Evaluating the Variable Stride Algorithm for Identifying Diabetic Retinopathy

Brian Danaher; Matthew Brown; Ying Zheng
Faculty Mentor: Dr. Mihhail Berezovski | Department of Mathematics. Embry-Riddle Aeronautical University
Funded by the NSF | Proposal by the Nevada National Security Site

Introduction, Problem Statement

- Neural Networks are a form of AI where layers of artificial “neurons” are connected with weights and biases
  - These weights and biases can be “trained” by showing them the correct data, passing the data through the network, and backpropagating (optimizing)
  - Convolutional Neural Networks (CNNs) use filters that create feature maps in place of weights and biases
  - These filters pass over the image, and are useful in image recognition because important features can be identified

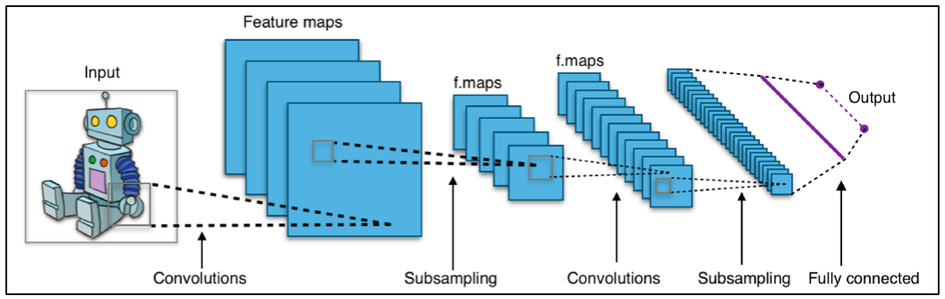

Figure #1 Illustration of a convolutional neural network

- Subsampling is a technique used to increase training speed by reducing the size of the dataset
  - This is mainly done with the avgpool and maxpool methods
- The Nevada National Security Site (NNSS) is a US Department of Energy facility dedicated to maintaining the US’ nuclear stockpile
- The NNSS has developed an algorithm in-house called Variable Stride (VS)
  - This algorithm aims to accelerate CNNs by dynamically subsampling images
  - If successful, variable stride will be used on the NNSS’ classified X-ray datasets
- The non-classified data that will be used to evaluate the algorithm is diabetic retinopathy data
  - Binary classification of the disease between stage 0 (healthy), and stage 4 (advanced)
  - 200 images, evenly split between stages 0 and 4

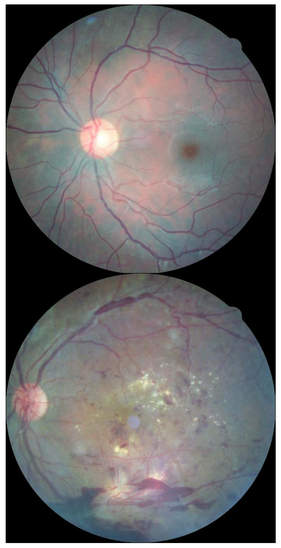

Figure #2 Comparison between a healthy retina (top), and a retina affected by stage 4 diabetic retinopathy (bottom)

Experimental Procedure

- A test was run on 130 CNN architectures, and the three with the highest accuracy were chosen for the experiment
- A test was also run on 10 striding schemes, the three which gave the highest accuracy were also chosen for the experiment
  - Looked at the activation maps to develop these schemes
- 15 combinations of networks were run 12 times each (3 network architectures, 3 striding schemes + maxpool and avgpool)
  - Test Losses, test accuracies, AUCs, epochs until convergence collected

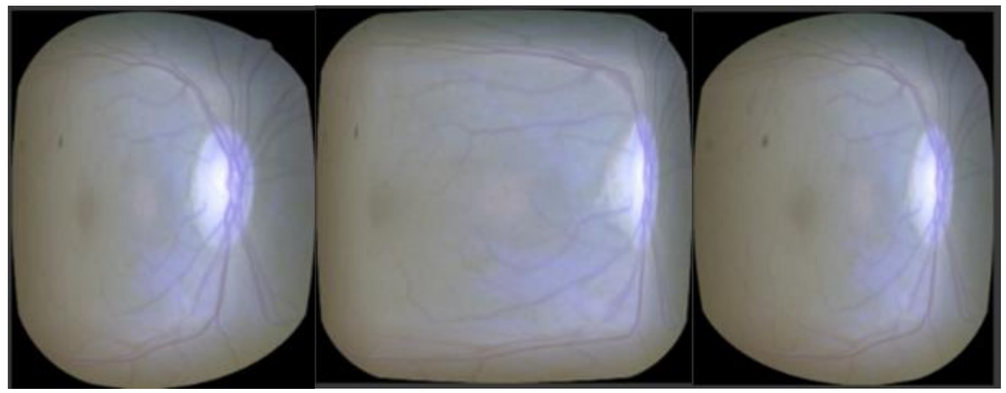

Figure #3 Demonstrations of the three striding methods tested: VS custom (left), VS center (center), and VS right (right)

Results

- T-tests for two independent sample means (95% confidence level) were run between each combination
- High variability within each set of runs led to few significant results on loss, accuracy, and AUC
- The most common significant result across these “sparse” metrics was that variable stride performed worse than both maxpool and avgpool
- Maxpool and avgpool frequently converged in less epochs than the VS schemes
  - Hypothesis 1: the entire eye is equally important for identifying diabetic retinopathy (evidence: scars can occur anywhere)
  - Hypothesis 2: the dataset is too small and overfitting occurs (evidence: in all cases, the test loss increases with the epochs)