REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
04-01-2023	Final Report	1-May-2021 - 30-Apr-2022

4. TITLE AND SUBTITLE	
Final Report: Fast and Reliable Object Identification via Guaranteed Learning on Neural Networks	5a. CONTRACT NUMBER W911NF-21-1-0255
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER 611102
6. AUTHORS	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES	8. PERFORMING ORGANIZATION REPORT NUMBER
Rensselaer Polytechnic Institute Rensselaer Polytechnic Institute 110 8th Street Troy, NY 12180 -3522	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	ARO
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) 78463-MI.8

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Meng Wang
UU	UU	UU	UU		19b. TELEPHONE NUMBER 518-276-3842

Standard Form 298 (Rev 8/98)
Prescribed by ANSI Std. Z39.18

RPPR Final Report
as of 13-Mar-2023

Agency Code: 21XD

Proposal Number: 78463MI **Agreement Number: W911NF-21-1-0255**

INVESTIGATOR(S):

Name: Meng Wang
Email: wangm7@rpi.edu
Phone Number: 5182763842
Principal: Y

Organization: **Rensselaer Polytechnic Institute**
Address: Rensselaer Polytechnic Institute, Troy, NY 121803522
Country: USA
DUNS Number: 002430742 EIN: 141340095

Report Date: 31-Jul-2022 Date Received: 04-Jan-2023
Final Report for Period Beginning 01-May-2021 and Ending 30-Apr-2022

Title: Fast and Reliable Object Identification via Guaranteed Learning on Neural Networks

Begin Performance Period: 01-May-2021 **End Performance Period:** 30-Apr-2022

Report Term: 0-Other

Submitted By: Meng Wang Email: wangm7@rpi.edu
Phone: (518) 276-3842

Distribution Statement: 1-Approved for public release; distribution is unlimited.

STEM Degrees: 3 **STEM Participants:** 5

Major Goals: Army networks and systems are generating and collecting vast amounts of data with different modalities, sampling rates, and accuracies continually. The obtained data often suffer from quality issues due to unreliable or defective sensors, poor atmospheric and lighting conditions, and cyber data attacks from malicious intruders. The processing and information extraction from these massive amounts of data for real-time decision making becomes exceptionally challenging. Anomaly detection and object identification are two tasks at the heart of situational awareness for Army applications. Anomaly detection sends alarms for abnormal events, and object identification locates human objects and predicts their movements. Compared with the classification and identification methods in civilian applications, the methods in a military application must have much higher reliability, as a minor error can potentially lead to the loss of lives of military personnel. Therefore, the deployed identification methods need to have analytical performance guarantees rather than numerical success only. This proposal will develop computationally efficient anomaly detection and object identification methods from large amounts of data. One distinctive feature of developed methods in this proposal is that they are accompanied by analytical performance guarantees which enable reliable implementation in military operations.

Accomplishments: See attachment.

Training Opportunities: Three PhD students and two undergraduate students have contributed to this project.

Results Dissemination: The results have been disseminated through two journal papers and five conference papers, including IEEE Transactions and top AI conferences like NeurIPS, ICML, and ICLR.

J1 Ming Yi, Meng Wang, Evangelos Farantatos and Tapas Barik. Bayesian Robust Hankel Matrix Completion with Uncertainty Modeling for Synchrophasor Data Recovery. ACM SIGENERGY Energy Informatics Review, 2022.

J2 Ming Yi and Meng Wang. Bayesian Energy Disaggregation at Substations with Uncertainty Modeling. IEEE Transactions on Power Systems, 2022, 37(1): 764-775.

C1 Hongkang Li, Meng Wang, Sijia Liu, Pin-Yu Chen and Jinjun Xiong, Generalization Guarantee of Training Graph Convolutional Networks with Graph Topology Sampling, in Proc. of 2022 International Conference on Machine Learning (ICML), July 2022. (acceptance rate: 21.9%)

C2 Ming Yi and Meng Wang, Joe H. Chow, Recent Results of Energy Disaggregation with Behind-the-Meter Solar Generation, in Proc. of the 11th Bulk Power Systems Dynamics and Control Symposium – IREP'2022, July 2022.

C3 Shuai Zhang, Meng Wang, Sijia Liu, Pin-Yu Chen and Jinjun Xiong, How Does Unlabeled Data Improve Generalization in Self-training? A one-hidden-layer Theoretical Analysis, in Proc. the Tenth International Conference on Learning Representations (ICLR), April 2022. (acceptance rate: 32.3%)

C4 Hongkang Li, Shuai Zhang, and Meng Wang, Learning and generalization of one-hidden-layer neural networks, going beyond standard Gaussian data, in Proc. 2022 56th Annual Conference on Information Sciences and Systems (CISS), Princeton, NJ, USA, 2022.

C5 Shuai Zhang, Meng Wang, Sijia Liu, Pin-Yu Chen and Jinjun Xiong, Why Lottery Ticket Wins? A Theoretical Perspective of Sample Complexity on Sparse Neural Networks, in Proc. of the Thirty-fifth Conference on Neural Information Processing Systems (NeurIPS), December 2021. (acceptance rate: 26%)

Honors and Awards: PI Meng Wang received the James M. Tien '66 Early Career Award and Grant for Faculty from Rensselaer Polytechnic Institute (RPI) in May 2022.
Graduate student Shuai Zhang won RPI ECSE Allen B. DuMont Prize for high scholastic ability in May 2021.

Protocol Activity Status:

Technology Transfer: PI Wang received the Army Polarimetric Thermal Face Database from Dr. Matthew Thielke at Army Research Lab in July 2022 to evaluate our developed learning methods on Army dataset and has obtained encouraging results.

PARTICIPANTS:

Participant Type: PD/PI
Participant: Meng Wang
Person Months Worked: 1.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Shuai Zhang
Person Months Worked: 6.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)

Participant: Hongkang Li
Person Months Worked: 6.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Ming Yi
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Undergraduate Student
Participant: Haolin Xiong
Person Months Worked: 3.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Undergraduate Student
Participant: Ruisi Jian
Person Months Worked: 3.00 **Funding Support:**
Project Contribution:
National Academy Member: N

ARTICLES:

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: ACM SIGEnergy Energy Informatics Review
Publication Identifier Type: DOI Publication Identifier: 10.1145/3527579.3527580
Volume: 2 Issue: 1 First Page #: 1
Date Submitted: 12/29/22 12:00AM Date Published: 2/1/22 5:00AM
Publication Location:
Article Title: Bayesian robust hankel matrix completion with uncertainty modeling for synchrophasor data recovery
Authors: Ming Yi, Meng Wang, Evangelos Farantatos, Tapas Barik
Keywords: PMU data recovery, robust matrix completion, Bayesian matrix completion, uncertainty modeling, Hankel matrix
Abstract: This paper considers energy disaggregation at substations (EDS) where the objective is to estimate the consumption of each load from aggregate measurements, in which whether or not some loads are consuming power is unknown to the operator. The existing EDS method cannot provide any reliable measure of the disaggregation results, while the disaggregation accuracy can vary significantly for different data due to the volatility of loads such as solar generation. This paper proposes a Bayesian-dictionary-learning-based approach to disaggregate the loads and provides an uncertainty measure of the returned estimation. Our approach learns the probability distributions of the load patterns and the decomposition coefficients from recorded data with partial labels at the offline stage. Our approach computes the mean and covariance of the probability distribution of each load consumption, estimates the load using the mean, and computes the uncertainty index based on the covariance.
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Transactions on Power Systems
Publication Identifier Type: DOI Publication Identifier: 10.1109/TPWRS.2021.3095047
Volume: 37 Issue: 1 First Page #: 764
Date Submitted: 12/29/22 12:00AM Date Published: 1/1/22 10:00AM
Publication Location:
Article Title: Bayesian Energy Disaggregation At Substations With Uncertainty Modeling
Authors: Ming Yi, Meng Wang
Keywords: Energy disaggregation, behind-the-meter solar generation, Bayesian dictionary learning, uncertainty modeling
Abstract: This paper considers energy disaggregation at substations (EDS) where the objective is to estimate the consumption of each load from aggregate measurements, in which whether or not some loads are consuming power is unknown to the operator. The existing EDS method cannot provide any reliable measure of the disaggregation results, while the disaggregation accuracy can vary significantly for different data due to the volatility of loads such as solar generation. This paper proposes a Bayesian-dictionary-learning-based approach to disaggregate the loads and provides an uncertainty measure of the returned estimation. Our approach learns the probability distributions of the load patterns and the decomposition coefficients from recorded data with partial labels at the offline stage. Our approach computes the mean and covariance of the probability distribution of each load consumption, estimates the load using the mean, and computes the uncertainty index based on the covariance.
Distribution Statement: 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

CONFERENCE PAPERS:

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: the 39th International Conference on Machine Learning
Date Received: 29-Dec-2022 Conference Date: 17-Jul-2022 Date Published: 17-Jul-2022
Conference Location: Baltimore, USA
Paper Title: Generalization Guarantee of Training Graph Convolutional Networks with Graph Topology Sampling
Authors: Hongkang Li; Meng Wang;Sijia Liu;Pin-Yu Chen; Jinjun Xiong
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: 11TH BULK POWER SYSTEMS DYNAMICS AND CONTROL SYMPOSIUM (IREP 2022)
Date Received: 29-Dec-2022 Conference Date: 25-Jul-2022 Date Published: 25-Jul-2022
Conference Location: Banff, Canada
Paper Title: Recent Results of Energy Disaggregation with Behind-the-Meter Solar Generation
Authors: Ming Yi and Meng Wang
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: The Tenth International Conference on Learning Representations (ICLR)
Date Received: 29-Dec-2022 Conference Date: 25-Apr-2022 Date Published: 25-Apr-2022
Conference Location: Virtual
Paper Title: How unlabeled data improve generalization in self-training? A one-hidden-layer theoretical analysis
Authors: Shuai Zhang; Meng Wang; Sijia Liu; Pin-Yu Chen; Jinjun Xiong
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: 2022 56th Annual Conference on Information Sciences and Systems (CISS)
Date Received: 29-Dec-2022 Conference Date: 09-Mar-2022 Date Published: 09-Mar-2022
Conference Location: Princeton, NJ, USA
Paper Title: Learning and generalization of one-hidden-layer neural networks, going beyond standard Gaussian data
Authors: Hongkang Li; Shuai Zhang; Meng Wang
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: Thirty-fifth Conference on Neural Information Processing Systems (NeurIPS) 2021
Date Received: 29-Dec-2022 Conference Date: 06-Dec-2021 Date Published: 06-Dec-2021
Conference Location: Virtual
Paper Title: Why Lottery Ticket Wins? A Theoretical Perspective of Sample Complexity on Pruned Neural Networks
Authors: Shuai Zhang; Meng Wang; Sijia Liu; Pin-Yu Chen; Jinjun Xiong
Acknowledged Federal Support: **Y**

Partners

,

I certify that the information in the report is complete and accurate:
Signature: Meng Wang
Signature Date: 1/4/23 10:03AM

Accomplishments

This project has developed efficient and reliable AI methods for object detection. Our major technical contribution is to reduce computational costs with guaranteed improved performance. Our major accomplishments include the following three aspects:

1. Neural Network pruning to reduce the computational cost and improve the detection accuracy [P6].

Network pruning removes unnecessary weights in neural networks, which is beneficial to building efficient AI learners that require less storage space, less expensive hardware, and less inference time. For example, earlier numerical experiments [1] suggest that over 90% of the parameters can be removed without harming the test accuracy. When using VGG-16 (a 16-layer Convolutional Neural Network) on the ImageNet dataset, training on the pruned model requires only 21% floating point operations per second (FLOPS) and 15.6% storage overhead in storing the model [2]. Moreover, recent studies [3–7] indicate that sparsification can speed up the convergence rate and even improve test accuracy, which leads to further efficient training of neural networks. For example, a matched sub-network with sparsity varying from 40% to 90% achieves higher test accuracy than the original dense network in various network architectures like ResNets [8, 9], BERTs [6], and GNNs [4]. Such numerical findings are summarized as the *lottery ticket hypothesis* (LTH) in [3], which hypothesises that there exists a good subnetwork, named "winning tickets," that achieves faster convergence rates and improved test accuracy compared with the original dense network.

Despite the empirical evidence of the winning tickets in many case studies, the theoretical explanation of LTH and its improved generalization is limited. The existing theoretical works are mainly from the scope of model compression, i.e., finding a sub-network that achieves a tolerable loss in either expressive power or training accuracy, compared with the original dense network [10–12]. None can provide theoretical support for the *improved* performance achieved by "winning tickets." As a first step to bridging the gap between numerical success and a theoretical understanding of pruned networks, we provided the first theoretical characterization of training a "winning ticket" and delivered a theoretical explanation for the widely-used LTH in [13].

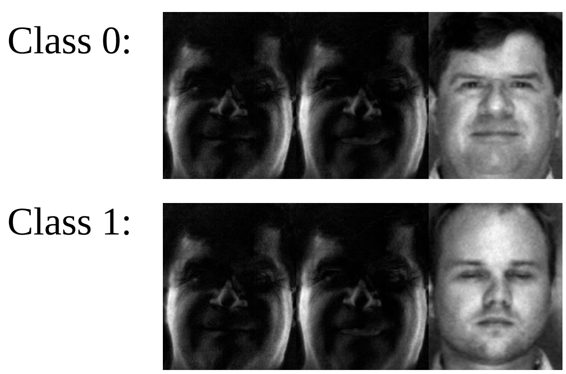

Figure 1: Illustration of data. Each input consists of two polarimetric thermal images from the same person and one visible image. The label is 0 if the visible image represents the same person as the thermal images, and 1 otherwise.

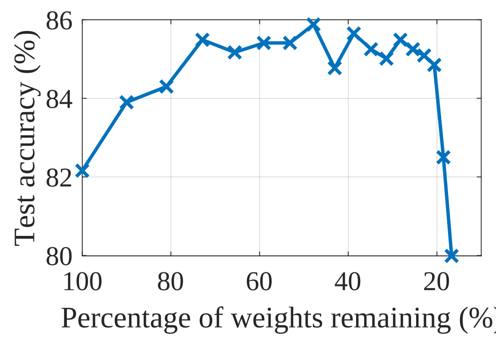

Figure 2: The test accuracy for iterative magnitude-based pruning method.

Specifically, compared with training the dense network, our theorems characterize that training a "winning ticket" needs fewer training samples, enjoys a faster convergence rate, and achieves improved test accuracy, which justifies the widely-used LTH. Also, our theorems provide the quantitative characterizations of the improved performance and the pruning rate. In addition, we provide numerical experiments on Army Polarimetric Thermal Face images [14]) to justify network pruning in improving test accuracy (see Figure 2 for an example). The dataset used in this project is adapted from the polarimetric thermal face database in [14]. The objective of the experiment is to determine whether the visible image represents the same person as the thermal images. Figure 1 shows two input images, and each input consists of two aligned thermal images from the same person and one visible image. The label is 0 if the thermal and visible images are from the same person, or 1 if otherwise. The training data contains 240 images based on 60 individuals with no expression. For each individual, we construct four images following Figure 1. Two images belong to class 0, and the other two belong to class 1, where the visible image is randomly selected from the other 59 individuals. The test data contains 240 images from the same group of individuals but with different expressions. The iterative magnitude-based pruning algorithm is applied via training on a 7-layer neural network. Figure 2 illustrates that the test accuracy increases from 82.16% to 85.88% with pruning, which verifies the improved performance when training on the "winning ticket". In addition, the pruned network with only 18.7% parameters remaining achieves better performance than the original dense network, and the computational complexity is significantly reduced when training on a "winning ticket."

2. Self-training to improve learning performance using unlabeled data with limited labeled samples [P5].

As one of the most widely used semi-supervised learning algorithms, the self-training approach utilizes unlabeled data to achieve improved test accuracy when the training process is only accessible to a limited number of labeled data. In practice, obtaining labeled data are costly and inconvenient. For example, labeling the ImageNet [15] took almost four years with around 49K workers from 167 countries. While we are short of data collection budget and time in most tasks, the quality of training data can hardly be guaranteed. Moreover, some data labels may contain sensitive personal information, e.g., personal accounts on smartphones and patient information in hospitals, that cannot be shared.

In contrast, unlabeled data are vastly available, and self-training has shown empirical success in diversified applications such as few-shot image classification [16–20], objective detection [21], robustness-aware model training against adversarial attacks [22], continual lifelong learning [23], and natural language processing [24, 25]. Despite the numerical success of self-training approaches, theoretical understanding of unlabeled data in improving the test accuracy is limited to impractical scenarios, e.g., linear models [26–29] or an infinite number of unlabeled data [30].

Therefore, this project aimed to bridge the gap between numerical success with non-linear neural networks and a theoretical understanding of self-training in improving test accuracy. Specifically, this project is centered on the commonly-used iterative self-training method. An initial teacher model (learned from the labeled data) is applied to the unlabeled data to generate pseudo labels. One then trains a student model by minimizing the weighted empirical risk of both the labeled and unlabeled data. The student model is then used as the new teacher to update the pseudo labels of the unlabeled data. This process is repeated multiple times to improve the eventual student model. This project addressed the following questions: how and when self-training improves test accuracy and how the number of unlabeled data affects the performance.

Specifically, we provided the analytical justification of the iterative self-training algorithm using unlabeled data over a supervised learning approach. First, we provided the quantitative justification of improved test accuracy and convergence rate by the number of unlabeled data used in training. Second, we provided insights into the parameter selection in balancing the improved performance and convergence. Third, we provided numerical justifications for the efficiency of self-training algorithms in object detection.

The objective of the experiment is to identify relevant changes in bi-temporal remote sensing images, namely, change detection. The dataset used in this project is from WHU dataset [31], which contains a satellite imagery data set of building samples extracted from aerial images. The training and test sets split follows the default settings in [31]. When implementing the self-training approach, we randomly pick a certain percentage of data in the training set as labeled data, and the remaining data in the training set are used as unlabeled data by discarding the labels. The supervised approach uses the same labeled data set as the self-training approach, which serves as the baseline. Figure 3 shows the intersection over union (IoU) value using the self-training algorithm (using labeled data and unlabeled data) is significantly improved compared with the supervised learning approach (labeled data only), especially when the number of labeled data is small. For example, with only 10% labeled data, the improvement of the IoU value is more than 44% (from 0.2683 to 0.3871) via the self-training algorithm. With 20% labeled data, the IoU value via the self-training algorithm is over 0.4, which is close to a good IoU value thresholding that is usually set as 0.5 [32].

Figure 4 shows an example of both self-training and supervised learning algorithms in detecting buildings.

The white region in the left sub-figure denotes the ground truth of the buildings, and the red area in the two right sub-figures indicates the detected buildings by self-training and supervised learning algorithms. We can see that with the help of unlabeled data, the self-training algorithm achieves higher IoU and better performance detecting buildings than the supervised approach.

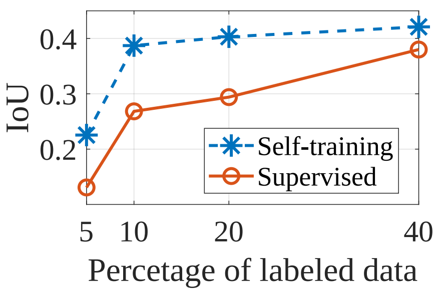

Figure 3: Intersection over union (IoU) using the self-training algorithm and supervised learning approach.

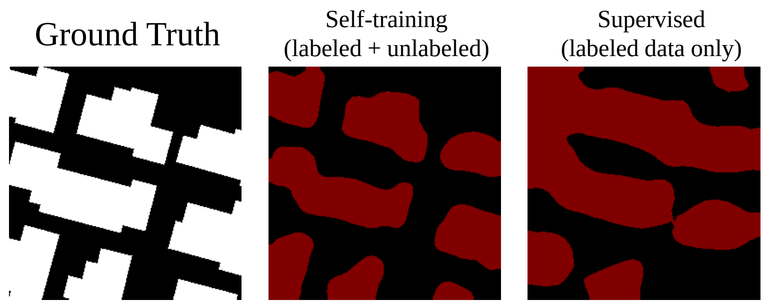

Figure 4: The improved performance of using unlabeled data. Self-training uses 40% of data with labels and 60% of data without labels. Supervised learning use only 40% of data with labels.

3. Graph topology sampling to reduce the training complexity [P3].

Graph convolutional networks (GCNs) can accurately model graph-structured data in many empirical tasks, including text analysis [33–35], recommendation systems [36, 37], and remote sensing [38, 39]. Such great success usually requires high computational and memory costs because the representation of one node is a recursive aggregation of its neighbors. This scalability issue motivates the development of graph topology sampling methods, which randomly select a subset of neighbors in each iteration when training GCNs to reduce the training complexity.

Existing sampling strategies can be categorized into three classes. See Figure 5 as an illustration. Node-wise neighbor-sampling methods such as GraphSAGE [33] and VRGCN [40] keep all the nodes and sample a subset of neighbors for each node. Layer-wise importance sampling methods such as FastGCN [7] and LADIES [41] sample a fixed number of nodes in each layer according to the estimation of the importance of nodes. Sub-graph sampling tools such as Cluster-GCN [42] and GraphSaint [43] extract appropriate multiple sub-graphs to represent the full graph without information loss during the training. These methods can significantly improve training time complexity compared with original GCNs. For instance, the training speed of FastGCN [7] is 2 times, 17 times, and 168 times that of a vanilla GCN on Cora, Pubmed, and Reddit dataset, respectively, showing a more remarkable performance on complexity reduction for larger graphs. Surprisingly, these sampling techniques can give comparable or even better empirical results than training with the entire graph. For example, GraphSAINT [43] shows new state-of-the-art F1 scores for PPI (0.995) and Reddit (0.970).

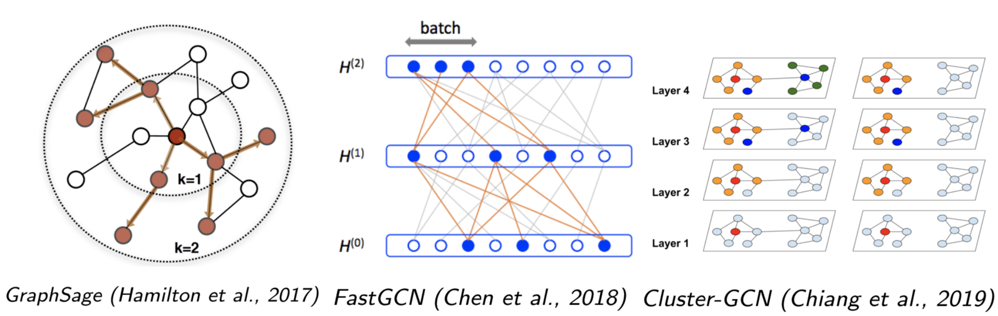

Figure 5: Different graph sampling methods. (a) GraphSage: Sampling a fixed number of neighbors for each node. (b) FastGCN: Sampling nodes in each layer by importance. (c) Cluster-GCN: Sampling sub-graphs using a graph clustering algorithm.

In contrast to the empirical success, the theoretical foundation of training GCNs with graph topology sampling is much less explored. Recent theoretical works either focus on GCNs trained without sampling [44–46] or only study the convergence rate of graph sampling [45] without generalization analysis. Therefore, this project establishes the theoretical basis to study what conditions are required for a three-layer GCN learned with graph topology sampling (as shown in Figure 6) to achieve satisfactory generalization. We apply a group-wise uniform sampling based on the degree of node groups. To be more specific, higher-degree nodes are sampled more during the training with importance scaling on the normalized adjacency matrix for aggregation. This sampling strategy is consistent with the intuition of the importance sampling of FastGCN [7]. We explicitly characterize the sample complexity and model complexity for the generalization. Our results also show that training with sampling methods has the same generalization performance as learning an effective adjacency matrix which is different from the entire graph.

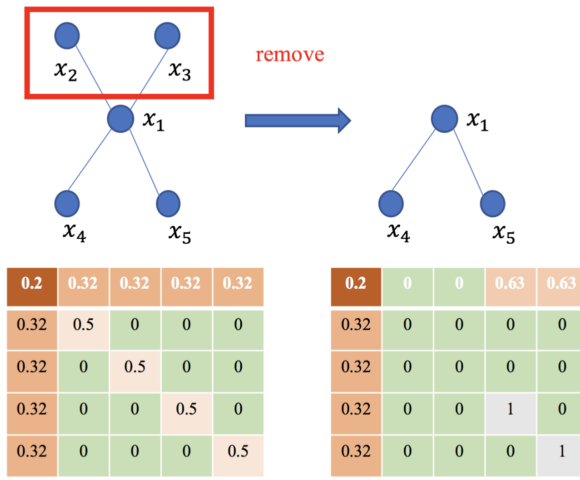

Figure 6: The studied graph topology sampling algorithm in this project. We set a higher sampling rate on higher-degree nodes $\boldsymbol{x}_1$, and a lower sampling rate on lower-degree nodes $\boldsymbol{x}_2$, $\boldsymbol{x}_3$, $\boldsymbol{x}_4$, and $\boldsymbol{x}_5$. For sampled nodes $\boldsymbol{x}_1$, $\boldsymbol{x}_4$, and $\boldsymbol{x}_5$, we scale the corresponding columns of the normalized adjacency matrix $\boldsymbol{A}$. For unsampled nodes $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$, we set the corresponding columns of $\boldsymbol{A}$ to 0.

Publications

We have the following published paper with full citations below.

[P1] Ming Yi, Meng Wang, Evangelos Farantatos and Tapas Barik, "Bayesian Robust Hankel Matrix Completion with Uncertainty Modeling for Synchrophasor Data Recovery," *ACM SIGENERGY Energy Informatics Review*, 2022.

[P2] Ming Yi and Meng Wang, "Bayesian Energy Disaggregation at Substations with Uncertainty Modeling," *IEEE Transactions on Power Systems*, 2022, 37(1): 764-775.

[P3] Hongkang Li, Meng Wang, Sijia Liu, Pin-Yu Chen and Jinjun Xiong, "Generalization Guarantee of Training Graph Convolutional Networks with Graph Topology Sampling," in *Proc. of 2022 International Conference on Machine Learning (ICML)*, July 2022. (acceptance rate: 21.9

[P4] Ming Yi and Meng Wang, " Recent Results of Energy Disaggregation with Behind-the-Meter Solar Generation," in *Proc. of the 11th Bulk Power Systems Dynamics and Control Symposium – IREP'2022*, July 2022.

[P5] Shuai Zhang, Meng Wang, Sijia Liu, Pin-Yu Chen and Jinjun Xiong, "How Does Unlabeled Data Improve Generalization in Self-training? A one-hidden-layer Theoretical Analysis," in *Proc. the Tenth International Conference on Learning Representations (ICLR)*, April 2022. (acceptance rate: 32.3%)

[P6] Shuai Zhang, Meng Wang, Sijia Liu, Pin-Yu Chen and Jinjun Xiong, "Why Lottery Ticket Wins? A Theoretical Perspective of Sample Complexity on Sparse Neural Networks," in *Proc. of the Thirty-fifth Conference on Neural Information Processing Systems (NeurIPS)*, December 2021. (acceptance rate: 26%)

[P7] Hongkang Li, Shuai Zhang, and Meng Wang, "Learning and generalization of one-hidden-layer neural networks, going beyond standard Gaussian data," in *Proc. 2022 56th Annual Conference on Information Sciences and Systems (CISS)*, Princeton, NJ, USA, 2022.

References

[1] S. Han, J. Pool, J. Tran, and W. Dally, "Learning both weights and connections for efficient neural network," in *Advances in neural information processing systems*, 2015, pp. 1135–1143.

[2] T.-J. Yang, Y.-H. Chen, and V. Sze, "Designing energy-efficient convolutional neural networks using energy-aware pruning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5687–5695.

[3] J. Frankle and M. Carbin, "The lottery ticket hypothesis: Finding sparse, trainable neural networks," in *International Conference on Learning Representations*, 2019. [Online]. Available: https://openreview.net/forum?id=rJl-b3RcF7

[4] T. Chen, Y. Sui, X. Chen, A. Zhang, and Z. Wang, "A unified lottery ticket hypothesis for graph neural networks," in *International Conference on Machine Learning*. PMLR, 2021, pp. 1695–1706.

[5] T. Chen, J. Frankle, S. Chang, S. Liu, Y. Zhang, Z. Wang, and M. Carbin, "The lottery ticket hypothesis for pre-trained bert networks," *arXiv preprint arXiv:2007.12223*, 2020.

[6] T. Chen, J. Frankle, S. Chang, S. Liu, Y. Zhang, M. Carbin, and Z. Wang, "The lottery tickets hypothesis for supervised and self-supervised pre-training in computer vision models," *arXiv preprint arXiv:2012.06908*, 2020.

[7] J. Chen, T. Ma, and C. Xiao, "Fastgcn: Fast learning with graph convolutional networks via importance sampling," in *International Conference on Learning Representations*, 2018.

[8] J. Frankle, G. K. Dziugaite, D. M. Roy, and M. Carbin, "Stabilizing the lottery ticket hypothesis," *arXiv preprint arXiv:1903.01611*, 2019.

[9] A. Renda, J. Frankle, and M. Carbin, "Comparing rewinding and fine-tuning in neural network pruning," in *International Conference on Learning Representations*, 2019.

[10] L. Liebenwein, C. Baykal, H. Lang, D. Feldman, and D. Rus, "Provable filter pruning for efficient neural networks," in *International Conference on Learning Representations*, 2019.

[11] M. Ben, M. Osadchy, V. Braverman, S. Zhou, and D. Feldman, "Data-independent neural pruning via coresets," in *International Conference on Learning Representations (ICLR)*, 2020.

[12] M. Ye, C. Gong, L. Nie, D. Zhou, A. Klivans, and Q. Liu, "Good subnetworks provably exist: Pruning via greedy forward selection," in *International Conference on Machine Learning*. PMLR, 2020, pp. 10 820–10 830.

[13] S. Zhang, M. Wang, S. Liu, P.-Y. Chen, and J. Xiong, "Why lottery ticket wins? a theoretical perspective of sample complexity on pruned neural networks," in *Thirty-fifth Conference on Neural Information Processing Systems (NeurIPS)*, 2021.

[14] S. Hu, N. J. Short, B. S. Riggan, C. Gordon, K. P. Gurton, M. Thielke, P. Gurram, and A. L. Chan, "A polarimetric thermal database for face recognition research," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2016, pp. 119–126.

[15] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[16] J.-C. Su, S. Maji, and B. Hariharan, "When does self-supervision improve few-shot learning?" in *European Conference on Computer Vision*. Springer, 2020, pp. 645–666.

[17] Q. Xie, M.-T. Luong, E. Hovy, and Q. V. Le, "Self-training with noisy student improves imagenet classification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10 687–10 698.

[18] T. Chen, S. Kornblith, K. Swersky, M. Norouzi, and G. E. Hinton, "Big self-supervised models are strong semi-supervised learners," *Advances in Neural Information Processing Systems*, vol. 33, pp. 22 243–22 255, 2020.

[19] I. Z. Yalniz, H. Jégou, K. Chen, M. Paluri, and D. Mahajan, "Billion-scale semi-supervised learning for image classification," *arXiv preprint arXiv:1905.00546*, 2019.

[20] B. Zoph, G. Ghiasi, T.-Y. Lin, Y. Cui, H. Liu, E. D. Cubuk, and Q. Le, "Rethinking pre-training and self-training," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[21] C. Rosenberg, M. Hebert, and H. Schneiderman, "Semi-supervised self-training of object detection models," in *Proceedings of the Seventh IEEE Workshops on Application of Computer Vision (WACV/MOTION'05)-Volume 1-Volume 01*, 2005, pp. 29–36.

[22] Y. Carmon, A. Raghunathan, L. Schmidt, J. C. Duchi, and P. S. Liang, "Unlabeled data improves adversarial robustness," *Advances in Neural Information Processing Systems*, vol. 32, pp. 11 192–11 203, 2019.

[23] K. Lee, K. Lee, J. Shin, and H. Lee, "Overcoming catastrophic forgetting with unlabeled data in the wild," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 312–321.

[24] J. He, J. Gu, J. Shen, and M. Ranzato, "Revisiting self-training for neural sequence generation," in *International Conference on Learning Representations*, 2019.

[25] J. Kahn, A. Lee, and A. Hannun, "Self-training for end-to-end speech recognition," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 7084–7088.

[26] Y. Chen, C. Wei, A. Kumar, and T. Ma, "Self-training avoids using spurious features under domain shift," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[27] A. Raghunathan, S. M. Xie, F. Yang, J. Duchi, and P. Liang, "Understanding and mitigating the tradeoff between robustness and accuracy," in *International Conference on Machine Learning*. PMLR, 2020, pp. 7909–7919.

[28] S. Oymak and T. C. Gulcu, "Statistical and algorithmic insights for semi-supervised learning with self-training," *arXiv preprint arXiv:2006.11006*, 2020.

[29] L. Oneto, D. Anguita, A. Ghio, and S. Ridella, "The impact of unlabeled patterns in rademacher complexity theory for kernel classifiers," *Advances in neural information processing systems*, vol. 24, pp. 585–593, 2011.

[30] C. Wei, K. Shen, Y. Chen, and T. Ma, "Theoretical analysis of self-training with deep networks on unlabeled data," in *International Conference on Learning Representations*, 2020.

[31] S. Ji, S. Wei, and M. Lu, "Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 1, pp. 574–586, 2018.

[32] T. Takahashi, K. Nozaki, T. Gonda, T. Mameno, M. Wada, and K. Ikebe, "Identification of dental implants using deep learning—pilot study," *International journal of implant dentistry*, vol. 6, no. 1, pp. 1–6, 2020.

[33] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Advances in neural information processing systems*, 2017, pp. 1024–1034.

[34] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. International Conference on Learning (ICLR)*, 2017.

[35] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *International Conference on Learning Representations (ICLR)*, 2018.

[36] R. Ying, R. He, K. Chen, P. Eksombatchai, W. L. Hamilton, and J. Leskovec, "Graph convolutional neural networks for web-scale recommender systems," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, pp. 974–983.

[37] R. Van den Berg, T. N. Kipf, and M. Welling, "Graph convolutional matrix completion," in *KDD*, 2018.

[38] Y. Ding, X. Zhao, Z. Zhang, W. Cai, and N. Yang, "Graph sample and aggregate-attention network for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2021.

[39] ——, "Multiscale graph sample and aggregate network with context-aware learning for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 4561–4572, 2021.

[40] J. Chen, J. Zhu, and L. Song, "Stochastic training of graph convolutional networks with variance reduction," in *International Conference on Machine Learning.* PMLR, 2018, pp. 942–950.

[41] D. Zou, Z. Hu, Y. Wang, S. Jiang, Y. Sun, and Q. Gu, "Layer-dependent importance sampling for training deep and large graph convolutional networks," *Advances in Neural Information Processing Systems*, vol. 32, pp. 11 249–11 259, 2019.

[42] W.-L. Chiang, X. Liu, S. Si, Y. Li, S. Bengio, and C.-J. Hsieh, "Cluster-gcn: An efficient algorithm for training deep and large graph convolutional networks," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 257–266.

[43] H. Zeng, H. Zhou, A. Srivastava, R. Kannan, and V. Prasanna, "Graphsaint: Graph sampling based inductive learning method," in *International Conference on Learning Representations*, 2019.

[44] S. Verma and Z.-L. Zhang, "Stability and generalization of graph convolutional neural networks," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 1539–1548.

[45] W. Cong, M. Ramezani, and M. Mahdavi, "On provable benefits of depth in training graph convolutional networks," *Advances in Neural Information Processing Systems*, vol. 34, 2021.

[46] X. Zhou and H. Wang, "The generalization error of graph convolutional networks may enlarge with more layers," *Neurocomputing*, vol. 424, pp. 97–106, 2021.