DOE/NV/03624--1013

NEVADA NATIONAL NNSS SECURITY SITE

Managed and operated by Mission Support and Test Services

Algorithms to Improve Training for Deep Learning with Diabetic Retinopathy Images

Margaret Lund, PhD

SIAM-CSE March 1, 2021

This work was done by Mission Support and Test Services, LLC, under Contract No. DE-NA0003624 with the U.S. Department of Energy and supported by the Site-Directed Research and Development Program. DOE/NV/03624--1013.

National Nuclear Security Administration

Missions

Maintain the safety, security, and effectiveness of the U.S. nuclear weapons stockpile through application of unparalleled science, technology, engineering, and manufacturing

Respond to nuclear and radiological emergencies in the U.S. and abroad

Prevent nuclear weapon proliferation and reduce the threat of nuclear and radiological terrorism around the world

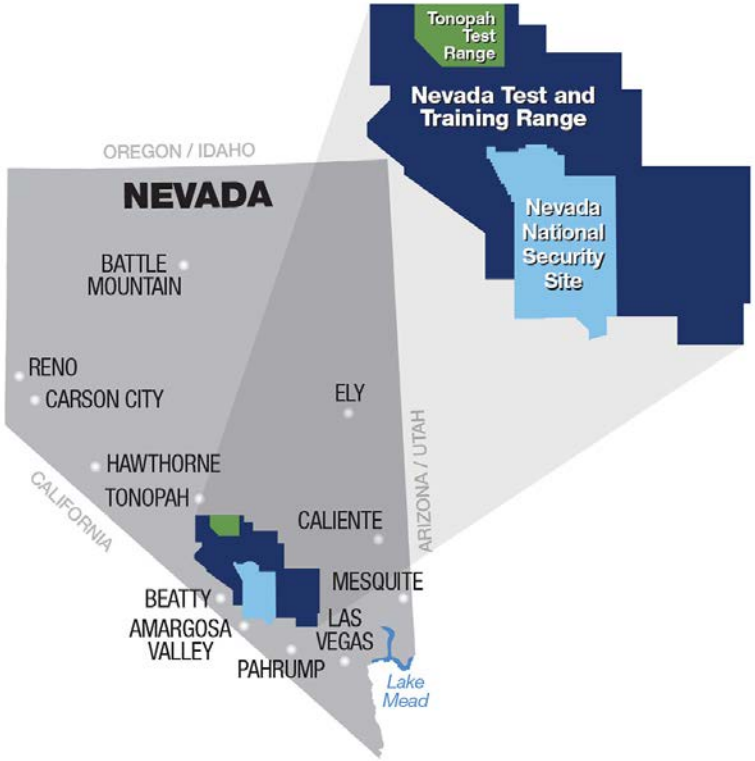
Provide the U.S. Navy with militarily effective nuclear propulsion plants and ensure their safe, reliable, and long-lived operation

Workforce & Infrastructure

Nevada National Security Site

Mission Statement

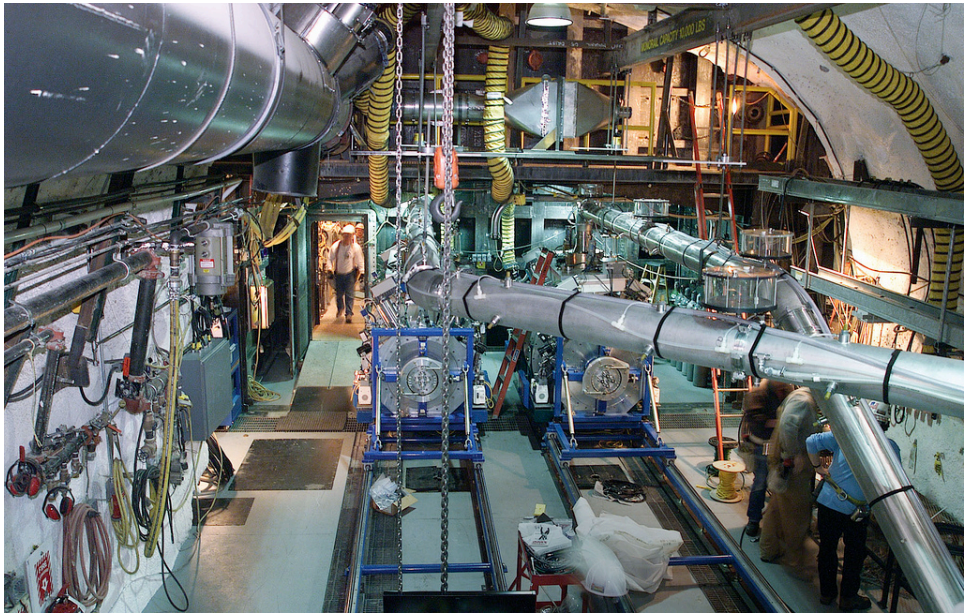
- Support the stewardship of the nation's nuclear deterrent
- Provide nuclear and radiological emergency response capabilities and training
- Contribute to key nonproliferation and arms control initiatives



- Execute national-level experiments in support of the National Laboratories
- Work with national security customers and other federal agencies on important national security activities
- Provide long-term environmental stewardship of the NNSS's Cold War legacy.

Neural Networks & Data Sets

- Our larger radiographic data sets contain ~2 million images
- Neural networks model data association with qualitative or quantitative metrics
  - Particularly useful for extremely large, labeled training sets

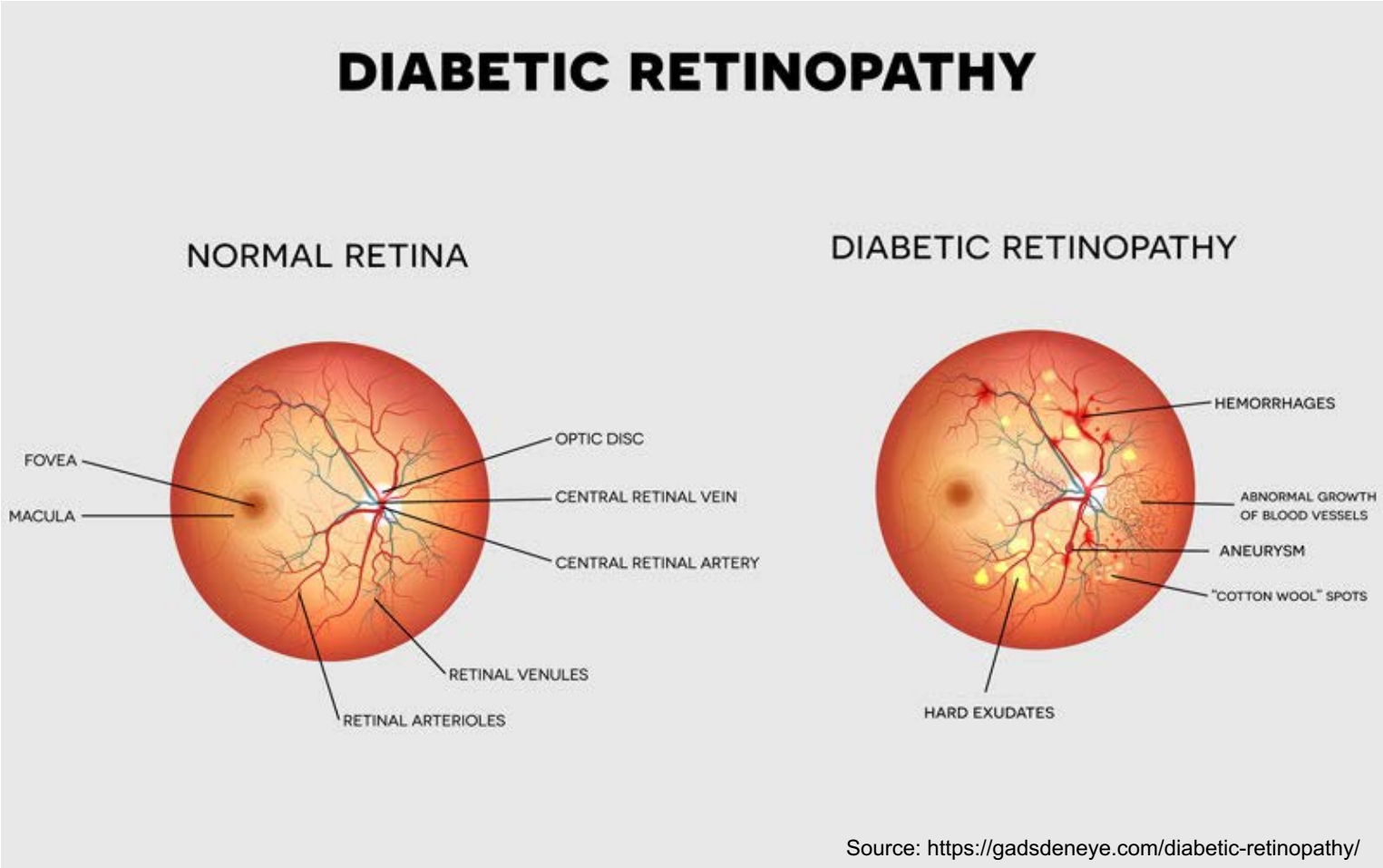


Cygnus radiography machine

http://nnsa.energy.gov/cygnus

Neural Networks & Data Sets

- The diabetic retinopathy data set is a collection of 10,000+ publicly available eyeball images



Source: https://gadsdeneye.com/diabetic-retinopathy/

Diabetic Retinopathy Data

- 5 Categories: Medical diagnosis ranges from 0 (no evidence of disease) to 4 (disease progression is advanced)

Class 0	Class 1	Class 2	Class 3	Class 4
Patient 3148, right	Patient 11505, left	Patient 818, left	Patient 3096, right	Patient 22568, left

Motivation for Better Pooling Method

- Convolutional neural networks (CNNs) are commonly used on image data sets of this size
- CNNs require significant downsampling/ pooling of the data in order to run on personal computers
  - limited memory and computing power
- Traditional pooling methods such as MaxPool and AvgPool downsample with the same stride across the entire image, even if the user knows a priori that some image regions contain more predictive power than other regions

Variable Stride Algorithm

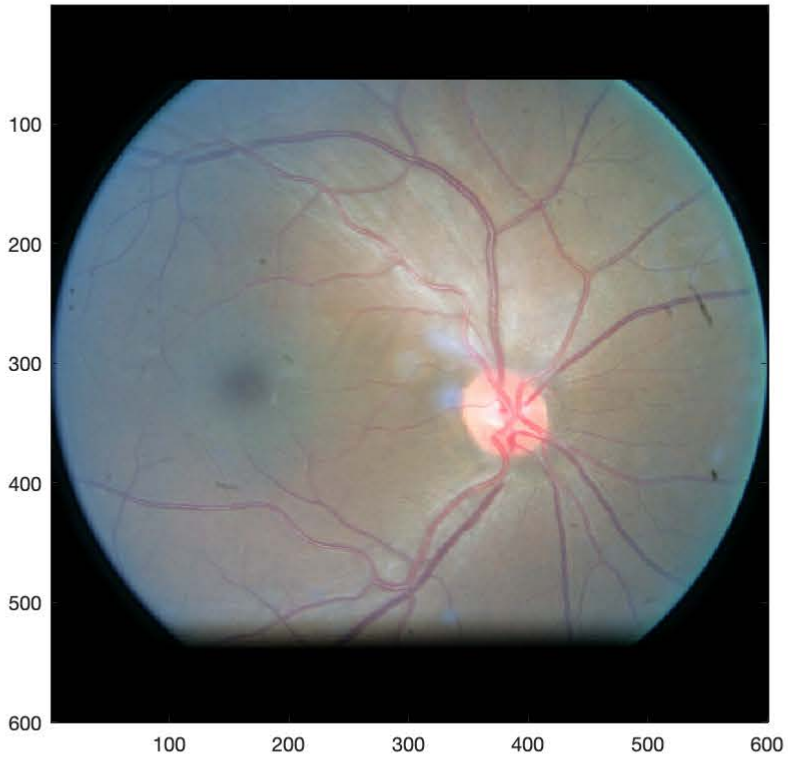
- Goal: develop a more efficient pooling method: Variable Stride
  - Use larger stride over “unimportant” regions
  - Use smaller stride over “important” regions

- Restrictions:
  - Limit how user can choose strides to ensure output image is rectangular
  - Entire dataset must have same “important” and “unimportant” regions

- Process:
  - For horizonal and vertical stride:
    - Partition image into fractions of the whole
    - Define stride for each fractional component
  - Starting in corner of image, pool using strides defined by each region

Example:
Determine horizontal stride
Break image into “left half” and “right half”
Stride on left = 2, stride on right = 4

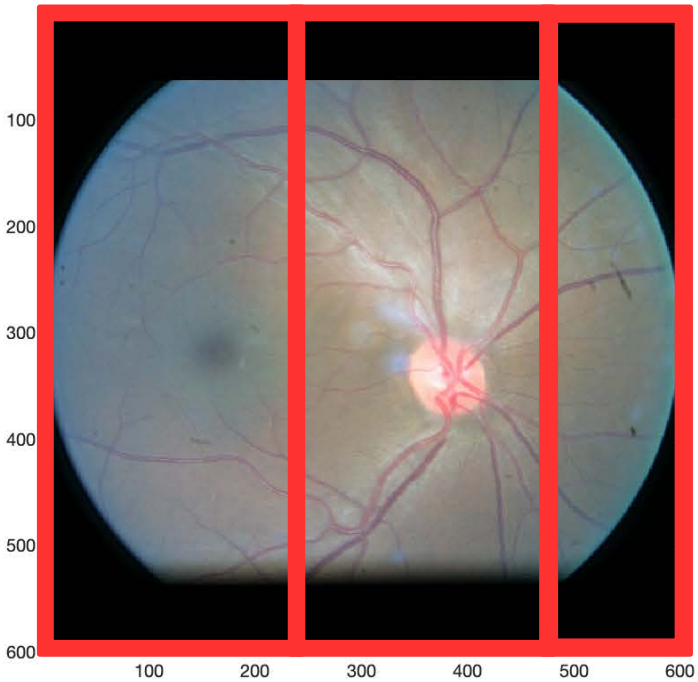
Variable Stride Example

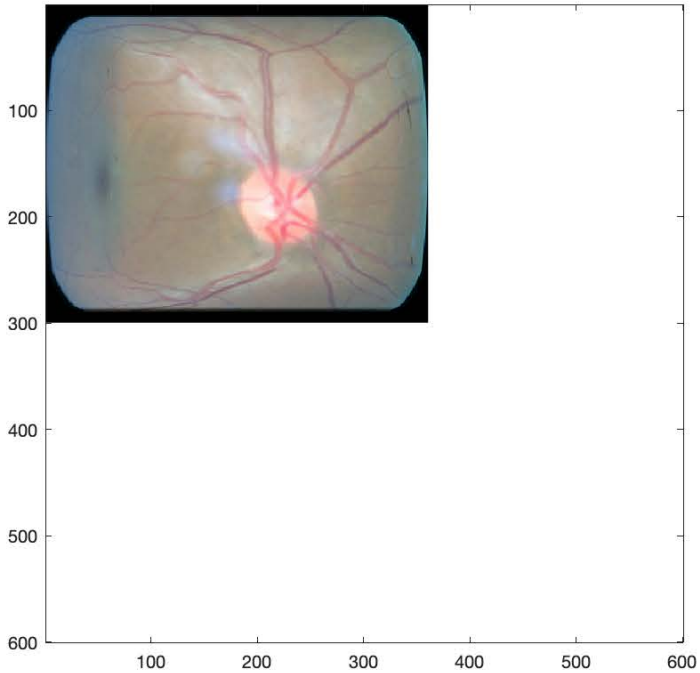


Class 2

Patient 818, left

Variable Stride Example



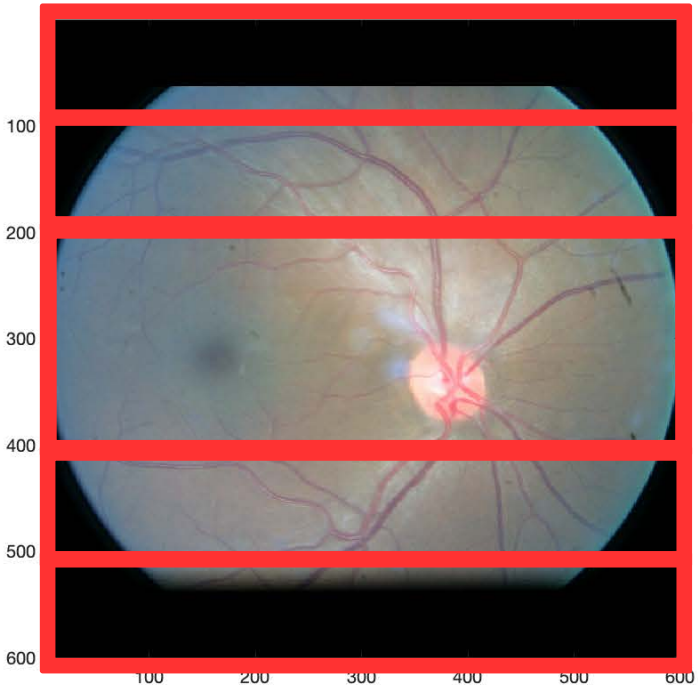


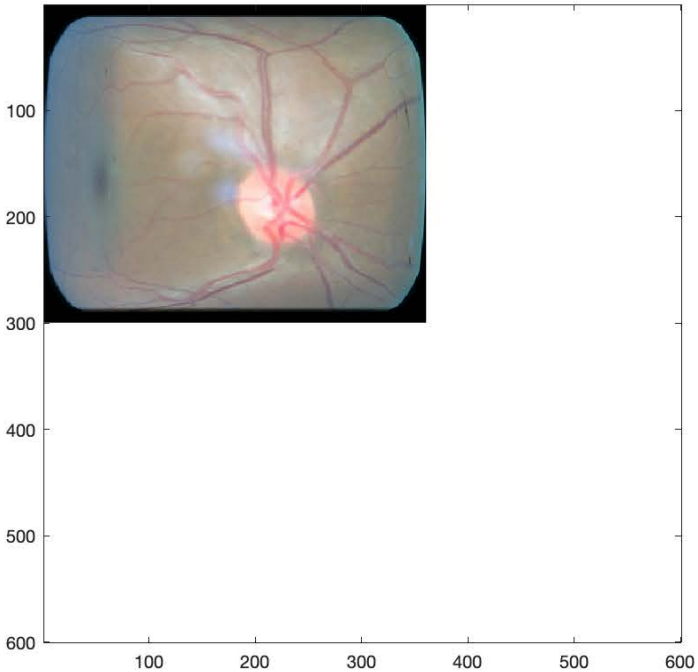
Horizontal Strides = [3, 1, 3];
Horizontal Regions = [2/5, 2/5, 1/5];

Vertical Strides = [6, 3, 1, 3, 6]
Vertical Regions = [1/6, 1/6, 1/3, 1/6, 1/6];

NEVADA NATIONAL
NNSS
SECURITY SITE
Managed and operated by
Mission Support and Test Services LLC

Variable Stride Example



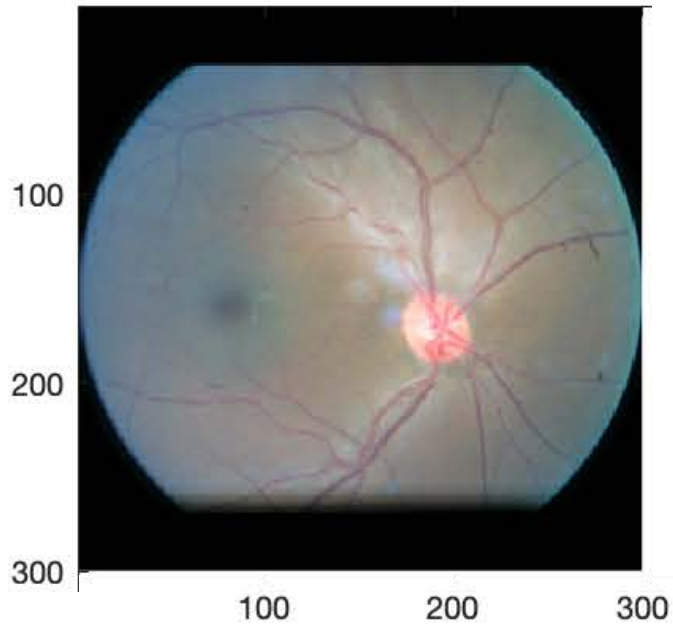


Horizontal Strides = [3, 1, 3];
Horizontal Regions = [2/5, 2/5, 1/5];

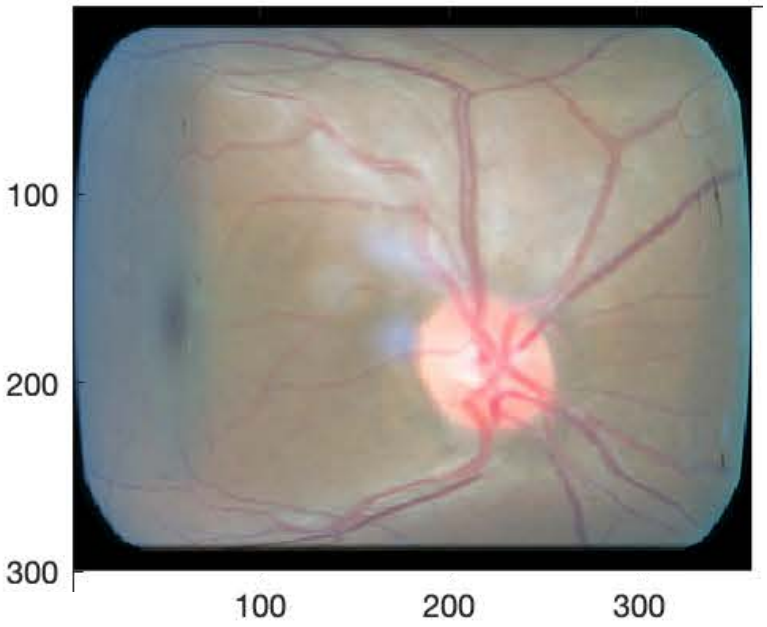
Vertical Strides = [6, 3, 1, 3, 6]
Vertical Regions = [1/6, 1/6, 1/3, 1/6, 1/6];

NEVADA NATIONAL
NNSS
SECURITY SITE
Managed and operated by
Mission Support and Test Services LLC

Variable Stride Example



AvgPool, stride=2



Variable Stride

Conclusion

- Currently partnering with Embry-Riddle Aeronautical University to study effects on training
  - Does it improve training speed?
  - Does it improve training efficiency?
  - Does it make network more robust to over-fitting?
  - Is it a valid alternative to MaxPool and AvgPool, or something else entirely?

- Questions?

- Thank you!