学 位 論 文 内 容 の 要 旨

博士の専攻分野の名称　博士（情報科学）　氏名　李 宗曜

学 位 論 文 題 名

A Study on Practicality Improvement of Image Recognition Technologies by Mitigation of Label Dependence

（ラベル依存緩和による画像認識技術の実用性向上に関する研究）

This thesis focuses on a common problem in applications of image recognition technologies, the dependence on labeled data, and summarizes the results of my studies on solutions to this problem.

In the past years, deep learning technologies have been developed rapidly and shown increasing application potential for real-world problems. In the field of computer vision, deep learning methods have made remarkable achievements, especially for the basic image recognition tasks such as image classification, object detection, and semantic segmentation. However, there is still an obstacle to the applications of deep learning to a wide range of real-world problems, the dependence on labeled data (i.e., label dependence) for training the deep learning models. Typically, the deep learning models with outstanding performance are trained based on supervised learning, thereby heavily relying on a large amount of well-labeled data which are unavailable in many real-world applications.

In some fields such as medical imaging, the annotation requires specialized knowledge and is thus difficult to obtain. Moreover, for image recognition tasks that are more complex than image-level classification, such as object detection and semantic segmentation, it may take a long time to precisely annotate an image. Constructing a large-scale dataset for such tasks is expensive and can considerably improve the costs. Therefore, it is necessary to mitigate the label dependence for applying the deep learning technologies to a wide range of real-world problems.

To mitigate the label dependence, I conduct studies leveraging unlabeled data in three directions: semi-supervised learning, unsupervised domain adaptation, and model adaptation, where the label knowledge is acquired from labeled data of the same distribution as the unlabeled data, labeled data of a different domain from the unlabeled data, and a pre-trained model of a different domain from the unlabeled data, respectively. I propose methods that solve the following problems that remained in previous studies, thereby obtaining superior performance and practicality. First, for semi-supervised learning, I apply the pseudo-label learning technology to a medical image task and tackle the limitation of the self-produced pseudo labels faced by the typical pseudo-label learning methods by introducing a tri-training architecture to achieve progressive performance improvements. Second, for unsupervised domain adaptation, I propose a divergence-based guidance mechanism for the adaptation process of object detection, aiming to improve the previous feature alignment methods that are unaware of foreground and poorly-aligned regions. Moreover, for semantic segmentation, I propose several methods to solve the following problems: 1) noise in the pseudo labels, 2) difficulty of performing adversarial learning, and 3) neglect of diversity within the target domain. To be specific, for the first problem, I propose a symmetric adversarial learning architecture to reduce the noise in the pseudo labels. For

the second problem, I propose variational autoencoder-based adaptation as an alternative, as well as a complement to the adversarial learning-based adaptation. For the third problem, I propose a novel concept of learning intra-domain style-invariant representation which considers the target-domain style diversity, leading to enhancement of generalization in the target domain. Finally, I propose the first solution for multi-source model adaptation of semantic segmentation, building on a novel concept of model-invariant feature learning which harmonizes the source-domain characteristics to produce more generalizable features. Moreover, to expand the application scenarios of multi-source model adaptation, I propose a union-set multi-source setting that relaxes the requirement for label space and develop a method for this setting still based on the model-invariant feature learning. By tackling the problems mentioned above, my studies realize better results for mitigating the label dependence, leading to improved practicality of the image recognition technologies.

The proposed methods of this thesis are briefly described as follows. For chronic gastritis classification using gastric X-ray images, I develop a semi-supervised learning method based on the tri-training architecture where disagreements of three heterogeneous models are exploited to improve the pseudo label generation. As to unsupervised domain adaptation, the method for object detection employs image-to-image translation to produce prediction divergence maps, which are used to guide the previous adversarial learning-based feature alignment, enabling it to be aware of the important foreground and poorly-aligned regions. Moreover, for unsupervised domain adaptation of semantic segmentation, I propose three methods as follows: 1) the first method trains two models by performing symmetric adaptation via adversarial learning and uses the prediction consistency between the two models to reduce the pseudo label noise; 2) the second method trains a variational autoencoder to learn the source-domain feature distribution and aligns the target-domain distribution with that learned by the variational autoencoder; 3) the third method generates intra-domain style-diversified images with multimodal image-to-image translation to learn intra-domain style-invariant representation with a self-ensembling architecture. Finally, for model adaptation of semantic segmentation in the two multi-source settings, I develop two methods both on the basis of the model-invariant feature learning which aims to produce features of similar distributions from different source-domain models, thereby harmonizing the model characteristics derived from different source domains.

This thesis consists of six chapters. Chapter 1 presents the background and proposition of this thesis. Chapter 2 presents the related works and clarifies the problems to be solved in this thesis. Chapter 3 proposes a semi-supervised learning method for chronic gastritis classification using gastric X-ray images. Chapter 4 proposes several methods for unsupervised domain adaptation of object detection and semantic segmentation. Chapter 5 proposes two methods for multi-source model adaptation, of which one method is developed for a novel union-set multi-source setting. Finally, Chapter 6 makes a conclusion of this thesis and discusses the future directions.

In summary, this thesis presents methods for mitigating the label dependence in three different directions. The methods solve some remaining problems of the previous works and can consequently achieve superior performance. By conducting experiments for tasks including gastritis classification, object detection, and semantic segmentation, the effectiveness of the proposed methods is validated.