.org/pdf/2102.04306.pdf.

A. Hatamixadeh, Y. Tang and V. e. a. Nath, "UNETR: Transformers for 3D Medical Image Segmentation," arxiv.org, 9 October 2021. [Online]. Available: https://arxiv.org/pdf/2103.10504.pdf.

"DRIVE," 2019. [Online]. Available: https://drive.grand-challenge.org.

A. Tragakis, C. Kaul, R. Murray-Smith and D. Husmeier, "The Fully Convolutional Transformer for Medical Image Segmentation," Arxiv.org, 1 June 2022. [Online]. Available: https://arxiv.org/pdf/2206.00566.pdf.

S. Kamran, K. Hossain and e. al., "RV-GAN: Segmenting Retinal Vascular Structure in Fundus Photographs using a Novel Multi-scale Generative Adversarial Network," arxiv.org, 14 May 2021. [Online]. Available: https://arxiv.org/pdf/2101.00535v2.pdf.

E. Uysal, B. Zaza and e. al., "Exploring the Limits of Data Augmentation For Retinal Vessel Segmentation," arxiv.org, 30 May 2021. [Online]. Available: https://arxiv.org/pdf/2105.09365v2.pdf.