INFO4990 Research Methods
Literature Review

Cross-domain Mitochondria Segmentation in Electron Microscopy Images

Yifei Yue 530840789

1 Introduction

With the rapid advancements in deep learning and computer technology, medical image analysis has emerged as a pivotal domain in biomedical research. Especially in electron microscopy images, deep learning techniques are capable of extracting more complex features from raw data without the need for prior domain knowledge. Adaptable to various scales and types of data, these techniques enhance robustness and scalability, thereby becoming the leading methodology. As a result, they offer us the potential for a deeper understanding of cellular structures and functions.

Despite significant technological advancements, a primary challenge in medical image analysis is the scarcity of annotated data. Labelling medical images not only demands the expertise and experience of professionals such as doctors and radiologists but is also expensive, time-consuming, and intricate. Particularly in electron microscopy images, the ultra-high spatial resolution reveals meticulous details of minute cellular and sub-cellular structures, making manual segmentation especially challenging and laborious. Hence, developing a method capable of effective large-scale electron microscopy image segmentation, even in the absence of annotated data, has become a pressing necessity.

To address this issue, researchers have gradually shifted focus to a cross-domain strategy: domain adaptation. Its core principle involves training machine learning models on source datasets, and then transferring this knowledge to target datasets that might differ from the source. The emergence of this approach offers a fresh perspective to alleviate discrepancies between datasets, allowing models to excel across multiple diverse datasets. This strategy, encompassing image processing, image segmentation, and the application of deep learning in medical images, has been widely recognized and explored by researchers.

However, even as domain adaptation opens new horizons, the challenge of effectively implementing it, especially when target datasets lack sufficient annotations, remains unresolved. Furthermore, while breakthroughs in domain adaptation for natural images are evident, medical images often involve more complex, higher-dimensional data, such as volumetric information and temporal sequences. This suggests that current deep-learning strategies might not be entirely suitable for medical image applications.

In light of this, this paper will delve into the latest research findings in electron microscopy image analysis, with a particular focus on advancements in image processing, image segmentation, deep learning, and domain adaptation. We have adopted a systematic approach, deeply exploring relevant literature to comprehensively identify and summarize current research trends and future challenges.

2 Domain Adaption

In Domain Adaptation (DA), the source domain and the target domain share the same learning task. However, in practice, DA can be classified into different categories based on various application scenarios, constraints, and algorithms. When we focus on the application of DA in deep learning, based on the availability of labels, deep learning-based DA methods can be divided into supervised domain adaptation, semi-supervised domain adaptation, and unsupervised domain adaptation.

2.1 Supervised DA

In supervised domain adaptation strategies, there is a certain amount of labelled data in the target domain available for model training. The core idea of this approach is to utilize these labelled data to adjust or fine-tune the model, ensuring it adapts more precisely to the target domain. This often involves retraining or refining the model to perform better on labelled data from the target domain. A common strategy is to train the model in the source domain and then fine-tune it in the target domain.

Ghafoorian et al. [31] conducted a detailed evaluation of fine-tuning strategies for brain lesion segmentation. They employed a domain adaptation method based on a CNN model previously trained on brain MRI scans. What's unique about their method is that they transferred the trained weights from the source domain, froze the first i layers of the model, and fine-tuned the remaining d-i layers only on target domain data. This strategy significantly enhanced the model's generalization performance: with just a minimal amount of target samples for fine-tuning, the Dice score rocketed from a negligible 0.005 to 0.63. On the other hand, Samala et al. [82] adopted a different approach. They used a DCNN model similar to AlexNet pre-trained on ImageNet, aiming for breast cancer classification. After collecting 19,632 ROIs of 2,454 tumour lesions, they further fine-tuned the model. This fine-tuning strategy also yielded significant results, with the recognition AUC value for breast cancer increasing from 0.78±0.02 to 0.90±0.04. The study by Abbas et al. [85] is even more noteworthy. They focused on chest X-ray classification and pre-trained a CNN model on ImageNet. The challenge they faced was the irregularity of the dataset, prompting them to adopt a class decomposition strategy. In this strategy, they further divided each image category into k subsets and assigned new labels for each subset. The result was remarkable: although the pre-trained ResNet model already achieved an accuracy of 82.24%, after applying the class decomposition strategy, the accuracy of the same model significantly increased to 99.8%.

All of the aforementioned methods employ a single-step Domain Adaptation (DA) strategy, which directly transfers a pre-trained model to the target domain. However, this approach encounters issues when there are limited samples in the target domain: a small number of samples might not adequately represent the overall data distribution of the target domain. This could lead to a decline in the model's generalization performance in the target domain and make fine-tuning challenging. To address this, some researchers have proposed multi-step domain adaptation methods.

For instance, Gu et al. [86] devised a two-step adaptation strategy for skin cancer classification. Initially, they fine-tuned ResNet on two larger skin cancer datasets and then trained it on the target domain, which was a smaller medical imaging dataset. This adaptation approach yielded superior experimental results on both the MoleMap and HAM10000 datasets compared to the direct transfer method of single-step DA.

But it's not just the choice of transfer strategy that matters; the structure of the model itself also influences its performance in medical image analysis. While many researchers prefer to use models pre-trained on datasets like ImageNet, traditional 2D CNNs might struggle to capture the rich information in three-dimensional medical images. In contrast, 3D CNNs can extract features in three dimensions, such as width, height, and depth, thereby better capturing the spatial patterns in medical images. Hence, many researchers specifically design 3D CNNs for medical tasks, training them with medical images as the backbone to facilitate subsequent data adaptation tasks.

Hosseini-Asl et al. [87] designed a 3D CNN specifically for brain MR image classification. After pre-training on source domain MR images, they fine-tuned the network's fully connected layer in the target domain. Upon final testing, this model achieved a classification accuracy of 97.6% on the MRI dataset, distinguishing between Alzheimer's patients and a healthy control group. Similarly, Kaur et al. [90] proposed a specific training strategy for a 3D U-Net. Their method involved first pre-training the 3D U-Net on a source domain with a vast number of samples and then fine-tuning it using a limited amount of labelled data from the target domain. The advantage of this strategy was verified on the BraTS dataset [52], especially when tumour samples were scarce. With only 20 tumour samples, compared to other baseline methods, its Dice scores for core and enhancing tumours increased by 25.9% and 204.09%, respectively. Additionally, another study [91] adopted a strategy similar to Kaur's. In this research, the team utilized a large amount of X-ray Computed Tomography (CT) data and synthesized radial MRI data for pre-training. Subsequently, they only used a small amount of labelled target MRI data for network fine-tuning. Impressively, even with minimal labelled target data for fine-tuning, the accuracy achieved was still on par with baseline methods.

2.2 Semi-supervised DA

Indeed, supervised domain adaptation that relies on a significant amount of labelled data from the target domain can lead to high annotation costs in practical applications. In contrast, semi-supervised domain adaptation can leverage both a small amount of labelled data and a large amount of unlabeled data for training. This can reduce annotation costs while improving the model's generalization performance, hence attracting attention from many researchers.

Roels et al. [96] explored a semi-supervised domain adaptation method for the segmentation of electron microscopy images in their research. Their proposed "Y-Net" structure features a feature encoder and two decoders, where one decoder is dedicated to image segmentation and the other, termed the "reconstruction decoder", is aimed at reconstructing both source and target domain images. Notably, in the initial training phase of the model, it operates in an unsupervised manner. Later, for more precise adaptation to the target domain, they removed the reconstruction decoder and fine-tuned using target samples. Compared to the traditional finetuning baseline (FT) method, which only achieved an IoU of 28.7% on the Drosophila dataset, the approach of Roels et al. elevated the IoU to an impressive 49.9%. On the other hand, Madani et al. [97] proposed a Domain Adaptation (DA) framework based on a semi-supervised Generative Adversarial Network (GAN) for chest X-ray image classification. This is not a regular GAN model; it uniquely uses labelled source data, unlabeled target data, and generated images as inputs, enabling the discriminator to perform a three-class classification: normal, diseased, or generated image. The ingenuity of this strategy is that when the unlabeled target data is categorized as a generated image, it aids in the loss computation. As a result, both labelled and unlabeled data are harnessed together, exhibiting semi-supervised characteristics. Experimental results showed that this method, compared to the traditional supervised learning Convolutional Neural Network (CNN), required an order of magnitude less data on both the NIH PLCO dataset [98] and the NIH Chest X-Ray dataset [99]. Remarkably, this semi-supervised model needed only 10 labelled images per category to achieve an accuracy of 73.08%. In contrast, to attain the same accuracy, the conventional CNN required between 250 to 500 labelled images.

2.3 Unsupervised DA

In the field of medical image analysis, while Supervised Domain Adaptation (Supervised DA) and Semi-Supervised Domain Adaptation (Semi-Supervised DA) have received attention, their shared bottleneck lies in the need for labelled data from the target domain to adjust the model. In reality, acquiring such data is often challenging in many scenarios, presenting significant hurdles in data acquisition and annotation. In contrast, Unsupervised Deep Domain Adaptation stands out because of its characteristic of not requiring labelled target data. The essence of Unsupervised Domain Adaptation (Unsupervised DA) is to employ technological methods to narrow the distribution gap between the source and target domains, enabling the model to self-adapt in situations lacking target domain labels. This approach cleverly addresses the data annotation issue.

Specifically, based on knowledge transfer strategies, existing unsupervised deep DA methods can be categorized into several main avenues: feature alignment, image alignment, combined feature + image alignment, and feature learning. Among these, the research emphasis of feature alignment strategies is on capturing cross-domain domain-invariant features through specially designed CNN models. This method aims to help the model overcome differences between data distributions, thereby enhancing the model's generalization capability and performance in new domains.

Kamnitsas et al. [33] proposed a multi-adversarial network based on DANN for brain lesion segmentation. In this design, the domain discriminator is trained synchronously with the segmentation network. Moreover, the authors believed that merely adjusting the final layer of the segmenter might not be comprehensive enough, so domain discriminators were introduced at multiple layers of the network, making the model more robust against image quality variations across different domains. Astonishingly, using this unsupervised domain adaptation method, the model, trained on source domain S and tested on target domain T, saw segmentation accuracy skyrocket from 15.7% to 62.7%, almost equivalent to 63.5% from supervised training on target domain T. Similarly, Zhang et al. [106] focused on the classification task of Alzheimer's disease and mild cognitive impairment on ADNI, introducing a Domain Adaptation (DA) method based on adversarial learning. This approach adjusts brain MRI features of the source and target domains via a meticulously designed cyclic feature adaptation module, making them more aligned and thereby optimizing the recognition accuracy of brain diseases.

However, it's important to emphasize that feature alignment strategies come with some tough challenges. First, not all features from the source and target domains should be aligned. Blindly aligning irrelevant features might backfire, compromising the transfer outcome. Secondly, an overzealous pursuit of complex alignment strategies might result in the model overfitting to source domain data, thereby weakening its generalization ability in the target domain. Such challenges prompted the introduction of the "image alignment" strategy. Unlike merely aligning features, image alignment considers the overall visual distribution, not being restricted to a single feature, thus somewhat mitigating the risks of improper alignment.

Mahmoud and his team [126] adopted a groundbreaking GAN-based reverse domain adaptation method specifically designed for endoscopic image analysis. While traditional GANs aim to transform synthetic images into more realistic ones, this unique method goes against the norm by converting real images into synthetic forms. The core idea is to "synthesize" real medical images through adversarial training, making them more akin to synthetic images, and then use these "synthesized" real images for model training. The rationale behind this strategy is that when the network encounters images more similar to real data during training, it exhibits better generalization capability for real medical image data. As a result, the gap between the source and target domains is narrowed. Experiments by Mahmoud's team demonstrated that this reverse domain adaptation, when applied to depth estimation in real porcine colon endoscopy, led to a staggering 78.7% improvement in structural similarity compared to the traditional method that directly uses synthetic data. On a related note, Gulrajani et al. [123] expanded the training dataset for brain tumour segmentation by introducing CycleGAN. Their method first simulates the creation of synthetic MR images with tumours, then uses CycleGAN to transform them into nearly real MRI images, thus enriching the diversity of training samples. This innovative strategy was validated on the BraTS [52] dataset and yielded satisfying results for brain tumour segmentation. It's worth noting that the conventional 2D U-Net method only utilized the original training data from BraTS for medical image segmentation, whereas Gulrajani's team innovatively simulated tumour-bearing images to bolster the training dataset. As a result, the 2D U-Net segmentation's Dice scores on three key indicators improved by 5.29%, 1.32%, and 1.97% respectively compared to the traditional method.

While the image alignment strategy has demonstrated its advantages in certain applications, its limitations should not be overlooked. In some cases, images between two domains might visually appear similar, but semantically they could be vastly different. This implies that when we only align the overall image from a macro perspective, we might miss these crucial detailed differences. Even more concerning in the realm of medical imaging, these subtle differences often pertain to matters of life, as they are frequently closely tied to vital biomarkers and diagnostic information. Thus, relying solely on image alignment might result in the loss of this crucial information.

To address this issue, a strategy that combines both feature alignment and image alignment emerged. This strategy is dedicated to integrating the strengths of the two alignment methods: feature alignment focuses on high-level semantic information, while image alignment emphasizes low to midlevel visual patterns. This way, it aims to capture the differences between the source and target domains from multiple perspectives, achieving a more comprehensive match and transformation.

Chen et al. [36] employed this strategy for cross-modality cardiac image segmentation. They first utilized CycleGAN to transform labelled source images into visually similar target images. Subsequently, they designed a dual-stream CNN structure with a domain discriminator. This structure could accept both the transformed target images and actual target images, and further minimize the distance between the two domains through adversarial learning. This method achieved outstanding experimental results on the MM-WHS dataset [108]: the average Dice value reached up to 83.7% for CT images and an even higher 85.4% for MRI images. These results are only slightly off from the best results obtained through fully supervised training, especially for CT images, where the Dice score was merely 5 percentage points shy of the highest score from supervised training.

Feature learning provides another innovative approach to domain adaptation strategies. Its core idea is to try to mine common features from both the source and target domains in the absence of target domain labels. Compared to the strategies mentioned earlier, the distinct advantage of feature learning is that it isn't confined to specific features or alignment techniques, giving it greater flexibility and making it broadly applicable across various tasks and scenarios.

An and colleagues [133] proposed an ingenious approach in this regard. They designed a hierarchical unsupervised feature extractor that combined a convolutional autoencoder with a pre-trained CNN. The intent behind this design was to embrace the strengths of both feature types: on one hand, inheriting universal image features from the pre-trained CNN and, on the other hand, delving deep into the unique details of medical images through the convolutional autoencoder. This dual feature extraction strategy ensures that the learned features are more representative, allowing them to be effectively applied to subsequent classification tasks. The final experimental results demonstrated that compared to the traditional transfer learning method based on a pre-trained AlexNet, their method achieved noticeable performance improvement on the ImageCLEF 2016 dataset, raising the classification accuracy from 79.21% to 81.33%.

3 Conclusion

4 References