



Self Supervised Learning for Lower Limb Muscle Segmentation on MRI

Anoosha Pai S, Erick Siavichay, Nitish Gudapati
{anoopai}/{esiavich}/{gnitish9}@stanford.edu

CS31N: Deep Learning for Computer Vision
Stanford University, Spring 2022

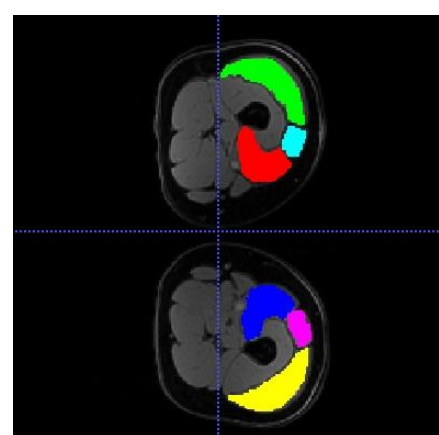
Introduction

- Medical imaging datasets consist of more unlabelled images than the labelled ones.
- Image annotations are limited by the available human experts.
- We use self-supervised learning (SSL) to boost segmentation performance.

Problem

- The goal of this work is to perform semantic segmentation of 6 muscles on a small lower limb MRI dataset.

Data

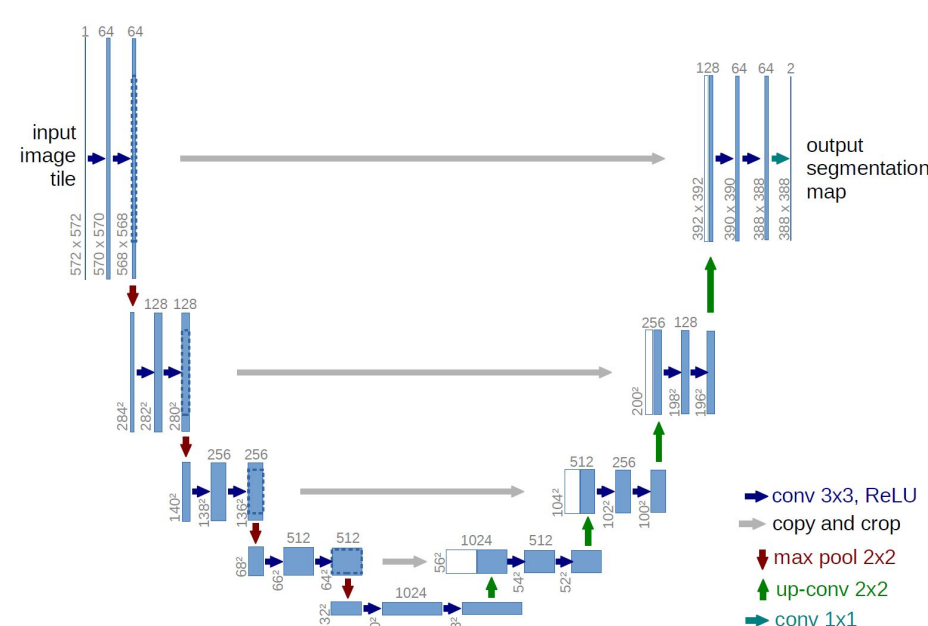


- 2D lower limb MRI scans across 5 subjects and 5 months (~1600 images).
- Images are normalized to have zero mean and unit standard deviation.
- Training - validation - test splits = 75% - 12.5% - 12.5%.

Methods

Baseline: Fully-Supervised Model

- Trained U-Net to perform instance segmentation (multiclass classification).
- Loss function: cross-entropy loss averaged for each class across all images.
- Similarity metric: batch aggregate Dice scores across all classes.



Ronneberger et al., 2015,
<https://github.com/milesial/Pytorch-UNet>

Self-Supervised Model (Pretext + Downstream):

Pretext task: Inpainting

- Masking percent: percentage of the original image to be masked.
- Kernel size: $K \times K$ pixels.
- Trained a U-Net to perform context prediction (regression) on masked images.
- Loss function: mean squared error averaged across all images.

Downstream task: segmentation

- Transferred weights from the pretext task.

Design Experiments

Hyperparameters considered for experiments

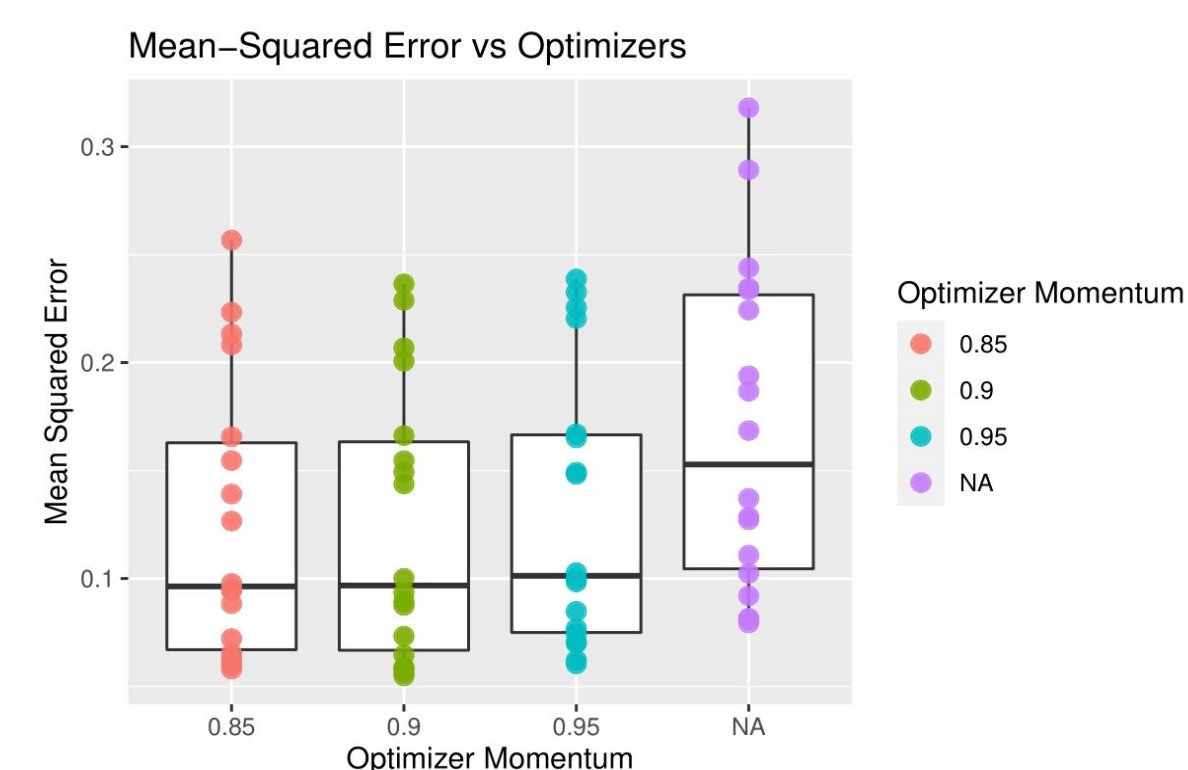
- Learning Rates (LR) : 1e-6, 5e-6, 1e-5
- Optimizer : RMSprop, Adam
- Momentum (M) : 0.85, 0.9, 0.95
- Mask Percentages (MP) : 10%, 25%, 50%
- Kernel Size (KS) : 32×32 , 16×16 , 8×8
- Batch Size: 1, 5, 10
- Transfer learning (freeze): encoder, decoder, both, none

Experiment 0: tested and selected the best network Parameters for the baseline model

- LR = 1e-5, optimizer = RMSprop, M = 0.9

Experiment 1: tested and selected best model parameters for pretext task

- LR = 1e-5, optimizer = RMSprop, M = 0.9

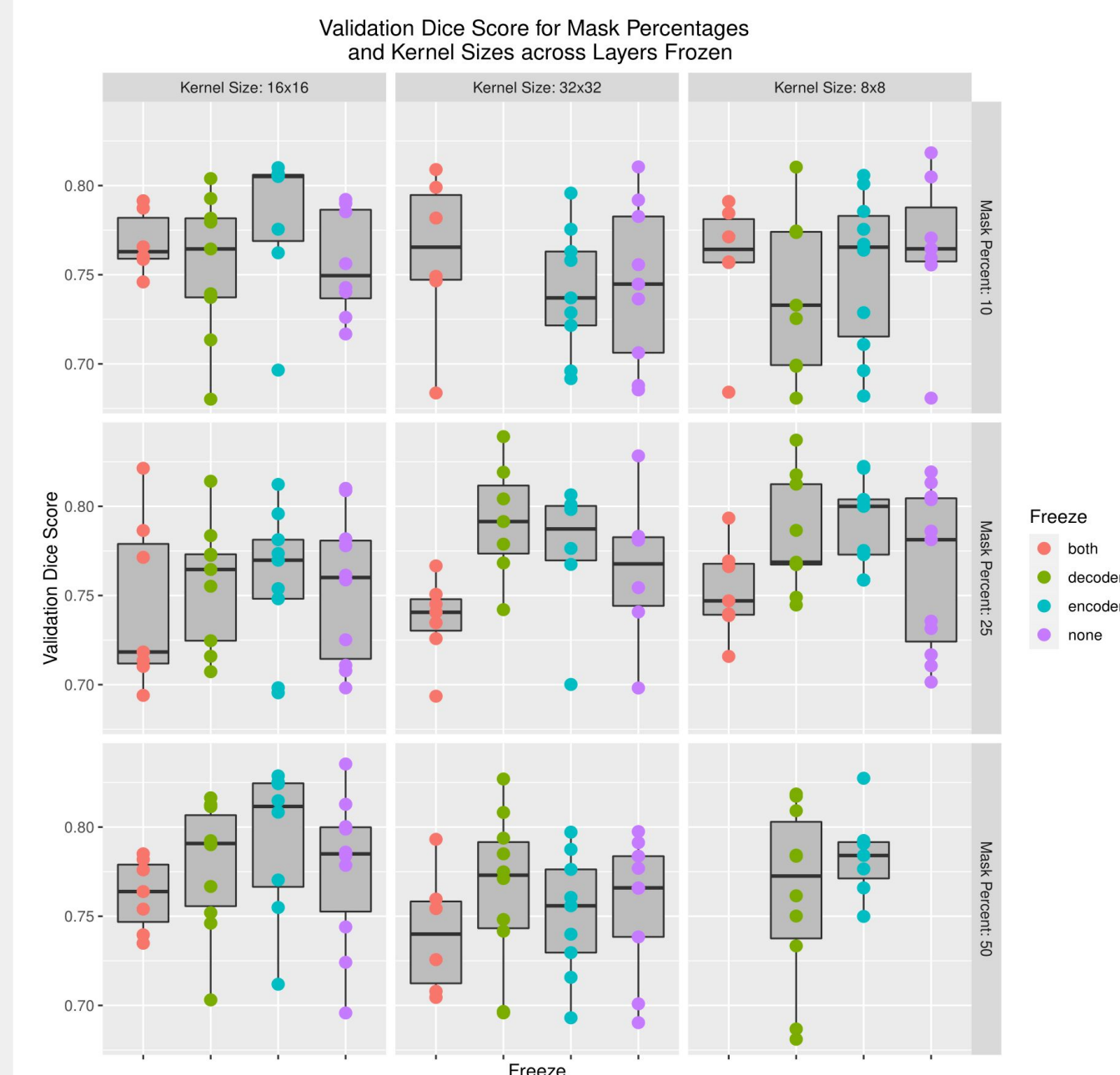


Experiment 2: tested and selected best masking parameters

- kernel size = 16×16 , mask percentages = 10 and 50

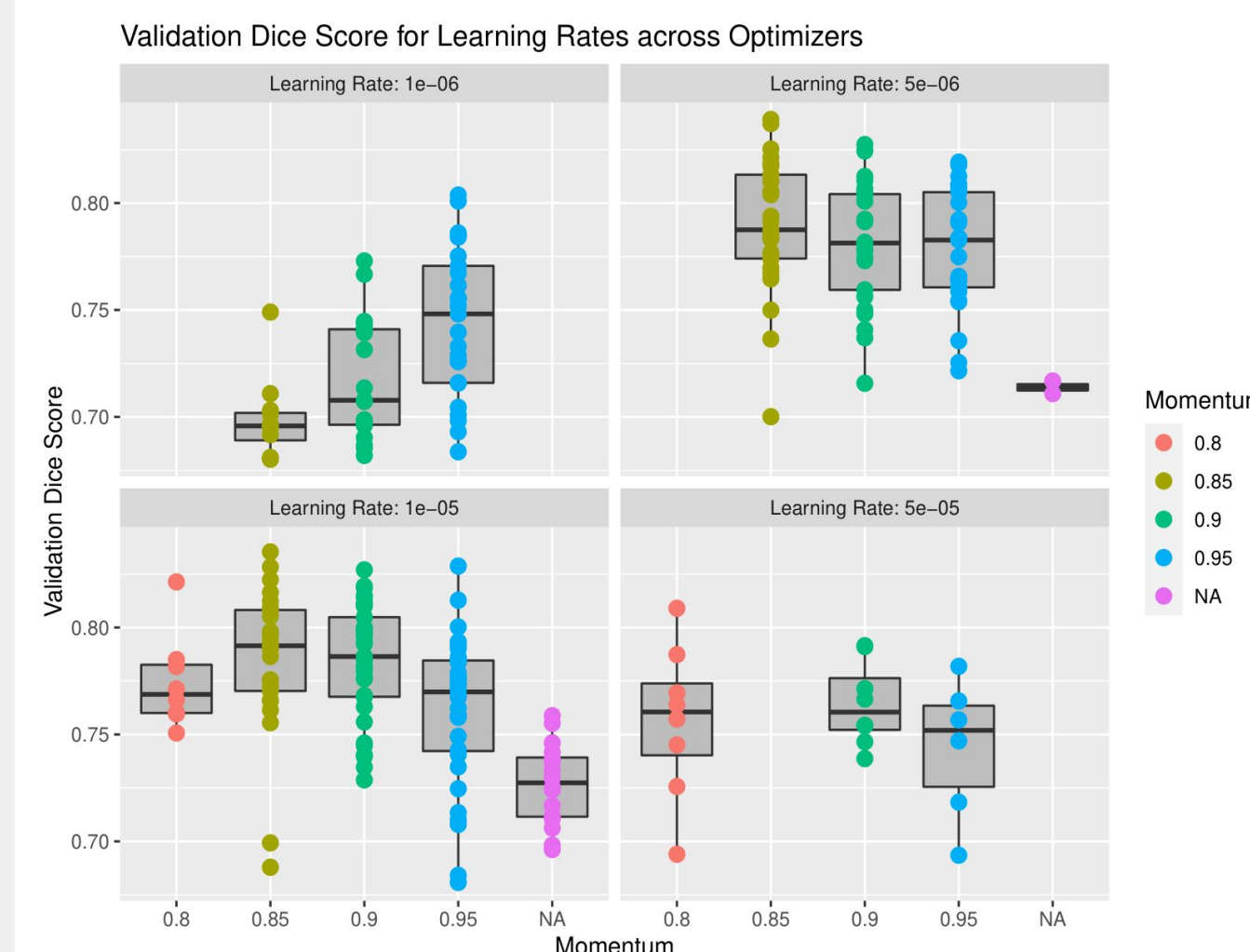
Experiment 3: tested and selected best transfer learning parameters

- only encoder weights frozen



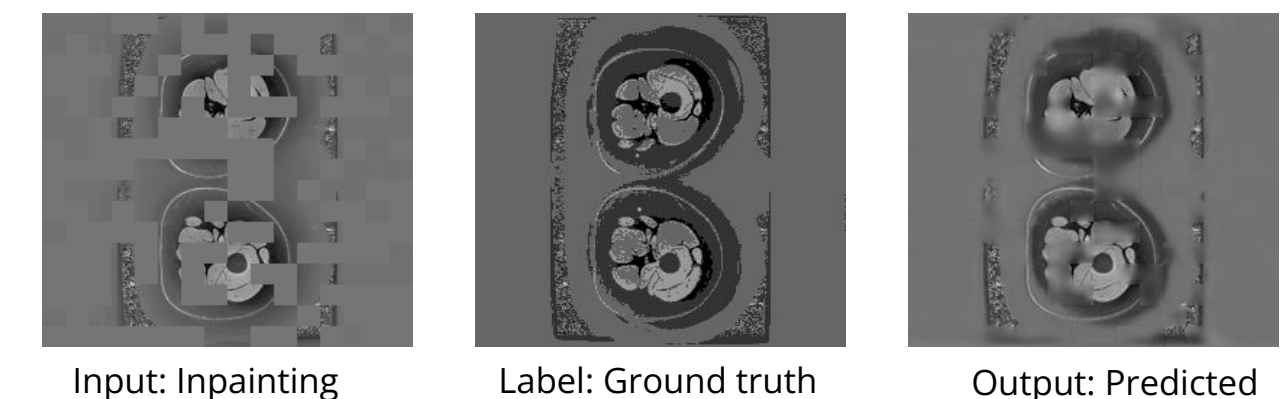
Experiment 4: tested and selected best model parameters for downstream task

- LR = 1e-5, optimizer = RMSprop, M = 0.9

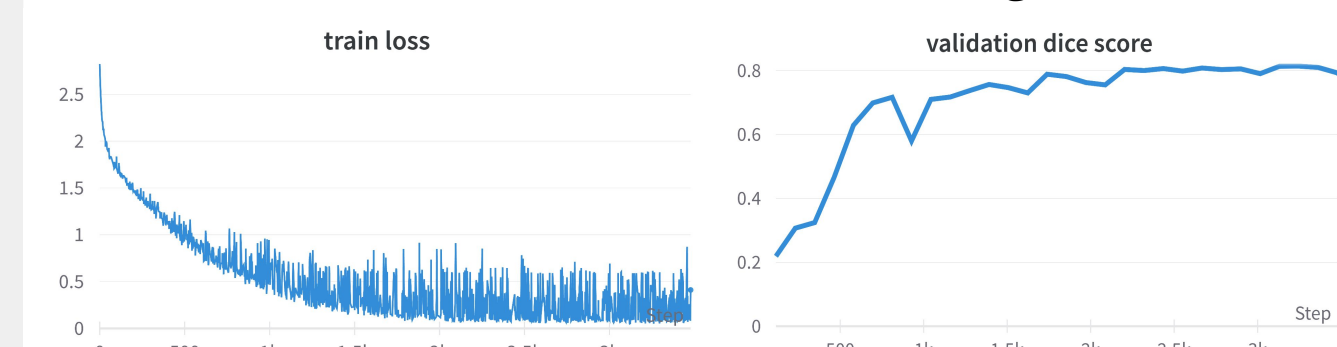


Results

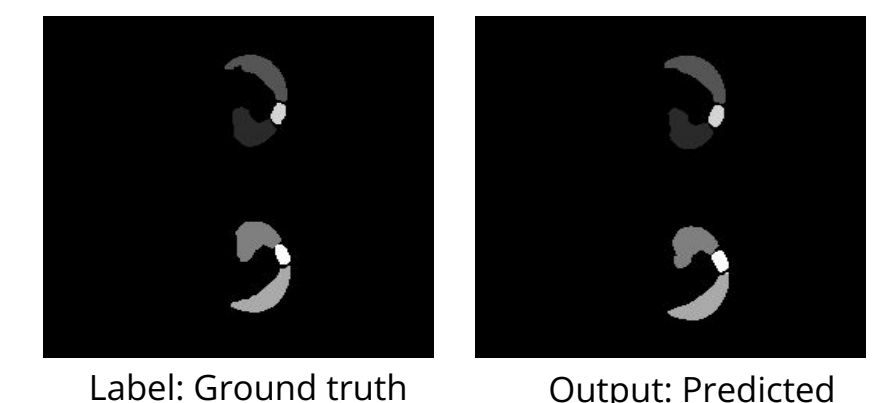
Pretext Task



Downstream Model Training



Downstream Task



U-Net encoder layers frozen for transfer learning

Dice Scores	Baseline	Mask = 10% Kernel size = 16	Mask = 50% Kernel size = 16
Validation	0.7916	0.7923	0.7912
Test	0.6226	0.6513	0.6250

Discussion

- Self supervised models tend to perform better than the fully-supervised model.
- Large gap between validation and testing Dice scores, suggesting high variance in segmentations masks.
- Loss functions for any of the training or validations sets are noisy, which is a signal for small datasets.

Future Work

- Make U-Net more deeper and larger.
- Experiment with different pretext tasks.
- Pretrain on larger, similar domain datasets.