氏　　名	シケツ　オウ Zhijie WANG
学位の種類	博　士（情報科学）
学位記番号	情　博　第　７８３　号
学位授与年月日	令和４年　９月２６日
学位授与の要件	学位規則第４条第１項該当
研究科、専攻	東北大学大学院情報科学研究科（博士課程）システム情報科学 専攻
学位論文題目	Semantic Segmentation with a Limited Amount of Training Data: Practical Domain Adaptation and Few-shot Learning (限られた学習データ下の画像セグメンテーション：実用的なドメイン適応と少数事例学習)
論文審査委員	（主査）東北大学教　授　岡谷 貴之 東北大学教　授　滝沢 寛之　　東北大学教　授　篠原 歩 東北大学准教授　鏡 慎吾

論 文 内 容 の 要 旨

第１章 Introduction

Due to recent advancements in deep learning, Convolutional Neural Networks (CNNs) have been applied in many computer vision tasks, such as classification, object detection, and semantic segmentation. However, proper training of CNN models usually requires a substantial quantity of labeled data. When the amount of data is insufficient, it will be hard to achieve adequate accuracy using CNN-based semantic segmentation methods, limiting the applications. For example, in autonomous driving scenarios, CNN models may not be able to recognize roads or pedestrians due to a lack of enough labeled data, thus causing accidents. On the other hand, it is laborious and time-consuming to obtain a good amount of labeled data. This is particularly true for pixel-wise classification tasks, such as semantic segmentation, which require annotators to pay enormous attention to images' details. In this thesis, to reduce the annotation cost in the tasks needing pixel-wise labeling, we study the problems of few-shot learning and unsupervised domain adaptation for semantic segmentation.

第２章 Redefining the Roles of Multi-level CNN Features for Few-shot Segmentation

Few-shot learning is a framework to train a model for an unseen class, given only several (e.g., one or five) labeled samples for the class, aiming to reduce the labeling cost. The current mainstream method for the few-shot learning of semantic segmentation is to place the mid-level features in the leading position. Specifically, it generates a similarity map between a query and labeled images using their high-level features and uses it as prior information to identify the coarse location of the target objects. In Chapter 2, we reinterpret this strategy as follows: the similarity map of the high-level features plays the central role, while the mid-level features play a secondary role. In particular, we interpret that the current methods use the similarity map as an

initial estimate of the object's location and then refine this estimate using the mid-level features. With this revised understanding, we show it is easy to extend the current methods, leading to further improvements. Specifically, we present a technique that uses current methods to update the estimate iteratively, beginning with the initial estimate. This strategy adds to performance improvements without bells and whistles. We show through experiments that this solution is applicable to almost all recent methods for few-shot segmentation and significantly enhances their performance.

第 3 章 Improved Unsupervised Domain Adaptation by Cross Region Alignment

As human annotation is expensive, researchers have proposed employing synthetic data generated from some 3D engines for which exact pixel-level annotations are easily accessible. However, it is often difficult to adapt neural networks trained on synthetic data to actual data due to the variation in distribution between synthetic and real images. Unsupervised domain adaptation is proposed for this problem, in which we have a large amount of annotated synthetic data and some unlabeled real data. There are two mainstream methods for this topic, the first one is self-training, in which we need to generate pseudo labels for real data and train a new model with them; the second one is adversarial training, in which we align the distributions between the synthetic data and real data; therefore, the model trained with the former can also achieve good performance in the latter. In Chapter 3, we offer a novel strategy that integrates both methods, in which the adversarial training is to align two feature distributions in the real data. It divides real images into two splits using a self-training framework, establishing the feature distributions to align. Our methods can continually improve the accuracy of current unsupervised domain adaptation. Experiments demonstrate the efficacy of our proposed method can always lead to better performance on several benchmark tasks and different baseline models.

第 4 章 Rethinking Unsupervised Domain Adaptation

Nearly all current unsupervised domain adaptation methods operate under the premise that there are no labeled label data available in the real scenario, and they adhere very closely to this assumption. In Chapter 4, first and foremost, we cast doubt on that notion. Because it is impossible to deploy any CNN-based models without evaluating how well they work, we feel that this is an unattainable goal to achieve in reality. The experiments of the unsupervised domain adaptation done in the past appear to go in that direction. However, except for a few scarce circumstances, we will be required to test our models appropriately before their deployment. In our investigation, using this practitioner's point of view as our guide, we situate ourselves in a data-centric posture. First, we assume that real labeled samples are accessible. However, we do not discount the cost of obtaining these samples. The next step is to analyze how we can or ought to use the labeled real samples, supposing that we have access to a few of them. Alternately, we consider the quantity of data required to reach a goal that we have established. This viewpoint is in contrast to the one presented in the earlier research, which may have emphasized the development of methodologies rather than the application of data.

論文審査結果の要旨

情景の画像から様々な物体等が画像内に占める領域を，その境界を含めて画素単位で正確に切り出す画像セグメンテーションは，画像認識の基本的な問題の一つであり，自動運転や医用画像診断など，複数の重要な応用が存在する。画像分類と比べて，教師データを人手で作るコストが高いことから，限られた量の学習データのみを使って高精度な推論を可能にする学習方法が求められている。本論文は，そのような方法の中から実用上重要な二つのアプローチ，すなわち少量事例学習（few-shot learning）とドメイン適応（domain adaptation）を扱ったもので，全編 5 章からなる。

第 1 章は序論であり，本研究の目的と背景を述べている。

第 2 章では，少数事例学習の方法を提案している。入力画像から抽出した高位特徴と中位特徴に関する発想を転換し，従来研究が想定していたそれぞれの役割をそっくり入れ替えることにより，提案方法は導き出されている．それによって推論の反復実行が可能となり，推論精度の向上をもたらしている。実験によって提案手法の有効性を示しており，車載カメラの画像を用いた標準的なベンチマークテストにおいて，世界最高精度を達成していることは重要な成果である。

第 3 章では，無教師ドメイン適応の方法を提案している。ソースおよびターゲットと呼ばれる二つのドメイン間で,取り出した画像特徴の分布を一致させることが課題となるが,提案手法は,敵対的学習の考え方に基づき，ドメイン間のみならず従来難しかったクラス間の分布をも接近させることを可能としている。実験によって提案手法の有効性を示しており，車載カメラの画像を用いた標準的なベンチマークテストにおいて，世界最高精度を達成していることは重要な成果である。

第 4 章では，従来研究が見過ごしてきた，無教師ドメイン適応におけるハイパーパラメータの決定に係るいくつかの課題を論じている。ターゲットドメインでの少量のラベル付きデータを用いてハイパーパラメータを決定する場合，ラベル付きデータ量が最終的な推定精度に与える影響を議論し，さらに同じラベル付きデータを流用して行う教師あり転移学習を，無教師ドメイン適応手法と比較し，どのような条件でどちらが優位となるかを論じている。無教師ドメイン適応を実世界の問題解決に利用する上でのベストプラクティスを示したものであり，重要な成果である。

以上要するに，本論文は，画像セグメンテーションを限られた量のデータで学習する問題に対し，条件に応じて使い分けることのできる複数の方法を提案しており，それらは同条件下で従来の類似手法よりも高い精度で推論を行える。この成果は，システム情報科学ならびにコンピュータビジョンの発展に寄与するところが少なくない。

よって，本論文は博士（情報科学）の学位論文として合格と認める。