

# Final Report: Reducing the burden of massive training data for deep learning

LESLIE N. SMITH

Email: [leslie.smith@nrl.navy.mil](mailto:leslie.smith@nrl.navy.mil)

Navy Center for Applied Research in AI (NCARAI)  
Information Technology Division

Keywords: Deep learning, few-shot learning, learning with limited labeled data, unsupervised learning, semi-supervised learning, Computer vision, image classification, evaluation metrics, noisy labels

August 30, 2021

This **NCARAI Technical Note** is the final project report for the NRL Base Program project reducing the burden of massive training data for deep learning.

**DISTRIBUTION STATEMENT A.** Approved for public release: distribution unlimited.

## **Introduction**

Over the past decade, the amount of effort in deep learning (machine learning with many layered neural networks) has increased exponentially due to the impressive performance of these networks on historically difficult problems, such as computer vision, understanding natural languages, and decision-making. Correspondingly, the importance of deep learning to the Navy has become increasingly clear in the past several years.

This impressive performance by neural networks is generally achieved by supervised learning, in which the model trains on large, balanced labeled datasets. Studies have shown that more training data allows networks to reach higher accuracy and generalize better, which has led to labeled datasets typically containing thousands to millions of labeled images.

Consequently, research in deep learning has exploded following the creation of large benchmark datasets, such as ImageNet and the Enron dataset.

While raw data is often plentiful for real world applications, labeling data is hard. Manually labeling thousands of data samples is labor intensive and can be a barrier in practice. Furthermore, in numerous fields (such as medicine, defense, and other scientific fields) a limited number of people who are experts in their fields can often only correctly classify data. Therefore, the goal of this project was to be able to achieve high performance without having to manually label massive amounts of data (i.e., learning with limited labeled data).

## **Technical Objective**

We created new methods for training deep neural networks that substantially reduced the current requirements for massive labeled training datasets and allow application of deep learning in situations where massive labeling of training data expensive or not possible.

## **Technical Approach**

### **Optimizing training Neural networks**

At the start of this Base Program, we were continuing our efforts in optimizing the training of neural networks with large labeled datasets. This effort led to important work in understanding the proper setting of the various hyper-parameters used by training algorithms [8].

### **Few-shot and semi-supervised learning**

Reducing the amount of training data required to enable deep networks to get state-of-the-art results is an open and active area of research. There are several methods in the literature, such as data augmentation, generating synthetic data, and transfer learning. We investigated the most relevant approaches for few shot leaning, such as metric learning and meta-learning methods. In summary, our experiments with few-shot metric learning and Meta-learning methods, such as Model Agnostic Meta-Learning (MAML) showed that these popular methods were inferior to semi-supervised learning (SSL) methods so we concentrated our efforts on maximizing performance with SSL.

## Technical Progress

### Optimizing training neural networks

Most important of the hyper-parameters is the learning rate. We invented the Cyclical Learning Rates (CLR) method [7] that is now widely recognized in the deep learning community, with over 1300 citations at this writing. These early efforts led to being one of the first to demonstrate speeding up training with large learning rates and large batch sizes, which I called super-convergence [9]. My experiments with super-convergence demonstrated up to an order of magnitude speed up in training without any loss in accuracy. Finally, I wrapped up these efforts with a report "Best practices for applying deep learning to novel applications" [10].

### Few-shot and semi-supervised learning

In FY18, we investigated several new architectures, loss functions, and data augmentation methods, such as analytically solving for the weights of the final linear layer [5], adding new heads to a network for new tasks (i.e., "HydraNets" [4]), and a multi-scalar contrastive loss function. We found the feature extractor network to be sensitive to its training data and adding a new top layer had mixed results. That is, in cases where transfer learning worked well, this method also worked. We also tested data augmentation methods, such as between class/mixup data augmentation [1, 2].

In FY19 our progress focused on the active area of few-shot meta-learning. In the realm of few-shot learning, prototypical networks [3] is a standard method that we have investigated extensively. We also developed a new incremental few-shot learning architecture that was presented at the Navy Applications of Machine Learning Workshop (NAML). Under the umbrella of meta-learning, the model agnostic meta-learner (MAML) [5] was also an integral part of our research. In addition, transfer learning and domain adaptation are commonly used methods when data for a specific task is scarce but does exist for similar tasks. In addition, our investigations included dynamic data augmentation, online batch selection, adding channels to RGB data, large batch size training, and upscaling the image resolution. The techniques we studied aligned well with the mainstream research and we made incremental progress in these fields.

In FY20 our investigations focused on semi-supervised learning, which proved to be particularly productive. Semi-supervised learning is a hybrid between supervised and unsupervised learning, which combines the benefits of both to better match the scenario of real-world problems. As with supervised learning, semi-supervised learning defines a task (i.e., classification) from labeled data but typically, it uses many fewer labeled samples. Like unsupervised learning, semi-supervised learning leverages feature learning from unlabeled data to the greatest extent possible. The term "one-shot" in our context means that only one example per class is labeled and the vast majority of the training data remains unlabeled.

In FY20 we demonstrated for the first time ever the potential for building one-shot semi-supervised (BOSS) learning up to attain test accuracies that are comparable to fully supervised learning [11, 12]. One-shot semi-supervised learning is generally viewed as a very challenging problem, and hence had not been studied.

Recently there have been a series of papers on semi-supervised learning and we built our study on a recent method called FixMatch [8] The FixMatch algorithm is primarily a combination of consistency regularization and pseudo-labeling. Consistency regularization utilizes unlabeled data by relying on the assumption that the model should output the same predictions when fed perturbed versions as on the

original image. Consistency regularization has recently become a popular technique in unsupervised, self-supervised, and semi-supervised learning. Several researchers have observed that strong data augmentation should not be used when inferring pseudo-labels for the unlabeled data but should be employed for consistency regularization. Pseudo-labeling is based on the idea that one can use the model to obtain artificial labels for unlabeled data by retaining pseudo-labels for samples whose probability are above a predefined threshold.

In our groundbreaking papers [11, 12] we observed that one-shot semi-supervised learning is plagued by class imbalance problems, even when equal numbers of labeled and unlabeled samples per class were used. In our context, class imbalance refers to a trained network with near 100% accuracy on a subset of classes and has poor performance on other classes. To combat class imbalance, we tested several variations of methods found in the literature for class imbalance problems, which refers to the situation where the number of training data per class vary substantially. We are the first to demonstrate that these methods significantly boost the performance of one-shot semi-supervised learning. In addition, we demonstrate that good prototypes are crucial for successful semi-supervised learning and propose a practical prototype replacement method for the poorly performing classes.

Furthermore, we followed up on the success of our project by applying these techniques to real world, Navy related unlabeled data. We applied our methods to train deep neural network classifiers using largely unlabeled data for hyperspectral imagery (HSI) and multispectral imagery (MSI) data. Unlike more traditional aerial photos, which have only three channels (e.g., red, green, blue), HSI has dozens or hundreds of channels across the visible and near-infrared spectrum. The precise information provided by these channels can accurately characterize materials. Each pixel of HSI imagery can thus be regarded as a high-dimensional vector whose entries correspond to the spectral reflectance of a specific material. Our Technical Report "Semi-supervised Learning on Hyperspectral Imagery" [13] describes the results of our effort.

In addition to limiting the number of labeled samples, working with hyperspectral data has numerous challenges, including per pixel classification, imbalanced data, and inconsistent performance metrics. Even with all these challenges, our team made significant progress in being able to train the hyperspectral dataset AeroRIT with few labeled examples. Our work compared four algorithms: fully supervised learning with limited labeled examples, the state-of-the-art in HSI semi-supervised learning (i.e., CCT), shallow networks, and conversion of the state-of-the-art in semi-supervised learning for semantic segmentation to work on the AeroRIT dataset. We concluded that improving HSI performance in the limited labeled pixel regime remains an open problem.

Our efforts in FY21 also included investigating class inequality in classification application for two scenarios: in the presence of noisy labels and in the limited labeled data scenarios. Our NRL Memorandum report "Towards High Performance Network Training with Noisy Label Datasets" [14] describes our effort in this direction. Follow-up work in this direction was studied based on the idea that some classes are easier to classify, while others are difficult due to factors such as confusion between class objects (i.e., cats and dogs) and presence (or absence) of distinctive features. Hence, noisy labels in these easy classes should have a substantially lower impact on performance than noisy labels in the hard classes. Our experiments demonstrated that noisy labels between high confusion classes potentially increased the class classification error by 4x the error of noisy labels between low confusion classes.

Finally, we investigated class inequality in learning with the limited labeled data scenario. We demonstrated that an imbalance in the numbers of training samples per class for both supervised and semi-supervised learning creates a tradeoff in the various classes' performances. Extra data samples in weak classes improves the performance of those classes but this happens at the expense of the performance of the strong classes. However, the mean test accuracy over all the classes remains a function of the total number of labeled samples, regardless of the distribution. Our experiments show that training with more labeled samples of weaker performing classes results in a trained model, which performs more uniformly across all classes. An NRL Memorandum Report describing this work is currently being written.

## References and Publications

### References

1. Tokozume, Yuji, Yoshitaka Ushiku, and Tatsuya Harada. "Between-class Learning for Image Classification." arXiv preprint arXiv:1711.10284 (2017).
2. Zhang, Hongyi, Moustapha Cisse, Yann N. Dauphin, and David Lopez-Paz. "mixup: Beyond empirical risk minimization." arXiv preprint arXiv:1710.09412 (2017).
3. Snell, Jake, Kevin Swersky, and Richard Zemel. "Prototypical networks for few-shot learning." In Advances in Neural Information Processing Systems, pp. 4080-4090. 2017.
4. Teja Mullapudi, Ravi, William R. Mark, Noam Shazeer, and Kayvon Fatahalian. "HydraNets: Specialized Dynamic Architectures for Efficient Inference." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8080-8089. 2018.
5. Finn, Chelsea, Pieter Abbeel, and Sergey Levine. "Model-agnostic meta-learning for fast adaptation of deep networks." In Proceedings of the 34th International Conference on Machine Learning-Volume 70, pp. 1126-1135. JMLR. org, 2017.
6. Sohn, Kihyuk, David Berthelot, Chun-Liang Li, Zizhao Zhang, Nicholas Carlini, Ekin D. Cubuk, Alex Kurakin, Han Zhang, and Colin Raffel. "Fixmatch: Simplifying semi-supervised learning with consistency and confidence." arXiv preprint arXiv:2001.07685 (2020).

### Publications

7. L. N. Smith, "Cyclical Learning Rates for Training Neural Networks," 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), 2017, pp. 464-472, doi: 10.1109/WACV.2017.58.
8. Smith, Leslie N. "A disciplined approach to neural network hyper-parameters: Part 1--learning rate, batch size, momentum, and weight decay." arXiv preprint arXiv:1803.09820 (2018).
9. Smith, Leslie N., and Nicholay Topin. "Super-convergence: Very fast training of neural networks using large learning rates." In Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications, vol. 11006, p. 1100612. International Society for Optics and Photonics, 2019.
10. Smith, Leslie N. "Best practices for applying deep learning to novel applications." arXiv preprint arXiv:1704.01568 (2017).

11. Smith, Leslie N., and Adam Conovaloff. "Empirical Perspectives on One-Shot Semi-supervised Learning." arXiv preprint arXiv:2004.04141 (2020).
12. Smith, Leslie N., and Adam Conovaloff. "Building One-Shot Semi-supervised (BOSS) Learning up to Fully Supervised Performance." arXiv preprint arXiv:2006.09363 (2020).
13. Smith, Leslie N., Kristen Nock, Krish Ganotra, Colin Olson, and Mark Snyder. Semi-supervised Learning on Hyperspectral Imagery. NAVAL RESEARCH LAB WASHINGTON DC, 2021.
14. Smith, Leslie N., and Elizabeth A. Gilmour. Towards High Performance Network Training with Noisy Label Datasets. NAVAL RESEARCH LAB WASHINGTON DC WASHINGTON United States, 2020.

## **Machine Learning Seminars**

As part of our research activities in machine learning, I organized and hosted a weekly (in-person and recently virtual) seminars on topics related to ML (or more broadly on AI). These events ranged from overviews of ML research topics to ML related applications. Over the four years of this Base Program, worldwide growth of interests, scientific effort, and ML applications has exploded, and along with it has interest in these seminars.

## **Naval/Marine Corps Needs**

We have identified several applications of interest to the Navy/NRL that potentially can benefit from deep learning methods, including:

1. Control of the electromagnetic spectrum with deep reinforcement learning (DeepRL).
2. Cyber-security and monitoring with recurrent neural networks (RNN) and cyber defense with DeepRL.
3. Intelligent decision aids and autonomous control of robots and UxVs with DeepRL.
4. Robotic perception with convolutional neural networks (CNN).
5. Situation awareness and target or threat identification and tracking with CNNs.
6. Autonomous control of distributed radar systems in a dynamic EM environment with DeepRL.
7. Fast and accurate approximation of meteorological simulations.
8. Discovery of protein binders in genes for precision biology.

These applications share the research challenge of training the deep networks when labeled data is scarce or expensive and when self-play is not possible for DeepRL.

## **Summary of Findings**

The following includes a summary of key issues, or findings, learned during our developments in this work unit. This summary clearly illustrates that the Learning with Limited Labeled Data community still has a number of challenges ahead before it will be more broadly adopted. These issues/findings include:

- The total performance of metric learning approaches, such as prototypical networks, and of meta-learning approaches, such as MAML, fall far lower than the performance for fully supervised methods. On the other hand, there are niche applications (i.e., handling novel classes) where these methods are valuable.

- Use of a frozen feature extractor and training only the last layers with few labeled examples gave mixed results. Domain shifts of the new small datasets significantly impacted the performance of the feature extractor.
- Substantial progress has been made over the past couple of years in unsupervised and semi-supervised learning. High performance with few-labeled training examples have been demonstrated for several benchmark datasets.
- Few-shot semi-supervised learning (SSL) is plagued with substantial performance differences between classes. Adding class imbalance algorithms to SSL improved training performance.
- We demonstrate that good prototypes are crucial for successful semi-supervised learning.
- Methods using contrastive representation learning is currently popular and has been part of several recent \SOTA algorithms appearing in the literature, including FixMatch [6].
- We found that improving HSI performance in the limited labeled regime remains an open problem.
- Our experiments show that training with more labeled samples of weaker performing classes results in a trained model, which performs more uniformly across all classes.

## Conclusions

Over the past four years, tremendous progress has been made to reduce the burden of manually labeling large numbers of training samples in order to train deep neural networks. Unsupervised and semi-supervised learning has shown the most promise and produces the best performance. Contrastive representation learning for unsupervised and semi-supervised learning has gain a great deal of attention and can also be used in a pretraining context.

Much of this progress has been limited to standard benchmark dataset but applications still lags far behind. Our work on applying these methods to hyperspectral imagery met with limited success and indicates that there is still much research and engineering to be done to use these methods on real-world applications. This is particularly true for Navy applications where less attention is given by academia and the high tech industry.

I will end with a general comment on what I see as a major challenge for NRL. One of the chief challenges for the Naval Center for Applied Research in Artificial Intelligence (NCARAI) is hiring the best and the brightest graduates in deep learning. Due to the worldwide growth and high demand, the highest quality graduates in DL are going to academia and the high Tech industry. The Chinese have openly stated that they plan to lead the world in areas of artificial intelligence and their scientist are learning in contributions to every major conference. Russian president Vladimir Putin said that whoever reaches a breakthrough in developing artificial intelligence would come to dominate the world. In recognition of the new arms race in AI, the 2018 National Defense Strategy states: ``The Department will invest broadly in military application of autonomy, artificial intelligence, and machine learning...to gain competitive military advantages''. Having the best and the brightest in deep learning at NCARAI is an open problem on which our country's military competitive advantage hangs.