Towards Learning with Limited Supervision: Efficient Few-shot and Semi-supervised Classification for Vision Tasks

Submitted in partial fulfillment of the requirements for

the degree of

Doctor of Philosophy

in

The Electrical and Computer Engineering Department

Ran Tao

B.S., Biomedical Engineering, Xi'an Jiaotong University
M.S., Biomedical Engineering, Carnegie Mellon University
M.S., Machine Learning, Carnegie Mellon University

Carnegie Mellon University
Pittsburgh, PA

Dec 2023

Acknowledgements

Pursuing my Ph.D. presented me with numerous hurdles, but the experience has proven to be exceptionally gratifying, and I have never second-guessed my decision to embark on this journey. The rigorous scientific training I received at Carnegie Mellon University positions me as a researcher eager to transfer my curiosity and passion for innovation into real-world applications and cutting-edge research topics. Within this academic environment, I've encountered dedicated and talented individuals who share a belief in their capacity to push the boundaries for the betterment of scientific progress. Motivated by the exceptional faculty and my industrious and gifted peers, I continually strive to learn, evolve, and contribute to the areas I am deeply committed to. This is a journey full of gratitude and appreciation.

I want to thank Professor Marios Savvides, my advisor, for consistently offering unwavering support, guidance, and motivation throughout my Ph.D. journey. Our paths first crossed in the Pattern Recognition course, where I became captivated by the facial detection and recognition techniques he introduced. Intrigued by the elegant algorithms used in these tasks, I joined Professor Savvides's lab as a research assistant during a summer internship. Over the subsequent years, Professor Savvides has been instrumental in encouraging me to embrace academic challenges, delve into complex topics, and pursue the goals I set for myself. I transformed into a researcher, contributing to numerous publications, and all of these are hard to achieve without Professor Savvides's support and trust.

I extend my sincere appreciation to my doctoral committee members, namely Prof. John Dolan, Dr. Saad Bedros, and Prof. Raied Aljadaany. Their invaluable insights and advice greatly enriched my thesis, and I am grateful for the flexibility they demonstrated in coordinating the timing of both the proposal and defense. Prof. John Dolan consistently exhibits exceptional consideration for students' needs, and his kindness and responsiveness have been a significant asset for my proposal and defense. Dr. Saad Bedros led the insightful discussions on extending my research on medical applications, which inspired me to dive deep into the imbalanced prediction issue for semi-supervised learning in specific domains. Having Prof. Aljadaany on my doctoral committee is like a fulfillment for my PhD. journey. When we were colleagues in the lab, he kindly offered a lot of support to help me with courses and preparing for the qualification exam. Prof. Aljadaany's academic success also serves as a guiding inspiration and a role model for me to emulate.

The warm, friendly and supportive environment in our lab, the CyLab Biometrics Center, has become a cherished memory that will remain with me for years to come. I received immense support and assistance from my fellow colleagues, and it brings me great joy to have had the opportunity to know and learn from such a talented group of individuals. I especially express my gratitude to: Khoa for guiding me through

my initial project and offering valuable advice on my professional development, Chi Nhan Duong and Kha Gia Quach for their help on my first project, Chandrasekhar Bhagavatula for the ice cream celebration of my summer intern, Chenchen Zhu and Yandong Wen for walking home together and chatting after a long day in lab, Yutong Zheng for inspiring me with his insights on facial recognition and many interesting conversations about birds, Dipan Pal for his guidance in writing academic papers, Hai for many small talks when we two workaholics met on Sunday in the lab, Uzair for always providing joy and cheers in the lab and also big credits to his boy's sweet smiles, Sreena Nallamothu for her dedication in coordinating the projects and making the lab space a lovely place during the holiday seasons, Magesh Kanna for helping me set up the poster session, Fangyi Chen and Yukai Huang for always bringing gifts and snacks after they went back home, Snow Zhang for helping with the patent applications and providing advice on my personal development, Han Zhang for being a fantastic guide for me on a new domain of knowledge and Hao Chen for numerous insightful discussions on research and those dedicated days working for papers. I would also like to thank Jonathan Dencker, a distinguished researcher from AI2C, for his help throughout the funding period.

In addition to my lab colleagues, I had the pleasure of collaborating with exceptional researchers who offered invaluable assistance. I want to express my deep appreciation for Guosheng Hu. As a distinguished researcher and team leader at Oosto, Inc., Guosheng consistently provided generous guidance to me in the field of computer vision. His advice on my publications and tutoring in developing academic ideas and drafting papers were instrumental to my progress. Jindong Wang, a Senior Researcher at Microsoft Research Asia, stands out as a great mentor who fosters collaboration among Ph.D. students from diverse institutions. Under his guidance, I had the opportunity to collaborate with a remarkable group of peers on the subject of semi-supervised learning. I am truly grateful for his pivotal role in enhancing my proficiency in machine learning.

The staff at CyLab and CMU have consistently provided invaluable assistance in handling numerous requests. Specifically, I appreciated the assistance and support from my academic advisors: Nathan Snizaski, Dorothy Holland-Minkley, and Greta Ruperto. Their professionalism and wholehearted assistance have been immensely beneficial, and it is truly a pleasure to collaborate with them.

My friends have been my pillar of strength. I am deeply thankful to them for providing the support and love that illuminated my life in Pittsburgh, helping me navigate through the challenging and isolating period of COVID. Special thanks to Boyuan Zhang, Mengfei Lv, Xin Gao, and Xinyu Li for the memorable girls' nights we spent together during holidays, New Year celebrations, and birthdays. These are precious memories that will stay with me for a lifetime. I express gratitude to Lulu Shao, Zhenyu Li, and their lovely kids for being my family in Pittsburgh. Appreciation also goes to Shaolei Shen, Chen Chen, Yuxuan

Zhang, and Zhiheng Ma for their unwavering support and care over the years. I extend my thanks to Tianyi Ouyang and Hongliang Luo for the unforgettable trips, traveling all the way from China to visit me in New York and Washington; you are truly among the best gifts life has given me. I also want to acknowledge Yang Zou, Zechun Liu, Yang Gao, Wenwen Wang, Yan Xu, Chaoyang Wang, Wenbo Zhao, and Yafei Hu for the enjoyable gatherings and game nights. Your presence has added so much joy to my life.

I extend my heartfelt gratitude to my family for their unwavering love and support. A special thanks to my parents for embracing me for who I am, being my best friends, providing everything they could that allowed me to explore life's possibilities, and placing full trust in every decision I've made. They are the source of my courage, honesty, and happiness. I want to express my appreciation to my partner, Jiati Le, for returning to Pittsburgh for me, enduring the solitude of not working alongside the team for the past two years, and for your love and support, the most cherished privileges life has given me.

As I anticipate the inevitable highs and lows in my future, the resilience, bravery, and problem-solving skills cultivated during my Ph.D. journey are the most precious assets I possess to confront the uncertainties ahead.

This work has been partly funded by BossaNova Inc. with grant number A022256, and U.S. Army Artificial Intelligence Integration Center (AI2C), with grant numbers W911NF20D0002, AWD00000001, and A025452.

Abstract

Vision classification tasks, a fundamental and transformative aspect of deep learning and computer vision, play a pivotal role in our ability to understand the visual world. Deep learning techniques have revolutionized the field, enabling unprecedented accuracy and efficiency in vision classification. However, deep learning models, especially supervised models, require large amounts of labeled data to learn effectively. The acquisition of large-scale datasets meets many difficulties considering the dynamics in real-world applications. Collecting and annotating data is a time-consuming and expensive process, which sometimes requires domain-specific expertise to provide a sufficient quantity of high-quality labeled data. Meanwhile, privacy and ethical concerns hinder data acquisition in certain domains, such as healthcare or finance. Learning with limited supervision addresses these challenges by developing techniques that allow models to learn and make predictions with only a partial or a small number of supervision.

In this presentation, we will introduce our research, which encompasses several advancements within the domain of learning with limited supervision. Initially, we introduce a novel fine-tuning method tailored to enhance the efficiency of few-shot learning, particularly in cross-domain scenarios. Building upon this, we extend our comprehension of few-shot fine-tuning into the transductive setting. Here, we present innovative weighting techniques to harness the potential of unlabeled data during the testing phase. In addition, we confront the intricate balance between data quality and quantity when leveraging unlabeled training data in semi-supervised learning. To address this challenge, we introduce the SoftMatch method, which allows for the adaptive integration of unlabeled data during training.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Learning with Limited Supervision: Difficulties and Challenges

Recent advancements in deep learning have propelled the field to new heights, reshaping the landscape of artificial intelligence and machine learning. With the advancement in large-scale data acquisition, unprecedented computational power, and innovative neural network architectures, deep learning has achieved remarkable breakthroughs in various domains, from natural language processing to computer vision. Vision classification tasks, a fundamental and transformative aspect of deep learning and computer vision, play a pivotal role in our ability to understand the visual world. The advent of deep learning techniques has revolutionized the field, enabling unprecedented accuracy and efficiency in vision classification. In particular, Vision classification tasks involve the process of training and deploying machine learning models to recognize and assign labels to objects, scenes, or patterns within visual data. The goal is to teach algorithms to differentiate between various categories or classes, allowing them to make decisions based on what they *see*. However, deep learning models, especially supervised models, require large amounts of labeled data to learn effectively.

The acquisition of large-scale datasets meets many difficulties considering the dynamics in real-world applications. Firstly, collecting and annotating data is a time-consuming process, which is expensive and sometimes requires domain-specific expertise to provide a sufficient quantity of high-quality labeled data. Also, the real-world data is diverse and can vary significantly under different scenarios. For instance, collecting data from remote or hazardous environments or historical data that no longer exists can be problematic. And as data distribution across different classes is naturally varied, data collection for those rare or unexpected events can be impractical. Meanwhile, in certain domains, such as healthcare or finance, data acquisition is hindered by privacy and ethical concerns. Sharing and using sensitive

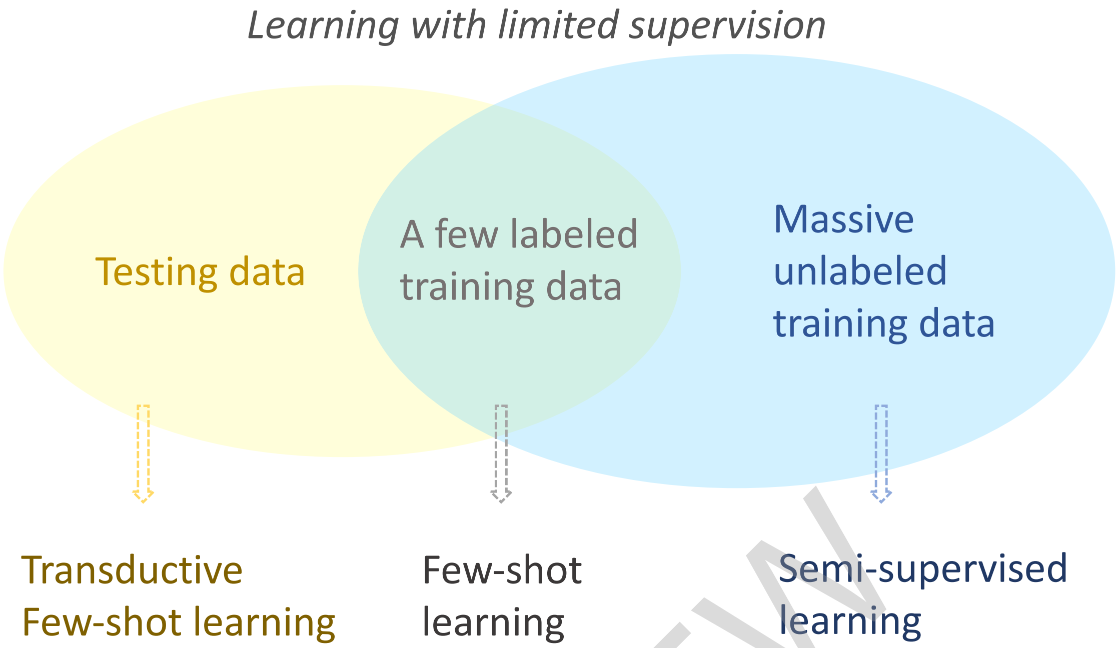
Learning with limited supervision
Testing data
A few labeled training data
Massive unlabeled training data
Transductive Few-shot learning
Few-shot learning
Semi-supervised learning

Figure 1.1: Illustration on research directions under Learning with Limited Supervision. With only a few labeled training data, learning with limited supervision is developed in different directions. Few-shot learning refers to only utilizing a few labeled training data for model learning. While considering the dynamics in the testing environment, combining the testing data and adapting models with a few labeled training data falls into transductive few-shot learning. Semi-supervised learning refers to the research where the data collection of massive unlabeled training data is available.

or personal data can be legally and morally complex. Last but not least, in applications like real-time object detection or autonomous navigation, the data is constantly changing. Continuously acquiring and labeling data to keep the model updated is a challenge.

Learning with limited supervision addresses these challenges by developing techniques that allow models to learn and make predictions with only a partial or a small number of supervision. As a compelling and challenging frontier in machine learning and artificial intelligence, learning with limited supervision can be broadly split into two areas: few-shot learning and semi-supervised learning. Few-shot learning, where the training data is limited to a few samples for each class, addresses the problem of generalization in scenarios where data is scarce or expensive to obtain. The most obvious challenge in few-shot learning is the scarcity of labeled examples. With only a handful of samples for each class, there are high demands to improve the generalization and avoid overfitting for model learning. While models leverage knowledge for related tasks, determining how to transfer this knowledge effectively and adapt it to the current few-shot tasks is particularly difficult which can be particularly difficult when dealing with complex patterns or concepts. Because of the limited data, few-shot models are prone to overfitting, where they perform well on the few training examples but struggle to generalize to new, unseen data. Despite these challenges, few-shot learning has gained attention and interest because of its potential applications in fields where acquiring large labeled datasets is impractical. Meanwhile, considering the dynamics in the real testing scenarios, conducting test-time adaptation of the models in deployment draws attention. Transductive few-shot learning, which combines a few labeled training data together with the testing data to train the models, improves the quantity of data acquisition and enhances the robustness of predictions for different testing scenarios. The other area of learning with limited supervision is known as semi-supervised learning, where a model is trained using a combination of labeled data (where input examples are paired with correct outputs) and unlabeled data (input examples without corresponding outputs). This approach leverages the advantages of supervised learning, where labeled data helps the model learn while also capitalizing on the abundance of unlabeled data, which is typically easier and cheaper to obtain. The goal of semi-supervised learning is to build accurate and robust models while mitigating the need for extensive labeled datasets. Specifically, for semi-supervised learning, effectively leveraging both labeled and unlabeled data is crucial to achieve meaningful generalization. Many challenges are under exploration in semi-supervised learning, such as balancing the utilization of labeled and unlabeled training data, identifying the most informative examples from the unlabeled data, improving the consistency of the label predictions for unlabeled training data, and ensuring the privacy and ethical use while dealing with potentially sensitive data. Despite these challenges, semi-supervised learning has shown great promise in various applications. Consistent efforts are made to improve the reliability and efficiency of semi-supervised learning models and to expand their applicability to real-world problems.

In the following sections, we will separately introduce the developments and related works in few-shot learning and semi-supervised learning. Further, we will discuss our contributions towards learning with limited supervision and introduce the contents of each section in the dissertation.

1.2 A Recent History of Few-Shot Learning

Few-shot learning has been quite an active research field in recent years. Methods in few-shot learning can be broadly split into two directions: meta-learning inspired methods and fine-tuning related methods. From the initial discussion of the few-shot learning topic, meta-learning, known as *learning to learn*, is correspondingly proposed to solve the special learning scenarios in which only a few training samples are available. Intuitively, meta-learning is designed to optimize the learning hypothesis based on its performance on testing. Thus, a general two-loop training diagram is proposed where the inner loop is optimized on the training data and the outer loop is initialized with the optimization of the inner loop, which is further optimized using the *testing* data. To satisfy this particular training diagram, the training data is designed in an episode way. From the data pool of training data, an episode is sampled with a meta-training set and meta-testing set to separately involve in the inner and outer loop of optimization on the model parameters. The branch of Meta-learning methods Finn et al. [2017], Ravi and Larochelle [2016], Rusu et al. [2018], Vinyals et al. [2016], Snell et al. [2017], Sung et al. [2018], Chen et al. [2020], Simon et al. [2018, 2020] on few-shot learning is designed to directly back-propagate the loss of the test set while the hypothesis for classification is proposed with the training set. There are optimization based meta-learning like MAML Finn et al. [2017], Ravi and Larochelle [2016], Rusu et al. [2018], metric-based meta-learning like Vinyals et al. [2016], Snell et al. [2017], Sung et al. [2018], Chen et al. [2020] and methods working on representations on sub-space like Simon et al. [2018, 2020]. Besides improving the meta-learning training diagram in few-shot learning, methods are also exploring the following topics: data argumentation with hallucinating more samplesHariharan and Girshick [2017], Wang et al. [2018], optimization with ridge regression or support vector machine Bertinetto et al. [2018], Lee et al. [2019], using graph neural networks Garcia and Bruna [2017], Kim et al. [2019], self/semi-supervised learning Ren et al. [2018], Gidaris et al. [2019], Li et al. [2019a], Wang et al. [2020], learning with semantic information Li et al. [2020], class weight generalization Gidaris and Komodakis [2018, 2019], Guo and Cheung [2020], modules working on attentive spatial features Li et al. [2019b], Hou et al. [2019], Doersch et al. [2020], knowledge distillation Tian et al. [2020]. The evaluation for few-shot learning is also an important development. At the beginning of few-shot learning, the datasets are limited to Omniglot and a small subset of Imagenet data such as mini-Imagenet. The recent development of Meta-DatasetTriantafillou et al. [2019] proposes a more realistic evaluation for few-shot learning where algorithms are evaluated over ten datasets from different domains with a large-scale meta-training set spanned from ImageNetKrizhevsky et al. [2012]. The 10 datasets covered in Meta-Dataset are namely ILSVRC-2012 Russakovsky et al. [2015a], OmniglotLake et al. [2015], AircraftMaji et al. [2013], CUB-200-2011Wah et al. [2011], Describable Textures Cimpoi et al. [2014], Quick Draw, FungiSchroeder and Cui [2018], VGG FlowerNilsback and Zisserman [2008], Traffic SignHouben et al. [2013] and MSCOCOLin et al. [2014]. The evaluation of meta-Dataset not only requires algorithms to obtain good performance on few-shot learning but also sets higher demands on generalization over different domains. It provides a more realistic comparison and inspires algorithms to apply to real-world applications.

The other direction in few-shot learning is fine-tuning related methods. Using the support set during testing for fine-tuning purposes is not a new idea, which originates from Qi et al. [2018] and is further discussed in Chen et al. [2019]. In Qi et al. [2018], using the average feature from the support set as the class prototype when computing the distance-based loss for fine-tuning the whole backbone is proposed. The utilization of direct training a large-scale classification on meta-training set has drawn more attention recently Triantafillou et al. [2019], Chen et al. [2020], Tian et al. [2020]. In doing so, a feature extractor is firstly trained under supervised learning, and in other words, the feature space is constructed by globally optimized distance metrics. Thus, this pre-trained feature extractor owns better generalization property when applying to novel classes Chen et al. [2020]. Tian et al. [2020] proposes to train a simple linear classifier upon the pre-trained feature extractor using the support set during testing. Besides, Dhillon et al. [2019] also proposes a transductive finetuning method that takes account of the query set in any unsupervised way. Transductive few-shot learning uses the unlabeled query set (testing images) and the support set (training images) to compensate for the lack of training data. Nichol et al. [2018] updates parameters of batch normalization layers using unlabelled query samples. Liu et al. [2018] propagates labels for unseen classes through episodic meta-learning and Bateni et al. [2022] presents the label refinement with a Mahalanobis-distance based classifier. TIM Boudiaf et al. [2020] designs a loss to encourage the marginal distribution of the query set to be uniform, and pseudo-labels are directly used without compressing the possibly wrong predictions. $\alpha-$TIMVeilleux et al. [2021] addresses creating different testing distributions to reflect real-world scenarios better and proposes to enhance TIM Boudiaf et al. [2020] by $\alpha-$convergence. Hu et al. [2021a] uses the Optimal Transport Algorithm (OTA) for pseudo-label mapping with entropy minimization on the OTA-based mapping. Lichtenstein et al. [2020] computes a linear projection space on features for each task when utilizing the query set, which focuses on different directions with TF-MP. Boudiaf et al. [2020], Hu et al. [2021a] enforce the testing distribution to be uniform and don't propose compressing the utilization of possibly wrong predictions. In Dhillon et al. [2019], a transductive framework is firstly proposed to involve the testing images during fine-tuning. Dhillon et al. [2019] builds the classification upon predicted logits other than directly on features. Previous works on transductive few-shot learning ignore compressing the utilization of wrong predictions.

1.3 A Recent History of Semi-Supervised Learning

Effectively utilizing unlabeled training data is the major concern discussed in the works for semi-supervised learning. Creating a *pseudo-label* based on the model for each unlabeled data is an effective way of utilizing supervised loss functions in semi-supervised learning. And pseudo-labeling [Lee et al., 2013] generates artificial labels for unlabeled data and trains the model in a self-training manner. Also, to improve the effective utilization of unlabeled data, the consistency of model predictions on the unlabeled data is another concern popularly discussed in different works. Consistency regularization [Samuli and Timo, 2017] is proposed to achieve the goal of producing consistent predictions for similar data points. Various works focus on improving the pseudo-labeling and consistency regularization from different aspects. Methods

[Samuli and Timo, 2017, Tarvainen and Valpola, 2017, Iscen et al., 2019, Ren et al., 2020] improving the loss weighting are exploring algorithms that can better balance between the utilization of labeled and unlabeled data during training. And data augmentation [Grandvalet et al., 2005a, Sajjadi et al., 2016, Miyato et al., 2018, Berthelot et al., 2019b,a, Xie et al., 2020, Cubuk et al., 2020, Sajjadi et al., 2016] is explored to improve the consistency by assuming that data augmentation would encourage diverse learning for models to avoid falling into a trivial solution while the predictions of models are directly used for unlabeled data. Methods focusing on improving feature consistency [Li et al., 2021b, Zheng et al., 2022, Fan et al., 2021] provides another viewpoint of consistency regularization in Semi-supervised Learning. A loss weight ramp-up strategy is proposed to balance the learning on labeled and unlabeled data. [Samuli and Timo, 2017, Tarvainen and Valpola, 2017, Berthelot et al., 2019b,a]. By progressively increasing the loss weight for the unlabeled data, which prevents the model from involving too much ambiguous unlabeled data at the early stage of training, the model, therefore, learns in a curriculum fashion. However, loss weight ramp-up strategy always assigns the same weight to all unlabeled samples, where the erroneous predictions result in degraded performance. Per-sample loss weight is utilized to better exploit the unlabeled data [Iscen et al., 2019, Ren et al., 2020]. Label propagation [Iscen et al., 2019] proposes to use the entropy of predicted probability to calculate the per-sample loss weight. The previous work "Influence" shares a similar goal with us, which aims to calculate the loss weight for each sample but for the motivation that not all unlabeled data are equal [Ren et al., 2020]. SAW [Lai et al., 2022] utilizes effective weights [Cui et al., 2019] to overcome the class-imbalanced issues in SSL. Modeling of loss weight has also been explored in semi-supervised segmentation [Hu et al., 2021b]. De-biased self-training [Chen et al., 2022, Wang et al., 2022a] study the data bias and training bias brought by involving pseudo-labels into training. Kim et al. [2022] proposed to use a small network to predict the loss weight.

And recently using confidence thresholding to select a certain portion of unlabeled training data during each iteration [Sohn et al., 2020, Zhang et al., 2021, Xu et al., 2021a] is a proven effective method to reduce the influence of learning from wrong pseudo-labels. Specifically, the consistency regularization framework proposed in FixMatch [Sohn et al., 2020] is a popular framework in Semi-supervised Learning, considering its elegance and efficiency. While FixMatch [Sohn et al., 2020] serves as the baseline method in our development, we introduce the details of FixMatch [Sohn et al., 2020] in the following.

FixMatch. There are two major functionalities in FixMatch: consistency regularization by weak-strong data augmentation and confidence thresholding used in the loss for unsupervised data. As shown in Fig. 1.2, a weak and strong data augmentation pipeline is used in FixMatch [Sohn et al., 2020]. The same image goes into both the weak and strong data augmentation. The feature of a weakly augmented image is used to obtain the label predictions, which are further used to generate the one-hot pseudo-label

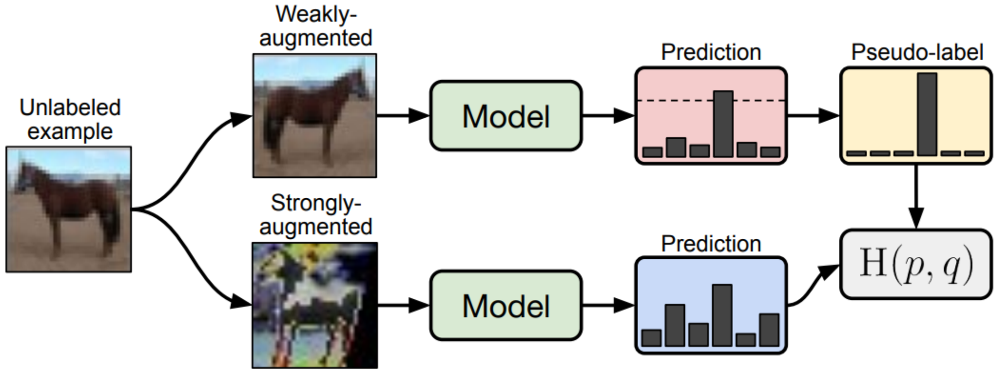

Figure 1.2: Figure and the following caption description is from [Sohn et al., 2020]. Diagram of FixMatch. A weakly-augmented image (top) is fed into the model to obtain predictions (red box). When the model assigns a probability to any class that is above a threshold (dotted line), the prediction is converted to a one-hot pseudo-label. Then, we compute the model's prediction for a strong augmentation of the same image (bottom). The model is trained to make its prediction on the strongly augmented version match the pseudo-label via a cross-entropy loss.

if the sample obtains an over-the-threshold confidence score. And the feature of a strongly augmented image is used to compute the model's prediction. To achieve consistency, the model is trained to make its prediction on the strongly augmented version match the pseudo-label via a cross-entropy loss.

We introduce the loss functions and the confidence threshold used in FixMatch [Sohn et al., 2020]. Denote the labeled and unlabeled datasets as $\mathcal{D}_L = \left\{\mathbf{x}_i^l, \mathbf{y}_i^l\right\}_{i=1}^{N_L}$ and $\mathcal{D}_U = \{\mathbf{x}_i^u\}_{i=1}^{N_U}$, respectively, where $\mathbf{x}_i^l, \mathbf{x}_i^u \in \mathbb{R}^d$ is the d-dimensional labeled and unlabeled training sample, and $\mathbf{y}_i^l$ is the one-hot ground-truth label for labeled data. We use N_L and N_U to represent the number of training samples in $\mathcal{D}_L$ and $\mathcal{D}_U$, respectively. Let $\mathbf{p}(\mathbf{y}|\mathbf{x}) \in \mathbb{R}^C$ denote the model's prediction. During training, given a batch of labeled data and unlabeled data, the model is optimized using a joint objective $\mathcal{L} = \mathcal{L}_s + \mathcal{L}_u$, where $\mathcal{L}_s$ is the supervised objective of the cross-entropy loss ($\mathcal{H}$) on the B_L-sized labeled batch:

$$\mathcal{L}_s = \frac{1}{B_L} \sum_{i=1}^{B_L} \mathcal{H}(\mathbf{y}_i, \mathbf{p}(\mathbf{y}|\mathbf{x}_i^l)). \tag{1.1}$$

For the unsupervised loss, $\mathcal{L}_u$ is the *cross-entropy* between the model's prediction of the strongly-augmented data $\Omega(\mathbf{x}^u)$ and pseudo-labels from the weakly-augmented data $\omega(\mathbf{x}^u)$:

$$\mathcal{L}_u = \frac{1}{B_U} \sum_{i=1}^{B_U} \mathbb{1}(\max(\mathbf{p} \geq \tau)\mathcal{H}(\hat{\mathbf{p}}_i, \mathbf{p}(\mathbf{y}|\Omega(\mathbf{x}_i^u))), \tag{1.2}$$

where $\mathbf{p}$ is the abbreviation of $\mathbf{p}(\mathbf{y}|\omega(\mathbf{x}^u))$, and $\hat{\mathbf{p}}$ is the one-hot pseudo-label $\operatorname{argmax}(\mathbf{p})$; B_U is the batch size for unlabeled data. $\mathbb{1}(\max(\mathbf{p} \geq \tau)$ refers to the confidence threshold that only samples under the condition that $\max(\mathbf{p} \geq \tau$ will be involved in the unsupervised loss.

Confidence thresholding methods [Sohn et al., 2020, Xie et al., 2020, Zhang et al., 2021, Xu et al., 2021a] adopt a threshold to enroll the unlabeled samples with high confidence into training. FixMatch [Sohn et al., 2020] uses a fixed threshold to select high-quality pseudo-labels, which limits the data utilization ratio and leads to imbalanced pseudo-label distribution. Dash [Xu et al., 2021a] gradually increases the threshold during training to improve the utilization of unlabeled data. FlexMatch [Zhang et al., 2021] designs class-wise thresholds and lowers the thresholds for classes that are more difficult to learn, alleviating class imbalance.

1.4 Summary of Key Contributions

Within the domain of learning with limited supervision, our research encompasses several critical advancements. Initially, we introduce a novel fine-tuning method tailored to enhance the efficacy of few-shot learning, particularly in cross-domain scenarios. Building upon this, we extend our comprehension of few-shot fine-tuning into the transductive setting. Here, we present innovative weighting techniques to harness the potential of unlabeled data during the testing phase. In addition, we confront the intricate balance between data quality and quantity when leveraging unlabeled training data in semi-supervised learning. To address this challenge, we introduce the SoftMatch method, which allows for the adaptive integration of unlabeled data during the training process.

- In recent works, utilizing a deep network trained on meta-training set serves as a strong baseline in few-shot learning. We move forward to refine novel-class features by finetuning a trained deep network. Finetuning is designed to focus on reducing biases in novel-class feature distributions, which we define as two aspects: class-agnostic and class-specific biases. Class-agnostic bias is defined as the distribution shifting introduced by domain difference, which we propose a Distribution Calibration Module (DCM) to reduce. DCM has good properties of eliminating domain differences and fast feature adaptation during optimization. Class-specific bias is defined as the biased estimation using a few samples in novel classes, which we propose Selected Sampling (SS) to reduce. Without inferring the actual class distribution, SS is designed by running sampling using proposal distributions around support-set samples. By powering finetuning with DCM and SS, we achieve state-of-the-art results on Meta-Dataset Triantafillou et al. [2019] with consistent performance boosts over ten datasets from different domains. We believe our simple yet effective method demonstrates its possibility to be applied to practical few-shot applications.

- Furthermore, we first observe that the few-shot fine-tuned methods are learned with the imbalanced class marginal distribution, leading to imbalanced per-class testing accuracy. This observation further motivates us to propose Transductive Fine-tuning with Margin-based uncertainty weighting and Probability regularization (TF-MP), which learns a more balanced class marginal distribution. Specifically, we first conduct sample weighting on unlabeled testing data with margin-based uncertainty scores and further regularize each test sample's categorical probability. TF-MP achieves state-of-the-art performance on in- / out-of-distribution evaluations of Meta-Dataset Triantafillou et al. [2019] and surpasses previous transductive methods by a large margin.

- To effectively leverage the limited labeled data and massive unlabeled data to improve the model's generalization performance in Semi-Supervised Learning, we first revisit the popular pseudo-labeling methods via a unified sample weighting formulation and demonstrate the inherent quantity-quality trade-off problem of pseudo-labeling with thresholding, which may prohibit learning. We propose SoftMatch to overcome the trade-off by maintaining both high quantity and high quality of pseudo-labels during training, effectively exploiting the unlabeled data. We derive a truncated Gaussian function to weight samples based on their confidence, which can be viewed as a soft version of the confidence threshold. We further enhance the utilization of weakly learned classes by proposing a uniform alignment approach. In experiments, SoftMatch shows substantial improvements across various benchmarks and achieves state-of-the-art performance.

In this dissertation, we introduce the key contributions separately in the following chapters. In Chapter 2, we discuss the works to enhance the effectiveness and efficiency of few-shot fine-tuning. In Chapter 3, we first introduce the transductive few-shot fine-tuning baseline and analyze the importance of reducing the utilization of wrongly predicted testing data during test-time adaption. In Chapter 4, we firstly analyze the quantity-quality trade-off presented in the recent works in Semi-Supervised Learning using a unified weighting scheme. We then introduce the methodology SoftMatch, which we developed to weight unlabeled data adaptively during training. In each chapter, comprehensive experimental results on ablation studies and comparison with the state-of-the-art performances are provided for evaluation purposes. In Chapter 5, we discuss the future work towards learning with limited supervision.

Chapter 2

Enhancing Finetuning in Few-shot Learning

2.1 Leverage a Feature Extractor to Few-Shot Problem

We first formalize the few-shot classification setting with notation. Let $(\mathbf{x}, y)$ denote an image with its ground-truth label. The episode is the basic unit in constructing the few-shot learning scenarios in few-shot learning. Specifically, training and test images are referred to as the support and query set respectively, and are collectively called a C-way K-shot episode. We denote the support set as $\mathcal{D}_s = \{(\mathbf{x_i}, y_i)\}_{i=1}^{N_s}$ and the query set as $\mathcal{D}_q = \{(\mathbf{x_i}, y_i)\}_{i=1}^{N_q}$, where $y_i \in C$ and $|C|$ is the number of ways or classes and N_s equals to $C \times K$. Meanwhile, from the perspective of evaluation datasets, the whole dataset is split into a (meta-)training set and a (meta-)testing set. (Meta-)training set, also known as base classes, is used for training the algorithms in few-shot learning. For evaluation, multiple episodes would be sampled from the (meta-)testing set, and the fine-tuning stage is also conducted using the support set from each episode during evaluation.

In recent works Chen et al. [2019], Tian et al. [2020], Chen et al. [2020], Dhillon et al. [2019], the importance of utilizing a good feature embedding in few-shot learning is well studied and addressed. As shown in Fig. 2.1, we compared the difference between baseline methods Chen et al. [2019, 2020], Tao et al. [2022] of leveraging a pre-trained feature extractor. Firstly, a feature embedding is pre-trained as a classification task using a meta-training set (base classes).

As shown in Fig. 2.1, we compare the difference between different two-stage works utilizing a feature extractor pre-trained on base classes. In Chen et al. [2019], the fine-tuning stage is only to fine-tune the randomly initialized weights in the classifier while the feature extractor is fixed. Finetuning on the meta-test set (novel classes) Tian et al. [2020], Yang et al. [2021], Dhillon et al. [2019] is shown to surpass most meta-learning methods. In Chen et al. [2019], the classifier using cosine distance in the softmax cross-