

Conditional Consistency Regularization for Semi-supervised Multi-label Classification

ZHENGNING WU

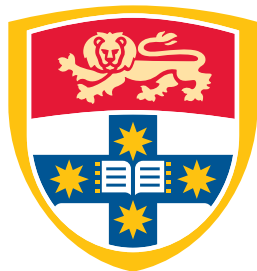
SID: XXXXXXXXXX

Supervisor: Dr Tongliang Liu

This thesis is submitted in fulfillment of
the requirements for the degree of
Master of Philosophy

School of Computer Science
The University of Sydney
Australia

27 April 2023



THE UNIVERSITY OF
SYDNEY

Authorship attribution statement

It should be noted that there are no chapters or materials in this paper that have been published.

Attesting authorship attribution statement

In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Student: Zhengning Wu

Signature:

Date: 27 April 2023

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Student: Tongliang Liu

Signature:

Date: 27 April 2023

Originality Statement

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes. I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Student: Zhengning Wu

Signature:

Date: 27 April 2023

Abstract

In practical scenarios, a sample may have multiple labels that reveal its classes instead of a single label, which is widely known as multi-label classification (MLC). However, some practical situations may lack reliable labels due to the high cost, time-consuming and professional labelling process. Although Semi-supervised classification may become a potential solution, most of the outstanding existing methods are customized for the single-label situation and ignore multi-label situations. Consistency regularization has performed great success in Weakly/Semi-supervised Single-label classification (SS-SLC), but few efforts have been devoted to semi-supervised Multi-label classification (SS-MLC). A simple solution for introducing consistency regularization to SS-MLC is to regularize predictions of models to be consistent with different augmentation of the same image. Nonetheless, the solution lacks attention to label relations which are crucial components of Multi-label classification.

In the thesis, I go beyond the consistency regularization in SS-SLC and propose Conditional Consistency Regularization (CCR) that is designed for SS-MLC. To be specific, we make potential labels (grand-truth label for labeled samples, pseudo-label for unlabeled samples) conditioned on different label states (i.e., positive, negative, or unknown for each class). By regularizing the two predictions to be invariant, the model can learn label relations implicitly between two different label states, which can boost classification performance. The comprehensive experiments that are conducted on different datasets show that the proposed method can surpass state-of-art SS-MLC and MLC methods by a large gap.

Acknowledgements

Firstly, I would like to thank my supervisor Dr Tongliang Liu sincerely. He provides professional guidance and advice to help me complete the research successfully. Secondly, I would like to thank Xiaobo Xia, Tianyu He and Xu Shen who gives me academic support during the research. And I also want to thank all the teammates in the Sydney AI center who provide outstanding academic discussions. Moreover, I would like to thank my parents and my girlfriend Jie Wang who gives me emotional support during the research. Finally, I want to express my sincere thanks to all teachers, staff and all classmates who help me at the University of Sydney. The 6 years at the University of Sydney is my most valuable experience in my academia.

CONTENTS

Authorship attribution statement	ii
Attesting authorship attribution statement	iii
Originality Statement	iv
Abstract	v
Acknowledgements	vi
List of Figures	ix
List of Tables	x
Chapter 1 Introduction	1
1.1 Research Background	1
1.1.1 Semi-supervised Single/Multi-label classification (SS-SLC/SS-MLC)	1
1.1.2 Consistency regularization	3
1.2 Current problems	3
1.3 Major work and contribution	4
1.3.1 Major work	4
1.3.2 Contribution	5
1.4 Structure of the thesis	6
Chapter 2 Literature Review	8
2.1 Multi-label classification (MLC)	8
2.2 Semi-supervised Single-label classification (SS-SLC)	9
2.3 Semi-supervised Multi-label classification (SS-MLC)	10
2.4 Consistency Regularization	10
2.5 Summary	11
Chapter 3 Method	12

3.1	Method Overview	12
3.2	Multi-View Image Consistency Regularization	15
3.3	Conditional Consistency Regularization with Label State.....	16
3.4	Training Objective	18
3.5	Comparison with the methods in SS-SLC.....	19
Chapter 4 Experiments		21
4.1	Experimental Setup	21
4.2	Comparison with State-of-the-Arts.....	22
4.3	Ablation Study	24
4.4	The Choices of Hyper-parameters	28
Chapter 5 Discussion		31
5.1	The proposed method	31
5.2	Comparison with previous work.....	32
5.3	Open problems of the proposed method	32
5.3.1	Limitation of labeled samples	33
5.3.2	Limitation of unlabeled samples	34
5.3.3	other methods of multi-label classification with limited supervision.....	34
5.4	Significance of the work	35
Chapter 6 Conclusion and future work		36
6.1	Conclusion	36
6.2	Future work	37
Bibliography		39

List of Figures

1.1	Difference between conventional consistency regularization and our proposed Conditional Consistency Regularization (CCR).	5
2.1	Difference between MLC, SS-SLC and SS-MLC.	9
3.1	Illustration of the proposed Conditional Consistency Regularization (CCR) framework for semi-supervised multi-label classification.	14
4.1	Illustrations of the examples of the predictions by CCR.	28
4.2	Ablation study on the thresholds r_u and r_l .	30

List of Tables

4.1	Performance comparison on COCO-80 dataset. The best results (%) are bolded.	23
4.2	Performance comparison on VOC-2007 dataset. The best results (%) are bolded.	24
4.3	Comparison with advanced MLC methods on VOC-2007 dataset. The best results (%) are bolded.	24
4.4	Ablation study on the consistency regularization on COCO-80 dataset. The best results (%) are bolded.	26
4.5	Ablation study on the consistency regularization on VOC-2007 dataset. The best results (%) are bolded.	26
4.6	Ablation study on various model architectures.	27
4.7	Ablation study on the weighting function $w(\cdot)$. The best results (%) are bolded.	29
4.8	Ablation study on the ratio between labeled examples and unlabeled examples for each training iteration. The best results (%) are bolded.	29

Introduction

1.1 Research Background

With the development of machine learning technology, many machine learning methods have been proposed for many tasks, such as object detection, image classification, regression, etc. A practical classification scenario in the real world is that a simple sample may have a great number of descriptions or classes. For example, an image shows "*A man rides a motorbike on the road.*" The image may contains **'person', 'motorbike'** and **'road'**. This kind of classification task is also known as multi-label classification (MLC), which aims to assign multiple labels to a single input sample and keep vibrant in recent years (Chen et al., 2019b; Xie and Huang, 2021; Cole et al., 2021; Hu et al., 2021; Gupta et al., 2021; Gao and Zhou, 2021). Compared with single-label classification, a single input sample only have one corresponding label (LeCun et al., 2015; He et al., 2016; Simonyan and Zisserman, 2015; Szegedy et al., 2016; Xia et al., 2020; Wu et al., 2021), MLC is more valuable and challenging because most practical scenarios tend to have multi-label in a sample and consideration of more labels and their relations (Gong et al., 2013; Wang et al., 2016; Li et al., 2016; Zhu et al., 2017; Xu et al., 2016). However, due to the high-expense, labour-exhaustive, time-consuming and speciality of labelling process (Liu et al., 2021), it is challenging to obtain reliable labels in practical cases. As a result of that, it is necessary and meaningful to make the model can utilize massive unlabeled samples, which is also widely known as Semi-supervised multi-label classification (SS-MLC).

1.1.1 Semi-supervised Single/Multi-label classification (SS-SLC/SS-MLC)

1.1.1.1 SS-SLC

SS-SLC is a task that set up a learning model and learning knowledge from a dataset that contains a few of labeled samples and a large quantity of unlabeled samples, where a single labeled sample has only a

single corresponding label. Normally, the most of SS-SLC tasks are inductive setting. It assumes that the test samples are unknown samples and the unlabeled samples are not the test samples (Zhou, 2018). The learned model is able to predict the label accurately when given unseen test samples. Formally, the goal of SS-SLC is to learn a model $f_\theta(\cdot)$ from N_u unlabeled images $D_u = \{(x_j)\}_{j=1}^{N_u}$ and N_l labeled images $D_l = \{(x_i, y_i)\}_{i=1}^{N_l}$, where x is the input samples and y is corresponding labels that can describe the input x . A good learned model $f_\theta(\cdot)$ is able to accurately predict y from an unseen instance x (Van Engelen and Hoos, 2020). There is also a special SS-SLC setting, which is named the transductive setting. It assumes that the test data is given in advance and the unlabeled samples are the test samples. The major aim is to increase the performance on test samples (Zhou, 2018). In the thesis, we will focus on inductive settings, which are closer to practical tasks.

This kind of learning paradigm could partially resolve the problems of lack labels. Compared with common single-label classification, it can leverage massive unlabelled samples with outstanding performance that reduce the requirements of massive accurate labels. Therefore, semi-supervised classification is one of the hottest and most popular research topics in the field of machine learning in recent years (Zhou, 2018; Yang et al., 2021). A great number of researchers have proposed impressive methods to deal with this task, like the self-training method (also known as pseudo-label) (Lee et al., 2013), consistency regularization method (Samuli and Timo, 2017; Xie et al., 2020a), generative method (Kingma et al., 2014), co-training method (Blum and Mitchell, 1998; Qiao et al., 2018) and hybrid method (Sohn et al., 2020). Although these methods achieve excellent performance, they have not been extended to SS-MLC, which is also a practical situation in real-world applications.

1.1.1.2 SS-MLC

SS-MLC is a task that proceeds MLC with a small number of labeled samples and a massive number of unlabeled samples. Some early age methods utilize label propagation methods to address SS-MLC under transductive setting (Lin et al., 2017a; Kong et al., 2011; Gong et al., 2016). However, these methods are difficult to generalize to unseen test samples. Some researchers also proposed methods in inductive setting (Jing et al., 2015; Wu et al., 2015), like manifold regularization, probabilistic framework (Chu et al., 2018), pseudo-labels (Wang et al., 2021), etc, which the model could generalize to unseen test samples. Although some methods are effective, these methods ignore the natural fact that the model predictions should be similar when inputting a perturbed version of the same image, which is also known as consistency regularization (Sajjadi et al., 2016; Tarvainen and Valpola, 2017).

1.1.2 Consistency regularization

Consistency regularization is one of the state-of-art methods that achieved promising performance in SS-SLC. Generally, this method assumes that the predictions of the learned model should be consistent when we input different image augmentations that originate from the same image (Xie et al., 2020a; Berthelot et al., 2019b; Huang et al., 2022). There are many image augmentation methods that have been proposed, like some weak augmentations, which only slightly add perturbations and strong augmentation, which makes large variations of the image (Cubuk et al., 2020). Consistency regularization has been verified in many previous papers (Sohn et al., 2020; Zhang et al., 2021) that it can achieve outstanding performance in weakly/semi-supervised learning areas. It is necessary to extend advanced MLC and SS-SLC methods to SS-MLC area to obtain better performance.

Specific literature on MLC, SS-SLC, SS-MLC and Consistency Regularization will be exhaustively discussed in section 2 below. This part only has a brief introduction of the background.

1.2 Current problems

Currently, there are no feasible and effective solutions to utilize a small number of labeled samples and a large number of unlabeled samples in multi-label learning, also named semi-supervised multi-label classification, which limits the applications of multi-label learning. Although some previous researchers have proposed some impressive methods in MLC, SS-SLC and SS-MLC, they have some weaknesses in practical situations. Methods in MLC (Lanchantin et al., 2021) achieve excellent performance under reliable supervision by considering multi-labels and their relations. However, it is difficult to obtain plenty of reliable labels due to the high-cost, time-consuming and professional labeling process in a practical situation. These methods can not utilize a great number of unlabeled samples that easy to obtain in the training phase, which is a practical challenge. Methods in SS-SLC (Berthelot et al., 2019a; Cascante-Bonilla et al., 2021) achieve outstanding performance via using unlabeled samples. However, they can not adopt multi-label situations and they also can not learn label relations between multi-label, which has been verified to be unique and crucial for multi-label situation (Gong et al., 2013; Wang et al., 2016; Li et al., 2016; Zhu et al., 2017). Although few methods in SS-MLC are effective, they ignore a natural fact: the model predictions should be similar when inputting the perturbed version of the same image, known as consistency regularization, which is an advanced method in SS-SLC. It is meaningful to be introduced and tailored for SS-MLC. One intuitive solution for introducing consistency regularization

to SS-MLC is to regularize the predictions of the model to be consistent under different augmentations of the same image. However, such a direct method is not suitable for SS-MLC because it lacks the learning process of label relations that have been proved vital for MLC. Therefore, it is meaningful and necessary to conduct research that proposes an algorithm adopted and customized for SS-MLC that can leverage unlabeled samples and learn label-relations between multi-label.

1.3 Major work and contribution

1.3.1 Major work

Compared to the intuitive solution discussed above, the thesis goes beyond the traditional consistency regularization and focuses on customizing for SS-MLC. The proposed method goes a step further by giving consistency regularization with the ability of modeling label relations, which we named Conditional Consistency Regularization (CCR). To be specific, the CCR models the input contains an image and a label state. The label state denotes the state of each class as positive, negative or unknown. In the training process, the model makes predictions of two augmented images conditioned on different label states respectively. Because of the minimization of the distance between these two predictions, the outputs obtained from different label states and augmented images are encouraged to be consistent. In this way, the model is able to learn label relations from different label states. Specifically, for labeled samples, the label state is a randomly masked version of the ground truth label, where parts of the label state are set to unknown. While, for unlabeled samples, because there is no ground truth label, we set up pseudo-label memory that utilizes the pseudo-label of each image in the latest epoch. Then, similar to labelled samples, the label is generated by randomly masking the pseudo-label with the unknown state. At test time, the label state is set to unknown for every class in order to obtain accurate predictions.

To better explain our method intuitively, we can take Fig 1.1 as an example, the label state of `Person` for the weakly-augmented image and the label state of `Motorbike` for the strong-augmented image is set to be positive. When we push the two predictions to be consistent, the model is encouraged to output positive predictions for both `Person` and `Motorbike`, which means the model can learn class `Person` and class `Motorbike` are likely to appear in one image. Similarly, class `Motorbike` and class `Cat` are not likely to appear together in one image due to the same reason. Under this training process, the model can learn label relations between multi-label and images that can reason the unknown class (e.g., `Person`, which is involved in another label state) based on the known label state (e.g., `Motorbike`) by pushing

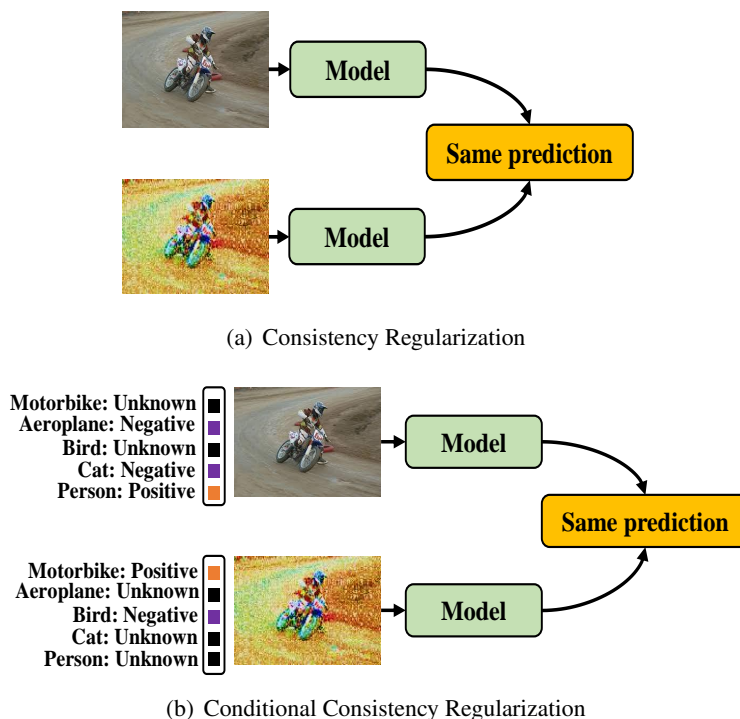


FIGURE 1.1: Difference between conventional consistency regularization and our proposed Conditional Consistency Regularization (CCR).

the model predictions to be the same to different label states. Namely, a higher confidence of a known class, like class `Person`, promotes a higher probability of an unknown class, like `Motorbike` that appears in the same image. In this way, the model can learn label relations implicitly by the proposed CCR framework. In order to verify the feasibility and the effectiveness of the CCR, I conduct the experiments on various practical datasets, like VOC-2007 and COCO-80. The CCR's performance could outperform the state-of-arts at about 10% on average.

1.3.2 Contribution

The major contributions are summarized as follows:

- The thesis is the first to introduce consistency regularization, which is known as the outstanding technology in SS-SLC to SS-MLC.
- In order to learn label relations and knowledge of unlabeled samples, we proposed CCR, which is a novel SS-MLC method with detailed explanations, discussions, and analysis.

- Extensive and comprehensive experiments have been carried out on different datasets and various baselines, with complete ablation studies to evaluate the effectiveness of the CCR and result in a performance boost in SS-MLC.
- Discussion of CCR has been introduced, which describes the advantages, disadvantages and future work of the CCR.

1.4 Structure of the thesis

The structure of the thesis is summarized as follows. Chapter 1 is the introduction of the thesis. Chapter 2 is the literature review of the thesis. Chapter 3 introduces the proposed CCR method in detail. Chapter 4 is about the experiments. Chapter 5 is a discussion of the proposed method. And Chapter 6 is the final conclusion of the thesis.

- Chapter 1: Introduction

This chapter provides brief introduction of the research background, including MLC, SS-SLC, SS-MLC and consistent regularization. On this basis, this chapter also summarizes the major work, performance and contribution of the proposed method.

- Chapter 2: Literature Review

This chapter briefly summarizes the relevant work of the thesis, including MLC, SS-SLC, SS-MLC and consistency regularization.

- Chapter 3: Method

This chapter gives a detailed explanation of the proposed method - CCR. It includes Problem Formulation, Method Overview, Multi-view Image Consistency Regularization, Conditional Consistency Regularization with Label State, Training Objective and Comparison with other baselines, etc.

- Chapter 4: Experiments

This chapter describes the experiment in detail, including experimental setup, performance comparison with other benchmarks, ablation research, etc.

- Chapter 5: Discussion

This chapter discusses the proposed method, Comparison with previous work, open problems, potential extension and significance of the proposed method.

- Chapter 6: Conclusion

This chapter provides a summary of the thesis and proposes some meaningful ideas that could be done in the future.

Literature Review

This chapter will discuss related work on Multi-label classification (MLC) in section 2.1, Semi-supervised Single-label classification (SS-SLC) in section 2.2, Semi-supervised Multi-label classification (SS-MLC) in section 2.3 and consistency classification in section 2.4. Moreover, Fig 2.1 shows the difference between MLC, SS-SLC and SS-MLC.

2.1 Multi-label classification (MLC)

MLC keeps vibrant in recent years with outstanding performance and benefits a massive number of practical applications, like text classification, image classification, protein classification, etc. It assigns multiple labels to a single input instance. In general, the model could learn knowledge from a dataset that contains a large number of training examples and their corresponding multi-labels (Liu et al., 2021). The learned model could predict accurately the multi-labels when the unseen samples are inputted. Compared with single-label classification, the effective MLC methods not only learned excellent features of the image but also explored the fund of knowledge of label relations to improve classification performance, which is a unique and pivotal factor to achieve excellent performance. Lots of methods have been proposed with impressive performance in MLC areas. Some methods estimate the joint probabilities of predicted labels from given inputs by using chain rules (Dembczynski et al., 2010; Nam et al., 2017; Read et al., 2009). Some methods leverage shared latent space of features and labels (Bhatia et al., 2015; Yeh et al., 2017). Some methods modified loss function that adapted MLC (Lin et al., 2017b; Ridnik et al., 2021; Wu et al., 2020). Some researchers modelled label relations and label dependencies by utilizing inference formulation (Guo and Gu, 2011; Li et al., 2016). Finally, some state-of-art methods introduced graph neural networks according to label occurrence frequency (Chen et al., 2021, 2019b,a). Those previous methods have achieved promising classification performance in MLC, they can not utilize massive unlabeled samples during the training process.

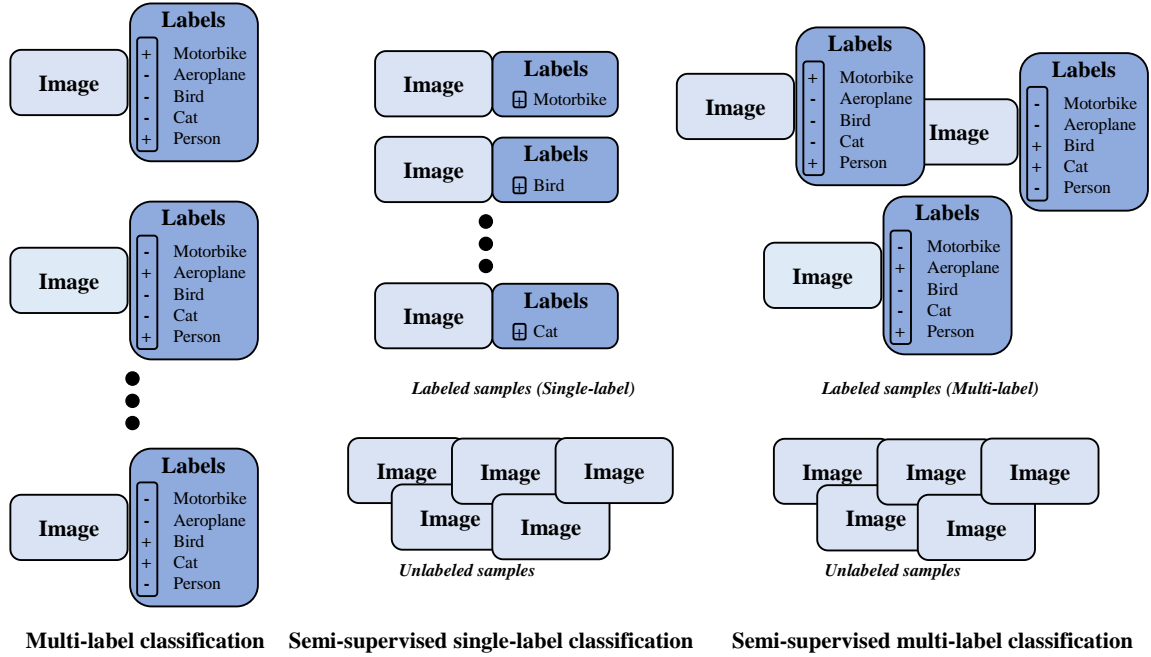


FIGURE 2.1: Difference between MLC, SS-SLC and SS-MLC.

2.2 Semi-supervised Single-label classification (SS-SLC)

Semi-supervised single-label classification (SS-SLC) is a problem setting that the training set that contains a small proportion of labeled examples and a large proportion of unlabeled examples, which is also known as semi-supervised classification or semi-supervised learning. It is a popular research topic that has a great number of methods in recent days. For example, Pseudo-label methods (also known as self-training) utilize the model's high confidence prediction of unlabeled samples which could be considered as labeled samples to increase classification performance (Lee et al., 2013; Xie et al., 2020b). Consistency regularization is a kind of method that states that the variations of input data should not affect the output of the model (Samuli and Timo, 2017; Xie et al., 2020a). Generative methods can set up a feasible connection between unlabeled samples and labels by using $P(x, y) = P(y)P(x|y)$ (Dai et al., 2017; Kumar et al., 2017; Liu et al., 2020). On top of those methods, There are other types of methods that can achieve outstanding performance, such as graph-based methods (Isken et al., 2019; Kipf and Welling, 2016; Liu et al., 2017; Wang et al., 2020), hybrid methods (Sohn et al., 2020) that mix the methods above. For example, Mixmatch (Berthelot et al., 2019b) produces pseudo-labels for augmented unlabeled samples and mixes the labeled and unlabeled samples via Mixup (Zhang et al., 2016) to learn better representations. Remixmatch (Berthelot et al., 2019a) extends from the Mixmatch

via adding novel augmentation skills and distribution alignment. Fixmatch (Sohn et al., 2020) proposed a simple but effective way to combine pseudo-label methods and consistency methods. PAWS (Assran et al., 2021) tailed novel algorithms for SS-SLC that can utilize both labeled samples and unlabeled samples under self-supervised learning. Although these semi-supervised single-label learning methods have promising performance, they can not adapt to multi-label scenarios and are unable to learn label relations which is unique and crucial for better performance in MLC.

2.3 Semi-supervised Multi-label classification (SS-MLC)

SS-MLC is conducting MLC with small parts of labeled samples and large parts of unlabeled samples which the thesis focuses on. label propagation methods are popular in the early days. For example, a curriculum-learning-based label propagation is proposed by (Gong et al., 2016) in 2016. DLP and DGFLP (Wang et al., 2013; Lin et al., 2017a) design a dynamic label propagation process. While the majority of these label propagation methods are under a transductive setting, which means that the algorithm is unable to generalize to unseen test examples. As for the algorithm that can be generalized to unseen test samples, which is also known as the inductive setting. SLRM (Jing et al., 2015) utilizes unlabeled examples by constructing a low-rank mapping from feature space to label space based on manifold regularization. COINS (Zhan and Zhang, 2017) maximizes the difference between two classifiers and updates the model by ranking predictions on unlabeled examples, which is an extension of the co-training strategy. Moreover, Chu et al. (Chu et al., 2018) constructs a generative model by using a sequence architecture. DRS (Wang et al., 2021) aligns the feature distribution of labeled and unlabeled examples into a latent space by proposing a two-classifier domain adaption network and learning the label relations by using a graph-based relation network. Although effective, the aforementioned methods ignore the natural fact that the model predictions should be similar for perturbed versions of the same image. In the thesis, I propose to introduce consistency regularization and go beyond it by building relations between different label states.

2.4 Consistency Regularization

Consistency regularization is a kind of technique that states that the variations of input data should not affect the output of the model based on the manifold assumption (Yang et al., 2021). It could also be regarded as utilizing unlabeled examples to find a smooth manifold on the datasets (Belkin and Niyogi,

2001). Recently, consistency regularization is used in SS-SLC and becomes a key module of the state-of-the-art methods in the semi-supervised learning area. Existing methods put efforts into different aspects of the variations and training framework. For example, some methods add augmentations to images. Sajjadi et al. (Sajjadi et al., 2016) generates two stochastic augmentations. Xie et al. (Xie et al., 2020a) investigates the input variation effects in consistency regularization and utilizes high-quality data augmentation for images, like AutoAugment (Cubuk et al., 2019) and RandAugment (Cubuk et al., 2020). Miyato et al. (Miyato et al., 2018) produces adversarial perturbations. The variations can also be added to the neural network, such as adding Gaussian noise in every network layer (Rasmus et al., 2015) and dropping some connections and layers (Zhang and Qi, 2020). Moreover, some variations are produced in the training framework. For example, Mean Teacher (Tarvainen and Valpola, 2017) builds a teacher-student framework that leverages the model parameter’s exponential moving averages. Temporal ensembling (Samuli and Timo, 2017) sets up a temporal training framework that computes the consistency loss on current epoch predictions and previous epoch predictions. The mentioned methods above achieved outstanding performance in SS-SLC. However, these methods are not specifically designed for MLC, since they lack the consideration of label relations and label dependencies. In contrast, the proposed CCR is tailored for SS-MLC and constructs the consistency by making the model predictions conditioned to different label states, which helps to make use of label relations for boosting classification as discussed.

2.5 Summary

Although the previous methods achieve impressive performance in MLC, SS-SLC, and SS-MLC, these methods have their specific weakness to address SS-MLC. The proposed method gives a feasible solution that takes the advantage of both MLC and semi-supervised classification by taking account of label relations and unlabeled samples which are unique and important in both areas.

Method

In this chapter, the thesis discusses the specific methods of conditional consistency regularization in SS-MLC. In section 3.1, I first introduce the overview of the method with formal problem formulation. In section 3.2, the thesis explains the consistency regularization that the method used. In section 3.3, the thesis discusses conditional consistency regularization via label state and its generation and discussion. In section 3.4, the thesis formally defines the training objective of the proposed method and finally compares the proposed method with the current advanced SS-SLC method in section 3.5.

3.1 Method Overview

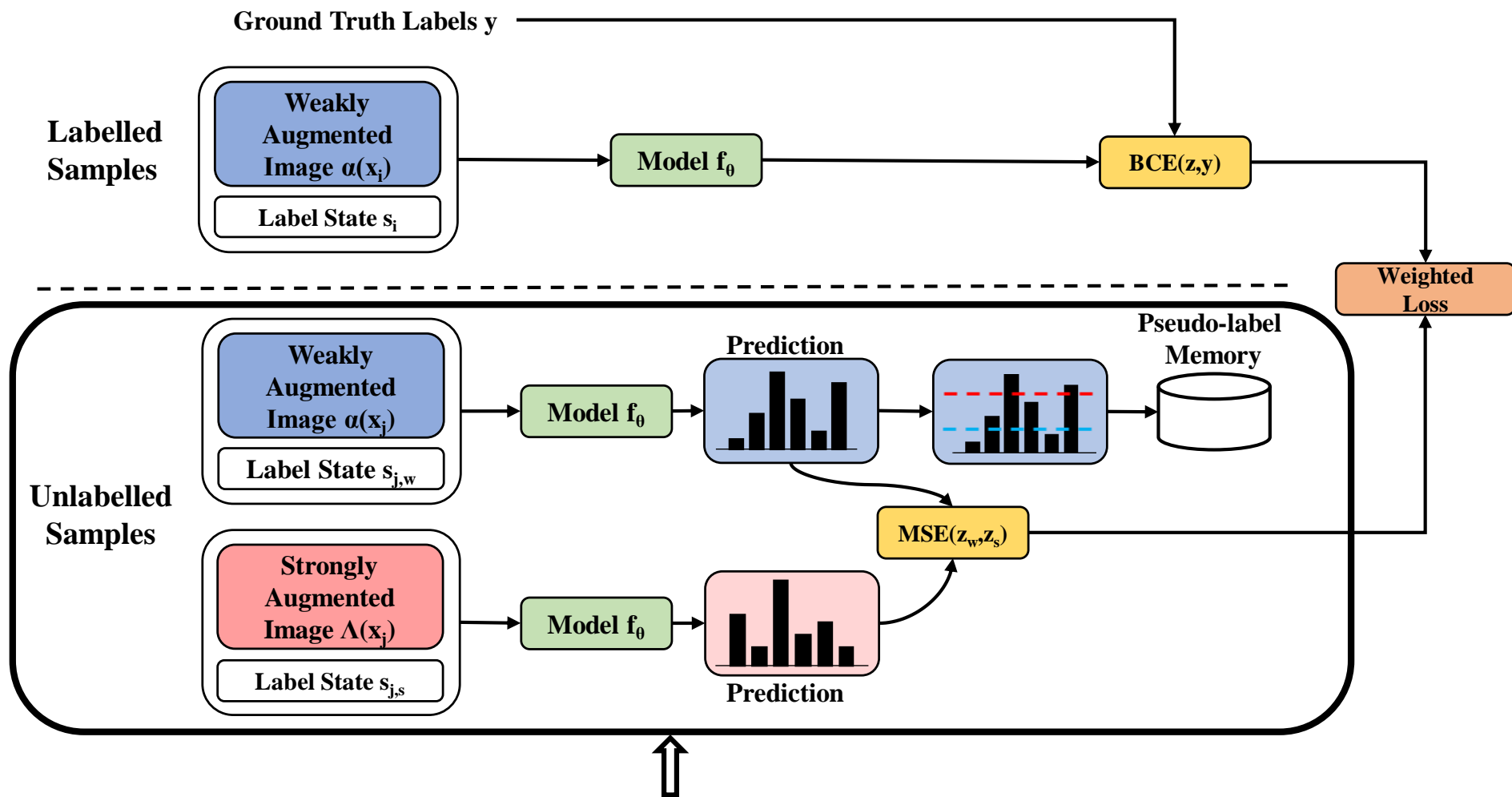
The thesis proposes Conditional Consistency Regularization (CCR) to deal with the SS-MLC problem. As shown in Fig. 3.1 and Alg. 1, the model input includes an augmented image and a label state that has three states of positive (P), negative (N), and unknown (U) for each class. For a labeled example x_i , following techniques in masked sequence modeling (Kenton and Toutanova, 2019; Lanchantin et al., 2021), the proposed method generates label states $s_i \in \{0, 1, U\}^C$ by masking a certain amount of ground truth labels as unknown. Then, the proposed method calculates the Binary Cross Entropy (BCE) loss only on those unknown classes. Hence, the model learns to use the known label states and the input image x_i , to predict the unknown classes.

For an unlabeled example x_j , since there is no ground-truth label, the proposed method builds a pseudo-label memory to cache the pseudo-label \bar{y}_j of each example x_j from the latest epoch. Subsequently, the label states s_j is generated from the masked version of \bar{y}_j like the labeled examples. As for image x_j , the proposed method applies a weak augmentation $\alpha(\cdot)$ and a strong augmentation $\Lambda(\cdot)$ to it respectively (Sohn et al., 2020), denoted as $\alpha(x_j)$ and $\Lambda(x_j)$. The two augmented views and their label states are then fed into the model $f_\theta(\cdot)$ to obtain the final predictions, which are encouraged to be consistent by a Mean Square Error (MSE) loss function in the training phase. It should be noted that

the masks for label states of different views are randomly generated, resulting in different label states for two augmented views. Therefore, when the proposed method pushes the two model predictions to be the same, the model learns the relations between the two different label states. On the other hand, the consistency between the two augmented views facilitates the learned models to be invariant under diverse data augmentation transforms.

Compared with previous methods, conditional consistency regularization makes predictions of two augmented images conditioned on different label states respectively which helps to learn the knowledge of unlabeled samples and their label relations. It should be noted that the proposed method is similar to most SS-SLC methods which only have one model, the labeled samples, unlabeled samples and augmented samples will be input into one model in every epoch.

As for the formal problem definition of SS-MLC, we can define that the goal of SS-MLC is to learn a feature embedding network (commonly neural network) $f_\theta(\cdot)$ from N_u unlabeled images $D_u = \{(x_j)\}_{j=1}^{N_u}$ and N_l labeled images $D_l = \{(x_i, y_i)\}_{i=1}^{N_l}$. For each image x_i , $y_i \in \{0, 1\}^C$ is the corresponding one-hot label, where C is the number of classes. We define $y[c] = 1$ if the image is associated with the c -th label, otherwise $y[c] = 0$. A good feature embedding network $f_\theta(\cdot)$ is able to accurately predict y from an unseen instance x .



CCR makes predictions of two augmented images conditioned on different label states respectively helps to learn knowledge of unlabeled samples and their label relations

FIGURE 3.1: Illustration of the proposed Conditional Consistency Regularization (CCR) framework for semi-supervised multi-label classification.

Algorithm 1 Pseudo-codes of our proposed method

Required: $(x_i, y_i) \rightarrow$ labeled training examples
Required: $x_j \rightarrow$ unlabeled training examples
Required: $f_\theta(\cdot) \rightarrow$ the model with parameters θ
Required: $\alpha(\cdot)$ and $\Lambda(\cdot) \rightarrow$ weak and strong augmentation functions
Required: $g(\cdot) \rightarrow$ the label state generation function
Required: $h(\cdot|r_u, r_l) \rightarrow$ the pseudo-label generation function with thresholds r_u and r_l
Required: $M \rightarrow$ pseudo-label memory for the unlabeled images
Required: $w(t) \rightarrow$ the ramp-up weighting functions of the t -th epoch
Required: B_i and $B_j \rightarrow$ batch sizes for labeled and unlabeled examples

- 1: **for** t **in** $[1, num_epochs]$ **do**
- 2: **for** each mini batch B **do**
- 3: $s_i \leftarrow g(y_i), s_{j,w} \leftarrow g(M_j), s_{j,s} \leftarrow g(M_j)$ {Get the label states}
- 4: $z_i \leftarrow f_\theta(\alpha(x_i)|s_i)$ {Feed-forward for labeled examples}
- 5: $z_{j,w} \leftarrow f_\theta(\alpha(x_j)|s_{j,w})$ {Feed-forward for unlabeled examples with weak augmentation}
- 6: $z_{j,s} \leftarrow f_\theta(\Lambda(x_j)|s_{j,s})$ {Feed-forward for unlabeled examples with strong augmentation}
- 7: $M_j \leftarrow h(z_{j,w}|r_u, r_l)$ {Update the pseudo-label memory}
- 8: $loss \leftarrow \frac{1}{B_i} \text{BCE}(z_i, y_i) + \text{BCE loss for labeled examples}$
- 9: $w(t) \frac{1}{B_j} \text{MSE}(z_{j,w}, z_{j,s})$ {MSE loss for unlabeled examples}
- 10: **end for**
- 11: **end for**

3.2 Multi-View Image Consistency Regularization

Inspired by FixMatch (Sohn et al., 2020) in SS-SLC, we introduce the multi-view image consistency regularization into SS-MLC. It generates two different views of the same unlabeled image by using *weak* augmentation $\alpha(\cdot)$ and *strong* augmentation $\Lambda(\cdot)$ respectively. In more detail, the weak augmentation consists of a horizontal flip with a probability of 50% and a random crop operation. While the strong augmentation consists of RandAugment (Cubuk et al., 2020; Xie et al., 2020a) and Cutout (DeVries and Taylor, 2017). RankAugment works by randomly selecting transformation on each input image, including AutoContrast, Brightness, Equalize, Sharpness, Posterize, Solarize, etc.

In the multi-view image consistency regularization, for each unlabeled example x_j , two augmented views $\alpha(x_j)$ and $\Lambda(x_j)$ are generated. They are then fed into the model to obtain two predictions. By minimizing the distance between the two predictions, we are able to impose a constraint on the model to make it harder to memorize the training data. Therefore, the model will be more robust when generalizing to unseen data (Berthelot et al., 2019b).

Although the multi-view image consistency regularization achieves great successes in SS-SLC, it lacks the consideration of label relations, which is vital for MLC (Gong et al., 2013; Wang et al., 2016; Li

et al., 2016; Zhu et al., 2017). To tackle this issue, in the following subsection, we will introduce our proposed CCR for SS-MLC.

3.3 Conditional Consistency Regularization with Label State

Label state

The label state s represents the state of labels for each input image. There are three possible states: positive (P), negative (N), or unknown (U). For example, if we have prior knowledge that the image is associated with the c -th label, $s[c]$ is positive. If not, $s[c]$ is negative. It also can be unknown if we give no prior knowledge of the c -th class.

Label state generation for labeled examples

During training, for the labeled example x_i , we have a ground truth label y_i , which can be used as the label state s_i . However, directly giving y_i to the model will inevitably make the model fall into a trivial solution since the model simply learns to output the given y_i . To handle this problem, we refer to masked sequence modeling (Kenton and Toutanova, 2019; Lanchantin et al., 2021). Specifically, we mask some percentages of y_i at random and use the remaining parts (via the label state) to predict the masked labels. In this case, the classes corresponding to the masked position are set as the unknown state.

Formally, for a given label $y_i = (y_i^1, y_i^2, \dots, y_i^C)$, where C is number of possible classes. We randomly mask 25% to 100% labels and replace them with a special symbol [unknown] as did in the literature (Lanchantin et al., 2021). Denote κ as the set of masked positions, y_i^κ as the set of masked labels, and s_i as the labels after masking. As shown in the example in the upper case of Fig. 1.1(b), $\kappa = \{1, 3\}$, $y_i^\kappa = \{y_i^1, y_i^3\}$, and $s_i = g(y_i) = ([\text{unknown}], y_i^2, [\text{unknown}], y_i^4, y_i^5) = ([\text{U}], [\text{N}], [\text{U}], [\text{N}], [\text{P}])$, where $g(\cdot)$ is the label state generation function. To integrate the generated label state and the input image, we transform each element in s_i to a d dimension feature vector by a learned embedding layer of a size $d \times 3$. Then, the embedded label state, acting as the condition, is added to the image feature as the model input.

Label state generation for unlabeled examples

As there are no ground truth labels for the unlabeled examples, one possible solution is to employ pseudo-labels as an alternative. To achieve this goal, for each training iteration, we should evaluate

the weakly-augmented image to obtain the pseudo-labels first. Then, the label state can be generated from the obtained pseudo-labels. However, there are two feed-forward runnings for each unlabeled example in this case, which makes the training process inefficient and time-consuming. To tackle this challenge, we propose to use pseudo-label memory with a size of $d \times N_u$, which stores the latest pseudo-label for each unlabeled example. More specifically, at each training iteration, we use the function $h(\cdot|r_u, r_l)$ to generate pseudo-labels from the prediction of the weakly-augmented image, where r_u is the upper threshold and r_l is the lower threshold. If the probability of a class is higher than the upper threshold r_u , we assign the pseudo-label of this class as positive. While, if the probability of a class is lower than the lower threshold r_l , we assign the pseudo-label of this class as negative. For the class with probabilities between r_u and r_l , we assign the pseudo-label of this class as unknown. In our implementation, the pseudo-label memory is initialized as unknown for all unlabeled examples. The reason why we set threshold r_l and r_u and choose relatively high confidence labels for pseudo-labels is that higher confidence usually results in more reliable pseudo-labels, which possibly reduce the side effect of noisy pseudo-labels, especially at the beginning of the training when the model’s predictions are unstable.

After acquiring the pseudo-label memory, we generate the label state for each unlabeled example by mimicking the masking process of the labeled example as described before. Note that, different from the ground truth labels that only have positive or negative states, there actually exists the unknown state in the pseudo-label memory (i.e., the classes with probabilities between r_u and r_l). Therefore, we only mask 25% to 100% positive or negative states for unlabeled examples.

It should be noticed that the label state generation process is conducted for each augmented image. Therefore, the label states for the two different views (i.e., weakly-augmented and strongly augmented views) of unlabeled examples are likely to be different. When we minimize the distance between the two model predictions that are obtained from these two different views, the model is encouraged to output the same distribution conditioned on the two different label states. The model hence learns the relations between the two different label states.

At inference time, since there is no prior label knowledge for the test data, we set the input label state as unknown for each class.

Discussions

One intuitive reason for introducing the label state is to provide the prior label observation to the model. It acts like Masked Language Modeling (MLM) (Kenton and Toutanova, 2019). In MLM, the model is trained to predict missing words from the given language context. As a result, it can learn the correlation between the words from the unlabeled training corpus. Our CCR framework works in a similar way for learning correlation between the classes, but differs from it in two aspects: 1) MLM typically learns from the unlabeled corpus, since the word correlation exists in every language context (e.g., a sentence or a paragraph). While our CCR learns correlation from the labels or pseudo-labels; 2) There are only three states (i.e., positive, negative, or unknown) for our CCR, while there are tens of thousands kinds of word embeddings in MLM. Therefore, our CCR does not need massive training data to learn the correlation like the MLM.

3.4 Training Objective

The proposed Conditional Consistency Regularization (CCR) contains two loss terms: L_l for labeled examples and L_u for unlabeled examples. Specifically, L_l is a standard Binary Cross Entropy (BCE) constraint (Lanchantin et al., 2021), i.e.,

$$L_l = \frac{1}{B_i} \sum_{c=1}^C BCE(f_{\theta}(\alpha(x_i)|s_i), y_i), \quad (3.1)$$

where BCE represents the binary cross-entropy loss function, s_i means the given label state for image x_i , C is the number of classes, and B_i is the batch size for labeled examples. Note that, since the label state involves a part of ground truth labels, calculating loss on these known classes is meaningless. Therefore, we discard the loss of these known classes and only accumulate the loss of unknown classes.

For the unlabeled examples, we use the Mean Square Error (MSE) loss function to constrain the distance between two model predictions that are obtained from different views and label states. Formally,

$$L_u = w(t) \frac{1}{B_j} \sum_{c=1}^C MSE(f_{\theta}(\alpha(x_j)|s_{j,w}), f_{\theta}(\Lambda(x_j)|s_{j,s})) \quad (3.2)$$

where $w(t)$ is a time-dependent weighting function that balances the loss weight for labeled and unlabeled examples, t indicates the current epoch, and B_j is the batch size for unlabeled examples. Similar to the BCE loss for the labeled examples, we only accumulate the loss of the classes that exist at least one unknown states in $s_{j,w}$ and $s_{j,s}$. Hence, the model learns to predict the unknown class from the observed known classes. In total, our loss can be formulated as $L = L_l + L_u$.

In addition, we empirically find that the weighting function $w(t)$, which decides the balance between the terms L_l and L_u , is important for the performance. We make it slowly increase from 0 to 1 in the early training stage (Samuli and Timo, 2017). Formally, we have

$$w(t) = \begin{cases} \exp\{-5[1 - (t/T)]^2\}, & t \leq T \\ 1, & t > T, \end{cases} \quad (3.3)$$

where t indicates the epoch number, T is a time threshold. Consequently, the loss is dominated by the labeled examples at the beginning, and gradually achieves a balance between the labeled examples and unlabeled examples during the training process. In this way, we can alleviate the unreliable and unstable predictions on unlabeled examples due to insufficient training. We also investigate the influence of T in the Section 4, experiments.

Discussion

It should be noted that the weighting function is commonly used in many SS-SLC methods, in the thesis we follow (Samuli and Timo, 2017). Both (Samuli and Timo, 2017) and I found that the gradual ramp-up of the unsupervised loss is important to the final performance, unsupervised loss introduced too fast or too slow will degrade model performance. As for the final weight after the t reach the threshold T , we also follow the (Samuli and Timo, 2017) which has been verified to be effective. It should be noted that this weight only represents the weight between supervised loss and unsupervised loss. Because the number of unlabeled samples is 5 times more than the number of labeled samples in every batch, as we return the average loss of each sample, every single labeled sample actually has more weight than every single unlabeled sample. The formal formulas and ablation study of the number of labeled samples and unlabeled samples are carefully discussed in Algorithm 1, line 8 & line 9 and table 4.8.

3.5 Comparison with the methods in SS-SLC

Recently, a large number of methods in SS-SLC have emerged such as FixMatch (Sohn et al., 2020), and temporal ensembling (Samuli and Timo, 2017). Compared with these methods in SS-SLC, the differences of the proposed method could be summarized in three aspects below:

- The proposed method targets SS-MLC, while these methods targets SS-SLC that can not be directly applied to our SS-MLC task.

- The proposed methods can utilize and learn label relations between multiple labels to enhance classification performance. It is well-known that the learning and use of label relations are crucial factors in MLC.
- We present experiment comparisons with some extensions of advanced semi-supervised single-label methods in experiments. CR is the extension of (Samuli and Timo, 2017) and CR-BCE is the extension of (Sohn et al., 2020). As we can see, our method surpasses these methods in most cases due to the acquirement of label relations.

Experiments

This chapter describes the experimental part in detail. There are 3 main sections in this chapter. The experimental setup are introduced in section 4.1, which explains the datasets used in experiments, implementation details of the CCR and evaluation metrics. The comparison with the advanced methods in semi-supervised multi-label classification and multi-label classification area are introduced in section 4.2. The ablation study discusses the effectiveness of every part of the CCR, including conditional consistency regularization, different model architectures and examples of prediction, in section 4.3. And finally the thesis also discusses the influence of the choices of hyper-parameters on the proposed method, including weighting function, the ratio between labelled and unlabeled samples and threshold, in section 4.4.

Finally, the thesis talks about the ablation study which verifies the effectiveness of every part of the CCR, including conditional consistency regularization, different model architectures, examples of predictions and the influence of the choices of hyper-parameters on the proposed method, including weighting function, the ratio between labelled and unlabeled samples and threshold.

4.1 Experimental Setup

Datasets

We use two large-scale real-world MLC datasets VOC-2007 (Everingham et al., 2015) and COCO-80 (Lin et al., 2014) to evaluate our method. VOC-2007 is a commonly used dataset for multi-label classification, object detection, and segmentation. It contains 9,963 images in realistic scenes, including 20 classes (e.g., person, bird, cat, and etc). We split about 50% data for training (5,011 images) and 50% for testing (4,952 images). COCO-80 contains 80 classes. We utilize 82,783 images as a training

set and evaluate all methods on a testing set consisting of 40,504 images. We randomly select a specific ratio and sample images from the training set with this ratio as unlabeled examples.

Implementation details

For both VOC-2007 and COCO-80, following (Wu et al., 2015), each image is resized to 256×256 pixels and randomly cropped to 224×224 pixels. We use pre-trained ResNet-101 (He et al., 2016) on ImageNet as a feature extractor following previous work (Lanchantin et al., 2021) for all methods. By default, following (Lanchantin et al., 2021), we employ a module with three-layer self-attention blocks (Vaswani et al., 2017) to output the prediction. For test images, the image is centred cropped. We conduct experiments on 5%, 10%, 20%, and 50% labeled examples for VOC-2007, and 1%, 3%, 5%, and 10% labeled examples for COCO-80 following (Chu et al., 2018). We set batch size as 16 for labeled examples and 80 for unlabeled examples. The learning rate is set to 10^{-5} . We use Adam optimizer for training. The hyper-parameters r_u and r_l are consistently set to 0.7 and 0.3.

Evaluation metrics

Following the previous works (Chen et al., 2019b; Lanchantin et al., 2021), we report three metrics to evaluate the performance for all methods: (1) the average per-class F1 score (CF1), where $CF1 = \frac{2CP * CR}{CP + CR}$. Here, CP is the average per-class precision, and CR is the average per-class recall. (2) the average overall F1 score (OF1), where $OF1 = \frac{2OP * OR}{OP + OR}$. Here, OP is the average overall precision, and OR is the average overall recall. (3) the mean average precision (mAP).

4.2 Comparison with State-of-the-Arts

For COCO-80, we compare the proposed CCR with multiple advanced methods, including DRS (Wang et al., 2021), West (Wu et al., 2015), DSGM (Chu et al., 2018), and DGM-Native (Chu et al., 2018). The results of West, DSGM, and DGM-Native refer to the numbers that are reported in literature (Chu et al., 2018). For VOC-2007, we take DRS (Wang et al., 2021) and COINs (Zhan and Zhang, 2017) as our baselines.

For fair comparison, we use the same training strategy for both baselines and our method that mentioned above. We use a pre-trained ResNet-101 (He et al., 2016) as the backbone to extract the features for all the methods.

Following (Chu et al., 2018), we randomly sample 1%, 3%, 5%, and 10% of training set as labeled examples for COCO-80. The detailed experimental results are shown in Tab. 4.1. We can observe that, for all ratios and all evaluation metrics, our proposed CCR surpasses the state-of-the-arts by a large margin. Specifically, our method achieves about 10% leads in CF1 and OF1, and more than 15% lead in mAP compared with the state-of-the-arts when the labeled ratio is low, e.g., 1% and 3%. With the increase of the ratio of labeled examples, the proposed method still surpasses state-of-the-arts clearly in CF1, OF1, and mAP.

For the experiments on VOC-2007, we randomly sample 5%, 10%, 20%, and 50% of a training set as labeled examples following the baseline (Chu et al., 2018). We demonstrate the results in Tab. 4.2. As we can see in the experiments, the performance of our method also has a significant increase over the state-of-the-arts in all labeled ratios and all evaluation metrics. To be specific, our method reaches 84.0% in CF1 and 86.2% in OF1 when the labeled ratio is relatively high at 50%. If we set the labeled ratio as 20%, for example, our method exceeds COINs (Zhan and Zhang, 2017) by 7.6% in CF1, 8.4% in OF1, and 7.6% in mAP. With the decrease in the labeled ratio, the task becomes more challenging. The proposed method still surpasses the state-of-arts (the best performance achieved by baselines) by 4.3% in CF1, 6.1% in OF1, and 9.3% in mAP when the labeled ratio is 10%.

TABLE 4.1: Performance comparison on COCO-80 dataset. The best results (%) are bolded.

Method	1% labeled			3% labeled		
	CF1	OF1	mAP	CF1	OF1	mAP
West (Wu et al., 2015)*	2.8	5.5	-	2.1	16.7	-
DSGM (Chu et al., 2018)*	36.1	48.3	-	41.1	52.2	-
DGM-Naive (Chu et al., 2018)*	42.5	49.9	-	42.6	50.1	-
DRS (Wang et al., 2021)	39.9	52.2	36.8	51.3	58.8	46.7
CCR (Ours)	51.9	60.0	54.4	60.2	66.0	62.5
Method	5% labeled			10% labeled		
	CF1	OF1	mAP	CF1	OF1	mAP
West (Wu et al., 2015)*	2.9	18.6	-	-	-	-
DSGM (Chu et al., 2018)*	42.9	53.8	-	-	-	-
DGM-Naive (Chu et al., 2018)*	45.1	51.1	-	-	-	-
DRS (Wang et al., 2021)	52.4	60.2	47.7	53.9	61.6	50.0
CCR (Ours)	62.9	67.0	65.7	64.8	69.1	68.6

The method with '*' is reported by (Chu et al., 2018)

In addition to the methods of semi-supervised multi-label classification, we also include a state-of-art multi-label classification method (Lanchantin et al., 2021) under the fully supervised setting (with

TABLE 4.2: Performance comparison on VOC-2007 dataset. The best results (%) are bolded.

Method	5% labeled			10% labeled		
	CF1	OF1	mAP	CF1	OF1	mAP
COINs (Zhan and Zhang, 2017)	60.9	65.7	66.3	69.0	71.7	75.6
DRS (Wang et al., 2021)	68.2	72.4	62.9	74.1	76.4	69.6
CCR (Ours)	69.1	76.6	81.4	78.4	82.5	84.9
Method	20% labeled			50% labeled		
	CF1	OF1	mAP	CF1	OF1	mAP
COINs (Zhan and Zhang, 2017)	74.2	76.3	80.0	76.6	78.3	82.9
DRS (Wang et al., 2021)	74.8	76.9	72.0	79.8	81.6	75.9
CCR (Ours)	81.8	84.7	87.6	84.0	86.2	89.6

100% labeled examples) on the VOC-2007 dataset for completeness of experiments. As we can see in Tab. 4.3, our method surpasses CTRANS (with 100% labeled examples) in all evaluation matrices with only 50% labeled examples on the VOC-2007 dataset. Moreover, we also provide CCR (with 100% labeled examples) performance, which uses labeled examples to calculate the unsupervised loss. Our method leads to improvements of 0.3% in CF1, 0.5% in OF1, and 1.1% in mAP. The results shows that the proposed method-CCR’s effectiveness and feasibility.

TABLE 4.3: Comparison with advanced MLC methods on VOC-2007 dataset. The best results (%) are bolded.

	CF1	OF1	mAP
CTTRANS(Lanchantin et al., 2021) (100% labeled)	83.9	86.0	89.5
CCR (50% labeled)	84.0	86.2	89.6
CCR (100% labeled)	84.2	86.5	90.6

4.3 Ablation Study

In this section, we conducted several experiments to verify the effectiveness of each part, including consistent regularization and network architecture.

Consistency regularization matters for SS-MLC

As mentioned in the introduction, we are the first to introduce consistency regularization into the SS-MLC. The consistency regularization imposes a constraint on the model, making the training data harder

to memorize. Therefore, the model will be more robust when generalizing to unseen data. Here, we construct the conventional consistency regularization on the same model architecture and training strategy of our method in order to verify the effectiveness of CCR, the proposed method. In detail, we implement a multi-view image consistency regularization, which is similar to (Samuli and Timo, 2017) (denoted as CR) as described in Section 3.2. The differences between CR and CCR lie in: (1) During training, CR uses unknown states for every unlabeled example, while CCR uses randomly generated label states. For the labeled examples, we employ label states in a similar manner for both CR and CCR for a fair comparison. (2) CR removes the pseudo-label memory module since it uses unknown states for every example instead of cached pseudo-labels. Note that, the weakly and strongly augmented operations are the same for CR and CCR. We demonstrate the results of CR on COCO-80 and VOC-2007 in Tab. 4.4 and Tab. 4.5 in terms of CF1, OF1, and mAP.

The results tell that the performance of CR is lower than our proposed CCR, but it outperforms all the state-of-the-art methods. For example, when using 5% labeled examples on COCO-80, CR achieves 55.7%, 64.4%, and 65.0% on CF1, OF1, and mAP respectively, while DRS (Wang et al., 2021) only obtains 52.4%, 60.2%, and 47.7%.

Furthermore, we implement a variation by replacing the MSE loss with BCE loss (denoted as CR-BCE). Specifically, there are three differences between CCR and CR-BCE: (1) For unlabeled examples, CR-BCE generates pseudo-labels for the prediction of the weakly-augmented image with $r_u = 0.7$ and $r_l = 0.3$, and then calculates the loss by using BCE between the pseudo-label and the prediction of the strongly-augmented image, which is similar to FixMatch (Sohn et al., 2020). (2) CR-BCE also uses unknown states for every example like CR. (3) CR-BCE removes the pseudo-label memory module. From Tab. 4.4 and Tab. 4.5, we can observe that CR-BCE realizes slightly better performance than CR in most cases. Actually, we have also implemented CCR with BCE loss for unlabeled examples, but have not observed performance gain. Therefore, we use MSE for simplicity. Compared with Tab. 4.4 and Tab. 4.5, the performance gains in all evaluation metrics on the VOC-2007 dataset are slightly lower than the performance gains on the COCO-80 dataset. We conclude that the performance gains come from the label relations knowledge that learns via the label state. While, the VOC-2007 dataset has a smaller size of classes of labels (only 20 possible classes for each image) compared with the COCO-80 dataset (80 possible classes for each image), resulting in fewer label relations. In a nutshell, the aforementioned experiments mean that consistency regularization can boost the performance of SS-MLC.

TABLE 4.4: Ablation study on the consistency regularization on COCO-80 dataset. The best results (%) are bolded.

	1% labeled			3% labeled		
	CF1	OF1	mAP	CF1	OF1	mAP
CR	47.1	58.1	53.7	58.2	64.8	62.0
CR-BCE	49.5	59.5	52.9	59.5	65.6	61.8
CCR (Ours)	51.9	60.0	54.4	60.2	66.0	62.5
	5% labeled			10% labeled		
	CF1	OF1	mAP	CF1	OF1	mAP
CR	55.7	64.4	65.0	45.7	57.7	67.8
CR-BCE	62.1	67.4	64.8	64.7	69.3	68.5
CCR (Ours)	62.9	67.0	65.7	64.8	69.1	68.6

TABLE 4.5: Ablation study on the consistency regularization on VOC-2007 dataset. The best results (%) are bolded.

	5% labeled			10% labeled		
	CF1	OF1	mAP	CF1	OF1	mAP
CR	64.4	74.6	80.9	75.1	80.6	84.6
CR-BCE	67.5	76.5	80.3	78.0	82.4	84.6
CCR (Ours)	69.1	76.6	81.4	78.4	82.5	84.9
	20% labeled			50% labeled		
	CF1	OF1	mAP	CF1	OF1	mAP
CR	77.4	82.1	87.3	83.2	86.0	89.8
CR-BCE	80.8	84.3	87.4	83.4	86.1	88.9
CCR (Ours)	81.8	84.7	87.6	84.0	86.2	89.6

It should be noted that, for the labeled examples, CR and CR-BCE still use the randomly generated label state, therefore, they are able to learn the label relations from the labeled examples. In this way, the performance gain of our proposed CCR mainly comes from the additional knowledge of the label relations in the unlabeled examples. The experimental results also verify this statement in that, when the ratio of labeled examples is low, the performance gain is larger than the higher ratio of the labeled examples.

Examples of predictions




TABLE 4.6: Ablation study on various model architectures.

	5% labeled			10% labeled		
Method	CF1	OF1	mAP	CF1	OF1	mAP
CCR-FC	72.7	76.1	77.4	75.9	79.5	81.9
CCR (Ours)	69.1	76.6	81.4	78.4	82.5	84.9
	20% labeled			50% labeled		
Method	CF1	OF1	mAP	CF1	OF1	mAP
CCR-FC	79.8	82.8	85.7	81.9	84.6	88.1
CCR (Ours)	81.8	84.7	87.6	84.0	86.2	89.6

In order to visualize the CCR method learned the label relations, we provide some image prediction examples of the COCO-80 dataset and the VOC-2007 dataset. As we can see in Fig. 4.1, our method’s predictions are more accurate and reflect the correlation between labels. For example, our method could predict more kitchenware labels based on the label relation in the first image. Even some labels that are not in the ground truth labels but appear in the image are also predicted, like `Bowl` in the first image. Similar to the COCO-80 dataset, we also provide some examples of the VOC-2007 dataset. As we can see, our predictions have stronger label relation compared with the baseline methods. For example, though `Person` is very hard to predict in the first image, our method still gives the `Person` prediction based on the label relation, because the car usually appears with a person. Moreover, our method could give an accurate prediction of the third image, because bicycles usually appear with a person.

Conditional consistency regularization works on various model architectures

Since our proposed method CCR is independent of the model architecture, we also implement a simple variation by using a 5-layer fully-connected network after the ResNet backbone (denoted by CCR-FC), where the fully-connected network is responsible for outputting the prediction. We conduct the experiments on VOC-2007 with 5%, 10%, 20%, and 50% ratios of labeled examples, and demonstrate the results in Tab. 4.6. In Tab. 4.6, CCR-FC is slightly inferior to CCR, indicating that the self-attention block has stronger abilities on learning features and label relations. We also find that CCR-FC still outperforms the state-of-the-art that are illustrated in Tab. 4.2, even with the simplest model architecture. This phenomenon strongly verifies that our CCR framework works on various model architectures.

			
Truth	Person, Handbag, Oven, Refrigerator, Clock	Person, Tie	Person, Bus, Backpack, Handbag
DRS	Person, Microwave	Person	Person, Car, Bus
Ours	Person, Bowl, Oven, Refrigerator, Clock	Person, Tie	Person, Car, Bus, Backpack, Handbag




			
Truth	Car, Person	Bottle, Person	Bicycle, Person
DRS	Car	Bottle	Person
COINS	Car	Bottle	None
Ours	Car, Person	Bottle, Person	Bicycle, Person

FIGURE 4.1: Illustrations of the examples of the predictions by CCR.

4.4 The Choices of Hyper-parameters

In this section, we investigate the hyper-parameters of the CCR, including the weighting function $w(\cdot)$, the value of B_j , and the values of thresholds r_u and r_l .

The weighting function $w(\cdot)$ affects the performance

As defined in Section 3.4, the weighting function $w(\cdot)$ is a time-dependent function that slowly increases from 0 to 1 by the T -th epoch, and is fixed to 1 after the T -th epoch. It provides the balance of the losses for labeled examples and unlabeled examples. To study the influence of $w(\cdot)$, we adjust T to

TABLE 4.7: Ablation study on the weighting function $w(\cdot)$. The best results (%) are bolded.

labeled Ratio \ T	CF1				OF1				mAP			
	0	10	30	50	0	10	30	50	0	10	30	50
5%	68.2	69.1	69.2	71.2	76.5	76.6	76.9	77.8	80.9	81.4	81.2	80.9
10%	79.8	78.4	78.0	78.9	82.9	82.5	82.4	82.6	85.4	84.9	84.6	84.5
50%	83.3	84.0	83.4	83.5	85.0	86.2	85.6	86.1	89.6	89.6	88.7	88.7

TABLE 4.8: Ablation study on the ratio between labeled examples and unlabeled examples for each training iteration. The best results (%) are bolded.

labeled Ratio \ B_j	CF1				OF1				mAP			
	16	48	80	112	16	48	80	112	16	48	80	112
5%	70.3	69.4	69.1	69.0	76.7	76.8	76.6	76.6	81.2	81.0	81.4	80.5
10%	77.3	79.3	78.4	78.8	82.0	82.6	82.5	82.4	84.8	84.9	84.9	85.1
50%	83.5	83.4	84.0	83.7	85.2	85.5	86.2	85.6	89.5	89.4	89.6	89.6

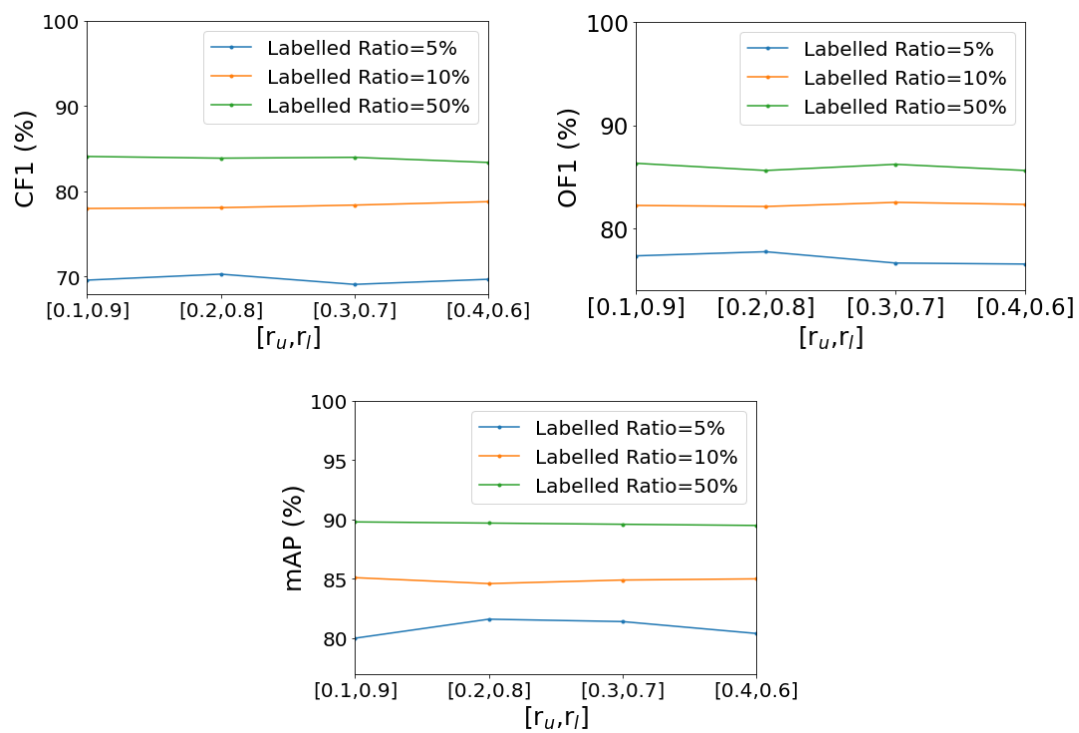
different numbers and report the results in Tab. 4.7. As we can see in the table, when the ratio of labeled examples is low, like 5%, the model results in better performance when T is larger. In the cases of the higher ratios, the model results in better performance when T is smaller. It is reasonable since the model may generate unreliable pseudo-label for the higher ratio of the unlabeled examples, the larger impact of these unreliable pseudo-label leads to larger noise in the optimization. In experiments, we set $T = 10$ by default.

The ratio between labeled examples and unlabeled examples matters

By default, we fix the batch size B_i for labeled examples as 16. In this paragraph, we adjust the batch size B_j for unlabeled examples. As shown in the Tab. 4.8, the model achieves the best performance when B_j is slightly larger than B_i . We argue that a small batch size for unlabeled examples limits the regularization ability of consistency. In experiments, we set B_j as 80.

On the thresholds r_u and r_l

As introduced in Section 3.3, the thresholds r_u and r_l are used in the pseudo-label generation function, which determines how we obtain the pseudo-label in the pseudo-label memory. We investigate different values for r_u and r_l in Fig. 4.2. As we can see, the performance is stable for different thresholds, which shows that our method is robust to these hyper-parameters.

FIGURE 4.2: Ablation study on the thresholds r_u and r_l .

Discussion

In this chapter, we will discuss the CCR, the proposed method, compare it with the previous method, and analyze why it is effective, the existing problems and the significance of this work. In summary, this chapter will critically analyze the proposed method.

5.1 The proposed method

As mentioned above, this work tries to resolve a practical problem that there are large numbers of unlabeled samples and small numbers of labeled samples in multi-label classification (MLC) due to the difficulties in the labelling process. It is also known as semi-supervised multi-label classification (SS-MLC). In order to solve this problem, Conditional Consistency Regularization (CCR) has been proposed and conducting comprehensive experiments with many baselines and ablation studies. The thesis introduces many advanced baselines from different aspects, like advanced methods in SS-MLC (Wang et al., 2021), MLC-CTRANS (Lanchantin et al., 2021) and extension of semi-supervised single-label classification (SS-SLC)-FixMatch(CR-BCE)(Sohn et al., 2020), Temporal ensembling(CR) (Samuli and Timo, 2017). According to the experiments, the proposed methods have better performance than baseline methods in most cases. These situations verify that the proposed method not only has outstanding performance in SS-MLC but also surpass some state-of-art MLC method by using 50% less labeled samples. Although some extensions of SS-SLC method have similar performance in some cases, like 50% labeled samples in COCO-80. It is possible that with the increase of labeled data, the relationships between labels will not bring much gain. Moreover, the success of CCR shows a feasible way of introducing advanced SS-SLC methods and MLC methods to SS-MLC. It improves classification performance by taking full advantage of leveraging unlabeled samples and modeling label relations. These performance of the experiments show that the CCR has a strong ability in addressing SS-MLC tasks.

5.2 Comparison with previous work

Compared with the previous method, the previous work focus on only MLC and SS-SLC. For instance, some researchers proposed many impressive methods in MLC, like conditional prediction (Cheng et al., 2010; Read et al., 2011), shared embedding space (Bhatia et al., 2015; Yeh et al., 2017), label graph formulation (Chen et al., 2019b, 2021), etc. Other researchers pay attention to SS-SLC, like self-training (Lee et al., 2013; Xie et al., 2020b), consistency regularization (Samuli and Timo, 2017; Xie et al., 2020a), graph-based method (Wang et al., 2020) and hybrid methods (Sohn et al., 2020; Berthelot et al., 2019b,a). Even though these methods perform well in their area, the previous methods are difficult to suit SS-MLC scenarios. It is clear that the MLC methods lack the strategy of leveraging the unlabeled samples which are unique and important in the semi-supervised classification area. And the proposed methods in SS-SLC lack the way of learning label relations which are seen as the key factors in the MLC area. Although few methods in SS-MLC also emerged, they may have some weaknesses like being unable to generalize to unseen test samples, out-of-date learning strategy, low performance, etc. However, SS-MLC scenarios have widely existed in practical challenges, which greatly limit the development of MLC and the performance of learning models. The proposed method fills the gap in this area via providing a way of introducing advanced SS-SLC methods and retaining key factors of MLC simultaneously. As a result of that, the proposed method can leverage unlabeled samples and learning label relations at the same time, which directly improves the utilizing efficiency of unlabeled samples in MLC and reduce the cost of labelling and learning process. In addition, the ablation studies on different model architectures demonstrate that the proposed methods' flexibility and expandability on the different models. In conclusion, the proposed methods provide a novel and effective way of combining MLC and SS-SLC methods.

5.3 Open problems of the proposed method

While there are many advantages and contributions in this work, there are some open problems and weaknesses that may appear in practical tasks. The section below will discuss 3 major problems that the author thinks out, including the limitation of labeled samples, the limitation of unlabeled samples and other methods of MLC with limited supervision.

5.3.1 Limitation of labeled samples

This subsection will discuss the labeled samples' impact on the proposed model in practical situations. There are 3 major limitations for labeled samples in real-world challenges, including unreliable labels, irrelevant multi-label and unadjustable number of labels.

Unreliable labels

Due to the lack of labeled samples in training samples, the proposed methods are very dependent on the reliability of labeled samples, especially at the beginning of the training process. If noisy labels or unreliable labels appear in the labeled samples which is quite common in the practical dataset, the model performance may strongly degrade or even have worsened performance than the unsupervised learning method. As a result of that, it is necessary to keep the reliability of labelled samples. Currently, the proposed method does not have any potential solutions for noisy labels, we may regard this problem as future work.

Irrelevant multi-labels

As mentioned above, the proposed method could have an increase in classification performance by leveraging the knowledge of unlabeled samples and modeling relationships between labels. If each given training sample has fewer corresponding labels (eg. 2-3 labels) or given labels do not have an obvious relationship, which may appear in the real-world task. The performance of the CCR will degrade to the simple extension of SS-SLC. While, if there are no unlabelled samples, the proposed method also achieves better performance than MLC methods via utilizing conditional consistency regularization, which provides better learning features and knowledge of label relations. The experimental results could be found in table 4.3.

Unadjustable number of labels

As mentioned in section 4.1, the goal of SS-MLC is to learn a feature embedding network (commonly neural network) $f_{\theta}(\cdot)$ from N_u unlabeled images $D_u = \{(x_j)\}_{j=1}^{N_u}$ and N_l labeled images $D_l = \{(x_i, y_i)\}_{i=1}^{N_l}$. For each image x_i , $y_i \in \{0, 1\}^C$ is the corresponding one-hot label, where C is the number of classes. We define $y[c] = 1$ if the image is associated with the c -th label, otherwise $y[c] = 0$. The proposed method needs to fix the number of potential labels, by that I mean, the C should be fixed by following the previous MLC methods (Lanchantin et al., 2021; Yeh et al., 2017). This situation raises the question that the proposed method is unable to add extra new labels that appear in unlabeled samples or in the new dataset. Namely, the number of labels may dynamically expand. We only can utilize the

labels we have already defined before training. This situation is also widely known as MLC with unseen labels(Liu et al., 2021).

5.3.2 Limitation of unlabeled samples

This subsection will discuss the unlabeled samples' impact on the proposed method in practical situations. There are 2 major limitation in unlabeled samples, including out-of-distribution samples and unreliable pseudo-labels cache in label state.

Out-of-distribution samples

According to many previous work(Wei et al., 2021; Guo and Li, 2022), there might be out-of-distribution samples in unlabelled samples, where the sample does not contain the corresponding objects of the providing labels. This situation may also decrease the performance of the model because the unlabeled samples may assign to the wrong class. Currently, the proposed method does not have solutions for out-of-distribution samples, we may set dealing with out-of-distribution samples as future work to better help the proposed method suit the practical situations.

Unreliable pseudo-labels cache in label state

Although the proposed methods leverage many measures to avoid unreliable pseudo-labels cache, like choosing high confidence samples, randomly masking the label state and reducing the weight of unsupervised loss at the beginning of training, it still may have wrong pseudo-labels cache. The wrong pseudo-labels may have a negative impact on learning models. There are many reasons for unreliable pseudo-label cache, like unwell-learned training models, unreliable labeled samples, etc.

5.3.3 other methods of multi-label classification with limited supervision

Expect SS-MLC scenarios that dataset contains completed labeled samples and unlabeled samples, there are many other multi-label with limited supervision situations(Liu et al., 2021). For examples, MLC with missing labels, which means the annotators only assign parts of labels in every sample, namely, there are incomplete labels in each sample; Weakly MLC, which means that there are fully labelled samples, incompletely-labelled samples and unlabelled samples in the dataset at the same time. These situations may also appear in practical tasks. The proposed methods currently do not have a solution to deal with these kinds of scenarios. The author may extend the proposed method to fit these practical situations.

5.4 Significance of the work

The significance of the proposed method is concluded below:

- The thesis is the first to extend consistency regularization which is an advance SS-SLC method to SS-MLC.
- The thesis is the first to propose a novel algorithm (CCR) that could boost classification performance via learning knowledge of unlabeled samples and label relationships. Normally, previous methods, like methods in semi-supervised classification and MLC only consider one aspect.
- The CCR provides a novel way to take full advantage of both SS-SLC and MLC methods by acquiring the knowledge of unlabeled samples and label relations, which are crucial to SS-SLC and MLC.
- Because SS-MLC has commonly existed in the practical task, the proposed novel algorithm helps the MLC method suit the real-world task and application and reduce the cost of labelling process.
- The thesis also presents many analyses, discussions, and insights into multi-label/single-label classification with limited supervision which may inspire researchers and students to have a better understanding of this area and encourage them to conduct in-depth research.

Conclusion and future work

This chapter will summarize the research problems, proposed methods, experiments, and discuss the necessary and meaningful work that may be carried out in the future.

6.1 Conclusion

The thesis proposed a novel method named Conditional Consistency Regularization (CCR) to resolve the problem in semi-supervised multi-label classification (SS-MLC). SS-MLC is a practical scenario that the training set has large numbers of unlabeled samples and small numbers of labelled samples due to the high-cost, time-consuming and professional labelling process, especially for the multi-label dataset that has more than one label for each sample. Normally, Previous methods only consider one aspect, like methods in semi-supervised single-label classification (SS-SLC) and methods in multi-label classification (MLC) that they can not extend to SS-MLC. Specific background knowledge of semi-supervised classification and MLC is introduced in chapter 1 (introduction) and chapter 2 (literature review).

In order to solve the problems, the thesis proposed a novel algorithm named Conditional Consistency Regularization by introducing consistency regularization which is an advanced technology in SS-SLC. Specifically, CCR utilizes consistency regularization and conditional label state to implicitly learning the knowledge of unlabeled samples and modeling the relationship between labels, which are verified to be pivotal in SS-SLC and MLC. Detailed methods, like method overview, algorithms, general framework, objective function, etc, are carefully explained in chapter 3. Comprehensive experiments on different datasets, baseline methods and ablation studies that show the effectiveness of the proposed method (CCR) describe in chapter 4(experiments) in detail. The proposed methods not only surpass all the baseline methods in SS-MLC, but also surpass state-of-art in MLC via only using half of labelled samples.

Finally, a comprehensive and critical discussion and analysis, like advantages, weakness, significance, reflection and summary, are explained in chapter 5.

To sum up, this thesis proposes Conditional Consistency Regularization (CCR) to deal with the SS-MLC. CCR is the first method that introduces the consistency regularization into SS-MLC. It is built upon the conventional consistency regularization that regularizes model predictions to be invariant to different augmented views of the same input image. In addition, to learn label relations for MLC, CCR leverages the label states that act as a condition for model training. By minimizing the distance between the two model predictions of two augmented images, the outputs that are obtained from different label states, are encouraged to be consistent. As a result, the model learns knowledge of unlabelled samples and the label relations to increase the performance.

6.2 Future work

Based on the experiments and discussion above, it is necessary and meaningful to conduct some extension based on this work in order to make the proposed method have better application in the real-world task, which could be regarded as future work. For example, dealing with out-of-distribution samples. It is necessary and meaningful to conduct further research to discover potential value in academics and applications. The words below demonstrate the future work that may extend from the proposed method-CCR:

- Create a standard dataset warehouse from real-world data and set the criterion to evaluate the samples from the dataset by their labels. To be specific, we can divide the samples into fully labelled samples, incomplete labelled samples, unlabelled samples and samples with noisy labels. These will help the academics and engineers obtain desired data before training the model.
- Extend the proposed method to make it suit other weakly-supervised multi-label classification scenarios, like the dataset with fully-labelled, incompletely-labelled and unlabelled samples. This extension will help the proposed method better fit the practical work.
- Extend the proposed method to adapt the out-of-distribution samples, which commonly existed in unlabelled datasets.
- Extend the proposed method to address dynamically expanding of the number of labels, which is also known as MLC with unseen labels.

- In the future, the author encourage other researchers to introduce other advanced semi-supervised classification methods and MLC method to SS-MLC in order to propose a better solution to deal with this practical scenario.

Bibliography

- Mahmoud Assran, Mathilde Caron, Ishan Misra, Piotr Bojanowski, Armand Joulin, Nicolas Ballas, and Michael Rabbat. 2021. Semi-supervised learning of visual features by non-parametrically predicting view assignments with support samples. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8443–8452.
- Mikhail Belkin and Partha Niyogi. 2001. Laplacian eigenmaps and spectral techniques for embedding and clustering. *Advances in Neural Information Processing Systems*, 14.
- David Berthelot, Nicholas Carlini, Ekin D Cubuk, Alex Kurakin, Kihyuk Sohn, Han Zhang, and Colin Raffel. 2019a. Remixmatch: Semi-supervised learning with distribution matching and augmentation anchoring. In *International Conference on Learning Representations*.
- David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin A Raffel. 2019b. Mixmatch: A holistic approach to semi-supervised learning. *Advances in Neural Information Processing Systems*, 32.
- Kush Bhatia, Himanshu Jain, Purushottam Kar, Manik Varma, and Prateek Jain. 2015. Sparse local embeddings for extreme multi-label classification. *Advances in Neural Information Processing Systems*, 28.
- Avrim Blum and Tom Mitchell. 1998. Combining labeled and unlabeled data with co-training. In *Proceedings of the Conference on Computational Learning Theory*, pages 92–100.
- Paola Cascante-Bonilla, Fuwen Tan, Yanjun Qi, and Vicente Ordonez. 2021. Curriculum labeling: Revisiting pseudo-labeling for semi-supervised learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 6912–6920.
- Tianshui Chen, Muxin Xu, Xiaolu Hui, Hefeng Wu, and Liang Lin. 2019a. Learning semantic-specific graph representation for multi-label image recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 522–531.
- Zhao-Min Chen, Xiu-Shen Wei, Peng Wang, and Yanwen Guo. 2019b. Multi-label image recognition with graph convolutional networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5177–5186.
- Zhaomin Chen, Xiu-Shen Wei, Peng Wang, and Yanwen Guo. 2021. Learning graph convolutional networks for multi-label recognition and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Weiwei Cheng, Eyke Hüllermeier, and Krzysztof J Dembczynski. 2010. Bayes optimal multilabel classification via probabilistic classifier chains. In *International Conference on Machine Learning*, pages 279–286.

- Hong-Min Chu, Chih-Kuan Yeh, and Yu-Chiang Frank Wang. 2018. Deep generative models for weakly-supervised multi-label classification. In *Proceedings of the European Conference on Computer Vision*, pages 400–415.
- Elijah Cole, Oisín Mac Aodha, Titouan Lorieul, Pietro Perona, Dan Morris, and Nebojsa Jojic. 2021. Multi-label learning from single positive labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 933–942.
- Ekin D Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V Le. 2019. Autoaugment: Learning augmentation strategies from data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 113–123.
- Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. 2020. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 702–703.
- Zihang Dai, Zhilin Yang, Fan Yang, William W Cohen, and Russ R Salakhutdinov. 2017. Good semi-supervised learning that requires a bad gan. *Advances in Neural Information Processing Systems*, 30.
- Krzysztof Dembczynski, Weiwei Cheng, and Eyke Hüllermeier. 2010. Bayes optimal multilabel classification via probabilistic classifier chains. In *International Conference on Machine Learning*.
- Terrance DeVries and Graham W Taylor. 2017. Improved regularization of convolutional neural networks with cutout. *arXiv preprint arXiv:1708.04552*.
- Mark Everingham, SM Eslami, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. 2015. The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111(1):98–136.
- Bin-Bin Gao and Hong-Yu Zhou. 2021. Learning to discover multi-class attentional regions for multi-label image recognition. *IEEE Transactions on Image Processing*, 30:5920–5932.
- Chen Gong, Dacheng Tao, Jie Yang, and Wei Liu. 2016. Teaching-to-learn and learning-to-teach for multi-label propagation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30.
- Yunchao Gong, Yangqing Jia, Thomas Leung, Alexander Toshev, and Sergey Ioffe. 2013. Deep convolutional ranking for multilabel image annotation. *arXiv preprint arXiv:1312.4894*.
- Lan-Zhe Guo and Yu-Feng Li. 2022. Class-imbalanced semi-supervised learning with adaptive thresholding. In *International Conference on Machine Learning*, pages 8082–8094. PMLR.
- Yuhong Guo and Suicheng Gu. 2011. Multi-label classification using conditional dependency networks. In *International Joint Conference on Artificial Intelligence*.
- Nilesh Gupta, Sakina Bohra, Yashoteja Prabhu, Saurabh Purohit, and Manik Varma. 2021. Generalized zero-shot extreme multi-label learning. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 527–535.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 770–778.

- Shu Hu, Lipeng Ke, Xin Wang, and Siwei Lyu. 2021. Tkml-ap: Adversarial attacks to top-k multi-label learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7649–7657.
- Zhuo Huang, Xiaobo Xia, Li Shen, Bo Han, Mingming Gong, Chen Gong, and Tongliang Liu. 2022. Harnessing out-of-distribution examples via augmenting content and style. *arXiv preprint arXiv:2207.03162*.
- Ahmet Iscen, Giorgos Tolias, Yannis Avrithis, and Ondrej Chum. 2019. Label propagation for deep semi-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5070–5079.
- Liping Jing, Liu Yang, Jian Yu, and Michael K Ng. 2015. Semi-supervised low-rank mapping learning for multi-label classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1483–1491.
- Jacob Devlin Ming-Wei Chang Kenton and Lee Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4171–4186.
- Durk P Kingma, Shakir Mohamed, Danilo Jimenez Rezende, and Max Welling. 2014. Semi-supervised learning with deep generative models. *Advances in Neural Information Processing Systems*, 27.
- Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.
- Xiangnan Kong, Michael K Ng, and Zhi-Hua Zhou. 2011. Transductive multilabel learning via label set propagation. *IEEE Transactions on Knowledge and Data Engineering*, 25(3):704–719.
- Abhishek Kumar, Prasanna Sattigeri, and Tom Fletcher. 2017. Semi-supervised learning with gans: Manifold invariance with improved inference. *Advances in Neural Information Processing Systems*, 30.
- Jack Lanchantin, Tianlu Wang, Vicente Ordonez, and Yanjun Qi. 2021. General multi-label image classification with transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16478–16488.
- Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature*, 521(7553):436–444.
- Dong-Hyun Lee et al. 2013. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on Challenges in Representation Learning, International Conference on Machine Learning*.
- Qiang Li, Maoying Qiao, Wei Bian, and Dacheng Tao. 2016. Conditional graphical lasso for multi-label image classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2977–2986.
- Guangfeng Lin, Kaiyang Liao, Bangyong Sun, Yajun Chen, and Fan Zhao. 2017a. Dynamic graph fusion label propagation for semi-supervised multi-modality classification. *Pattern Recognition*, 68:14–23.
- Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. 2017b. Focal loss for dense object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2980–2988.

- Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *Proceedings of the European Conference on Computer Vision*, pages 740–755.
- Qingshan Liu, Yubao Sun, Cantian Wang, Tongliang Liu, and Dacheng Tao. 2017. Elastic net hypergraph learning for image clustering and semi-supervised classification. *IEEE Transactions on Image Processing*, 26(1):452–463.
- Weiwei Liu, Haobo Wang, Xiaobo Shen, and Ivor Tsang. 2021. The emerging trends of multi-label learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Yi Liu, Guangchang Deng, Xiangping Zeng, Si Wu, Zhiwen Yu, and Hau-San Wong. 2020. Regularizing discriminative capability of cgans for semi-supervised generative learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5720–5729.
- Takeru Miyato, Shin-ichi Maeda, Masanori Koyama, and Shin Ishii. 2018. Virtual adversarial training: a regularization method for supervised and semi-supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(8):1979–1993.
- Jinseok Nam, Eneldo Loza Mencía, Hyunwoo J Kim, and Johannes Fürnkranz. 2017. Maximizing subset accuracy with recurrent neural networks in multi-label classification. *Advances in Neural Information Processing Systems*, 30.
- Siyuan Qiao, Wei Shen, Zhishuai Zhang, Bo Wang, and Alan Yuille. 2018. Deep co-training for semi-supervised image recognition. In *Proceedings of the European Conference on Computer Vision*, pages 135–152.
- Antti Rasmus, Mathias Berglund, Mikko Honkala, Harri Valpola, and Tapani Raiko. 2015. Semi-supervised learning with ladder networks. *Advances in Neural Information Processing Systems*, 28.
- Jesse Read, Bernhard Pfahringer, Geoff Holmes, and Eibe Frank. 2009. Classifier chains for multi-label classification. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 254–269.
- Jesse Read, Bernhard Pfahringer, Geoff Holmes, and Eibe Frank. 2011. Classifier chains for multi-label classification. *Machine learning*, 85(3):333–359.
- Tal Ridnik, Emanuel Ben-Baruch, Nadav Zamir, Asaf Noy, Itamar Friedman, Matan Protter, and Lihi Zelnik-Manor. 2021. Asymmetric loss for multi-label classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 82–91.
- Mehdi Sajjadi, Mehran Javanmardi, and Tolga Tasdizen. 2016. Regularization with stochastic transformations and perturbations for deep semi-supervised learning. *Advances in Neural Information Processing Systems*, 29.
- Laine Samuli and Aila Timo. 2017. Temporal ensembling for semi-supervised learning. In *International Conference on Learning Representations*.
- Karen Simonyan and Andrew Zisserman. 2015. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*.
- Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. 2020. Fixmatch: Simplifying semi-supervised learning

- with consistency and confidence. *Advances in Neural Information Processing Systems*, 33:596–608.
- Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. 2016. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2818–2826.
- Antti Tarvainen and Harri Valpola. 2017. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in Neural Information Processing Systems*, 30.
- Jesper E Van Engelen and Holger H Hoos. 2020. A survey on semi-supervised learning. *Machine Learning*, 109(2):373–440.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
- Bo Wang, Zhuowen Tu, and John K Tsotsos. 2013. Dynamic label propagation for semi-supervised multi-class multi-label classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 425–432.
- Haibo Wang, Chuan Zhou, Xin Chen, Jia Wu, Shirui Pan, and Jilong Wang. 2020. Graph stochastic neural networks for semi-supervised learning. *Advances in Neural Information Processing Systems*, 33:19839–19848.
- Jiang Wang, Yi Yang, Junhua Mao, Zhiheng Huang, Chang Huang, and Wei Xu. 2016. Cnn-rnn: A unified framework for multi-label image classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2285–2294.
- Lichen Wang, Yunyu Liu, Hang Di, Can Qin, Gan Sun, and Yun Fu. 2021. Semi-supervised dual relation learning for multi-label classification. *IEEE Transactions on Image Processing*, 30:9125–9135.
- Chen Wei, Kihyuk Sohn, Clayton Mellina, Alan Yuille, and Fan Yang. 2021. Crest: A class-rebalancing self-training framework for imbalanced semi-supervised learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10857–10866.
- Fei Wu, Zhuhao Wang, Zhongfei Zhang, Yi Yang, Jiebo Luo, Wenwu Zhu, and Yueting Zhuang. 2015. Weakly semi-supervised deep learning for multi-label image annotation. *IEEE Transactions on Big Data*, 1(3):109–122.
- Tong Wu, Qingqiu Huang, Ziwei Liu, Yu Wang, and Dahua Lin. 2020. Distribution-balanced loss for multi-label classification in long-tailed datasets. In *Proceedings of the European Conference on Computer Vision*, pages 162–178.
- Zhengning Wu, Xiaobo Xia, Ruxin Wang, Jiatong Li, Jun Yu, Yinian Mao, and Tongliang Liu. 2021. Lr-svm+: Learning using privileged information with noisy labels. *IEEE Transactions on Multimedia*, 24:1080–1092.
- Xiaobo Xia, Tongliang Liu, Bo Han, Nannan Wang, Mingming Gong, Haifeng Liu, Gang Niu, Dacheng Tao, and Masashi Sugiyama. 2020. Part-dependent label noise: Towards instance-dependent label noise. In *Advances in Neural Information Processing Systems*.

- Ming-Kun Xie and Sheng-Jun Huang. 2021. Partial multi-label learning with noisy label identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Qizhe Xie, Zihang Dai, Eduard Hovy, Thang Luong, and Quoc Le. 2020a. Unsupervised data augmentation for consistency training. *Advances in Neural Information Processing Systems*, 33:6256–6268.
- Qizhe Xie, Minh-Thang Luong, Eduard Hovy, and Quoc V Le. 2020b. Self-training with noisy student improves imagenet classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10687–10698.
- Chang Xu, Tongliang Liu, Dacheng Tao, and Chao Xu. 2016. Local rademacher complexity for multi-label learning. *IEEE Transactions on Image Processing*, 25(3):1495–1507.
- Xiangli Yang, Zixing Song, Irwin King, and Zenglin Xu. 2021. A survey on deep semi-supervised learning. *arXiv preprint arXiv:2103.00550*.
- Chih-Kuan Yeh, Wei-Chieh Wu, Wei-Jen Ko, and Yu-Chiang Frank Wang. 2017. Learning deep latent space for multi-label classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*.
- Wang Zhan and Min-Ling Zhang. 2017. Inductive semi-supervised multi-label learning with co-training. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1305–1314.
- Bowen Zhang, Yidong Wang, Wenxin Hou, Hao Wu, Jindong Wang, Manabu Okumura, and Takahiro Shinozaki. 2021. Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling. *Advances in Neural Information Processing Systems*, 34.
- Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. 2016. mixup: Beyond empirical risk minimization. In *International Conference on Learning Representations*.
- Liheng Zhang and Guo-Jun Qi. 2020. Wcp: Worst-case perturbations for semi-supervised deep learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3912–3921.
- Zhi-Hua Zhou. 2018. A brief introduction to weakly supervised learning. *National science review*, 5(1):44–53.
- Feng Zhu, Hongsheng Li, Wanli Ouyang, Nenghai Yu, and Xiaogang Wang. 2017. Learning spatial regularization with image-level supervisions for multi-label image classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5513–5522.