(U) Physics-informed deep learning with uncertainty quantification for weapons radiography

Project #: NLV-003-20 | Year 1 of 2

PI,[1,a] M. Lund, [a] M. Howard, [a,b] C. Noorda, [a,c] J. Slone

[1]lundmc@nv.doe.gov, (702) 295-3155
[a] NNSS; [b] Brigham Young University; [c] Rice University

Executive Summary

This project developed mathematically-defensible algorithms for making training more efficient and accurate for deep convolutional neural networks.

This work was done by Mission Support and Test Services, LLC, under Contract No. DE-NA0003624 with the U.S. Department of Energy, the Office of Defense Programs, supported by the Site-Directed Research and Development Program. DOE/NV/03624--1236

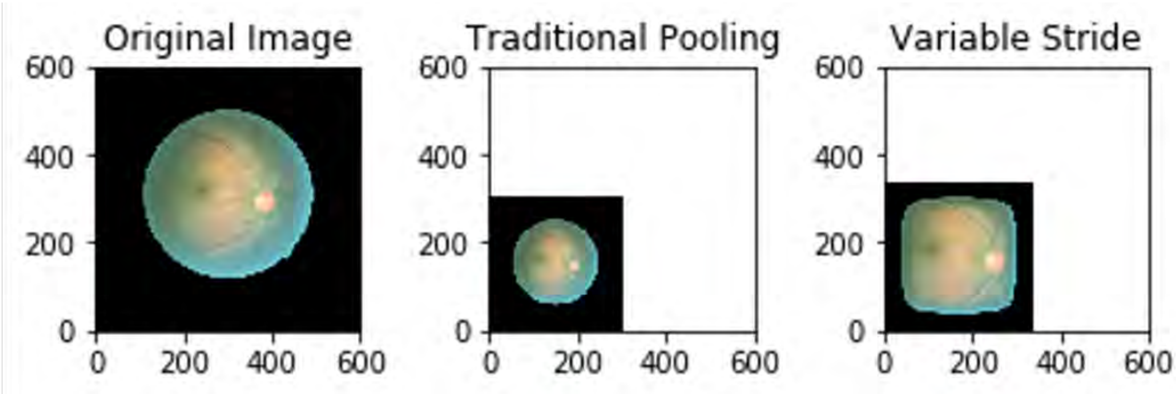
Description

Deep neural networks called Convolutional Neural Networks (CNNs) accept an image as input, use multiple layers to progressively extract higher level features from the images, and output some value/category of interest. Much can be learned by applying CNNs to image datasets within the Complex. For example, the Labs rely on the intuition of expert weapons modelers and costly simulations to predict performance of a weapon design but moving forward, deep learning will augment the need for heavy-handed supervised analysis. However, existing networks are typically for high-resolution categorical image data and need to be tailored to fit NNSA needs.

This project focused on developing novel ideas to improve training in CNNs and resulted in three main deliverables. First, we developed a more effective method for pooling images called *variable stride*, allowing networks to be trained using less memory space, fewer trainable parameters, and less computational power, all while retaining the most important parts of the image. While traditional pooling methods downsample equally across an entire image, a priori knowledge from SMEs of the most information-dense regions, combined with cutting-edge CNN activation mapping, allows us to inform variable stride to focus on certain regions of the image, as seen in Figure 1, increasing confidence in our models and aiding in network interpretability.

Second, a new, unsupervised method called *area under the margin* was adopted for eliminating training images which are detrimental to the training process- either mislabeled, poor quality, or otherwise misleading. This resulted in improved network accuracy for all 3 tested datasets and resulted in an additional discovery of determining network overtraining in real-time, allowing us to halt training once overtraining is discovered, saving valuable time and computing resources.

Finally, we began developing a methodology for identifying whether new images fall within the scope of a trained network, especially when the new images differ in some significant way from the training images. This is a major concern here at the NNSS, where image data is limited and there are often changes in design/ setup from experiment to experiment. Year 2 of this project will focus on statistical methods for quantifying uncertainty in network outputs.



Pooling methods downsample an image to summarize the features from the previous network layer. While traditional pooling methods downsample equally across an entire image, variable stride focuses on information-dense regions of the image, aiding in network interpretability.

Conclusion

Our 3-pronged approach to addressing issues with our datasets and CNNs improved training efficiency and network accuracy on 3 datasets. We collaborated with partners at 3 national labs, advised a research team of 4 undergraduate students at Embry Riddle Aeronautical University as their industry partner, advised 3 summer REU students, and supported 2 summer interns. Next year will focus on developing statistical methods for quantifying uncertainty in CNN output values, which will increase confidence in our models and aid in network interpretability. Additional goals include utilizing time on LLNL HPCs, presenting our work at 2 conferences, sponsoring a senior capstone project at BYU, and supporting 2 summer interns.

Mission Impact/Benefit

This year's SDRD laid the groundwork for deep learning capabilities within the radiography analysis team here at the NNSS. Cultivating a team of machine learnists who can not only apply existing CNN methods, but also develop new methods to suit our unique needs fills a knowledge gap and advances our abilities to extract vital information from our limited image data. Subsequent years of this project will further these developments and adapt to work on new experimental image data.

White Papers and Presentations

Three white papers: *Variable Stride: A New Approach to Pooling for Convolutional Neural Networks*, *Identifying Overtrained Networks using Area Under the Margin Algorithm*, and *Determining if New Data is Within the Scope of a Trained Network*.

Lund, M., et al. "Algorithms to improve training for deep learning with diabetic retinopathy data." Minisymposium talk at The Society for Industrial and Applied Mathematics' Conference on Computational Science and Engineering, March 2021.

Lund, M., et al. "Deep learning for diabetic retinopathy images." Presented in the NNSS Signal Processing and Applied Math Team's Science Project Series, March 2021.