Advancing Medical Imaging Informatics by Deep Learning-Based Domain Adaptation

Anirudh Choudhary*,†, Li Tong*,†, Yuanda Zhu‡, May D. Wang†,‡,§

1 Department of Computational Science and Engineering, Georgia Institute of Technology, GA, USA
2 Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, GA, USA
3 School of Electrical and Computer Engineering, Georgia Institute of Technology, GA, USA

* Equal Contributing First Authors

1 Introduction

Medical imaging informatics utilizes digital imaging processing and machine learning (ML) to improve the efficiency, accuracy, and reliability of imaging-based diagnosis [1]. During the past few years, medical imaging informatics has made remarkable progress due to the increasing availability of data and the rapid development of deep learning (DL) techniques [2]. However, fundamental challenges hinder the effective deployment of deep learning models to clinical settings. Annotated medical datasets are limited due to the tedious labeling process [2] and are not easily shared due to privacy concerns [3, 4]. While multicenter datasets can increase the amount of annotated data, these datasets suffer from heterogeneity due to varying hospital procedures and diverse patient populations [5, 6]. Due to a distribution shift (also known as domain-shift) between the available training dataset and the dataset encountered in clinical practice, pre-trained models trained by one dataset may fail for another dataset.

1.1 What is Transfer Learning and Domain Adaptation

Transfer learning (TL) [7] is a technique that applies knowledge learned from one domain and one task to another related domain and/or another task, when there is insufficient labeled data for traditional supervised learning. For medical imaging, a domain usually refers to images or features, while the task refers to segmentation, classification, etc. Mathematically, X and Y being random variables, where X is d-dimensional feature space with marginal probability distribution $p(X)$ and Y is a labeled vector with conditional probability distribution $p(Y|X)$, we use $D = \{X, p(X)\}$ to represent domain and $T = \{Y, \hat{p}(Y|X)\}$ to represent task, where $\hat{p}$ is learnt using a function (e.g., neural network). If both source ($D_s$) and target domains ($D_t$) are similar, i.e., $D_s \sim D_t$, then $D_s$ and $D_t$ can use the same ML model for similar tasks ($T_s \sim T_t$). However, if $D_s \not\sim D_t$, or $T_s \not\sim T_t$, the ML model trained on the source domain might have decreased performance on the target domain ($D_t$). TL can be categorized into three types based on the relationships between domains and/or tasks:

1. Inductive TL requires some labeled data. While the two domains may or may not differ ($D_s \sim D_t$, or $D_s \not\sim D_t$), the target and source tasks are different ($T_s \not\sim T_t$), for e.g., 3D organ reconstruction across multiple anatomies;

2. Transductive TL requires labeled source data and unlabeled target data with related domains ($D_s \sim D_t$) and same tasks ($T_s \equiv T_t$), while the marginal probability distributions differ ($p(X_s) \not\equiv p(X_t)$), for e.g., lung tumor detection across X-Ray and computed tomography images;

3. Unsupervised TL does not require labeled data in any domain and has different tasks ($T_s \not\equiv T_t$), for e.g., classifying cancer for different anatomies using unlabeled histology images.

Domain Adaptation (DA) is a transductive TL approach that aims to transfer knowledge across domains by learning domain-invariant...
transformations, which align the domain distributions (see Figure 1-b). DA assumes that the source data is labeled, while the target domain can be (a) fully labeled data (i.e., supervised setting); (b) a small set of labeled data (i.e., semi-supervised setting); or (c) completely unlabeled data (i.e., unsupervised setting).

1.2 Using Domain Adaptation to Improve Model Training in Medical Imaging

In biomedical imaging, due to the existence of multi-modality imaging (e.g., magnetic resonance imaging (MRI) and positron emission tomography (PET)), DA has advantages over conventional semi-supervised learning or unsupervised learning. Cross-modal DA transfers labels between distinct, but somewhat related, image modalities (e.g., MRI and computed tomography (CT)). Single-modality DA adapts different image distributions within the same modality [8] (Figure 2).

1.2.1 Challenge of Limited Training Data

Developing accurate DL models requires large scale training data covering a wide range of input variations. However, in biomedical imaging, due to concerns over patient privacy [3,4] and lack of manual annotation of images by clinical experts [9], few well-labeled datasets are available for training. This situation is worse for rare diseases, where a low number of positive cases lead to significantly unbalanced datasets [10].

DA can mitigate the lack of well-annotated data by augmenting target domain data, either by generating synthetic labeled images from source images or aligning source and target image features and training a task network on them [11, 12, 112]. For example, MRI achieves higher resolution for soft tissue imaging compared to CT [13]. As such, MRI is preferred for neuroimaging, and brain MRI annotations are easily accessible. On the other hand, CT imaging is fast and less expensive and may be preferred in trauma situations [14]. Thus, through DA, annotated MRI scans from historical subjects can be combined with CT to reduce the number of image acquisitions needed. As another example, Hematoxylin and Eosin (H&E) stained images are widely available, while immunohistochemistry (IHC) images, which clearly highlight nucleii via specific biomarkers [15], are not. DA methods can translate multi-stained H&E-stained images to the IHC domain, making nuclei detection easier [16].

1.2.2 Challenge of Dataset Variations

To train robust DL models, many studies rely on images aggregated from multiple institutes such as NCI/NIH The Cancer Genome Atlas (TCGA) [17] and Stanford’s large chest radiograph dataset, CheXpert [18]. The data in these repositories are heterogeneous due to varying hospital processes (image acquisition platforms or data preparation protocols), different demographics of patient populations (ethnicity, gender, age), or different pathological conditions [5]. Specifically, pathology images have stain variations [19] while MRIs are susceptible to varying magnetic fields and contrast agents [20]. Such intra- or in-

---

**Fig. 1** a) Transfer learning (TL) and its different types; b) Overview of domain adaptation; c) Organization of this survey paper.
ter-dataset variations cause the training and test dataset to have different distributions, resulting in a domain-shift which impacts model generalization [21, 22]. Diversifying the training data by creating larger datasets is a possible solution, but recent medical imaging studies [21, 23, 24] have shown that it does not guarantee improved generalization. DA methods try to minimize the dataset variation, while retaining the distinguishing aspects for task classifier, and have been shown to generalize well in image segmentation tasks for multiple modalities [25-27].

The organization of the survey paper is illustrated in Figure 1-c. Section 2 introduces our survey methodology in identifying and selecting relevant medical imaging studies. Section 3 presents various DL-based frameworks in the DA literature and the current best practices for medical imaging. Section 4 summarizes the current DA challenges and future opportunities.

2 Materials and Methods
In this survey, we examined publications between 2017-2020. We considered the proceedings of leading peer-reviewed journals and conferences, including IEEE Transactions on Medical Imaging, the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), Medical Image Analysis (Elsevier), IEEE International Symposium on Biomedical Imaging (ISBI), Conference on Computer Vision and Pattern Recognition (CVPR), Association for the Advancement of Artificial Intelligence (AAAI), and the International Conference on Medical Imaging with Deep Learning (MIDL). Additionally, we identified a few relevant works from arXiv and PubMed, which were not found in review proceedings. Our search keywords included ‘Domain Adaptation’, ‘Transfer Learning’, ‘Cross Modality’, ‘Multimodal’ and ‘Medical Image Adaptation’, and ‘Medical Images’. We found that radiology and pathology were the most common application areas (characteristics of our results are illustrated in Figure 3-a and Figure 3-c). Cross-modality segmentation is observed more extensively in radiology compared to other areas. MRI, CT, and PET maintain better relative morphological-consistency of organs and provide complementary information for disease detection [30]. On the other hand, histopathology with smaller objects, such as nuclei, are prone to artifacts during cross-modal translation [31].

3 Deep Learning-based Domain Adaptation
DL-based DA is achieved using various representation learning strategies such as aligning the domain distributions, learning
a mapping between domains, separating normalization statistics, and ensemble-based approaches [32-34]. As shown in Figure 2-a, there are two families of DA approaches for medical imaging: (a) Domain Transformation (DT-DA) translates images from one domain to the other domain, so that the obtained models can be directly applied to all images, and (b) Latent Feature Space Transformation (LFST-DA) aligns images from both domains in a common hidden feature space to train the task model on top of the hidden features. These two approaches can work together to improve adaptation performance by preserving finer semantic details [35,36]. We have summarized the application of these DA methods in medical imaging in Table 1.

3.1 Domain Transformation in Domain Adaptation

DT-DA translates images from one domain to another domain (i.e., image-to-image translation [37]). Such translation is typically done using generative models (e.g., generative adversarial networks (GANs)) [38], which achieve pixel-level mapping by learning the translation at a semantic level. The translation direction is usually decided by the relative ease of translation and modeling in a modality [39]. For example, Dou et al., [36] observed lower performance in adapting CT to MRI for cardiac images, since cardiac MRI is more challenging to segment. The task networks are trained independently or jointly, with the image-translation network, using the labeled source images [35]. DT-DA performs alignment in the image space instead of the latent feature space, leading to better interpretability through visual inspection of synthesized images [40], enforcing semantic consistency, and preserving low-level appearance aspects using shape-consistency [41] and structural-similarity constraints [42].

3.1.1 Unidirectional Translation

Unidirectional translation maps images from the source domain to the target domain or vice versa using GANs (e.g., vanilla GAN and conditional GAN (cGAN)) [43]. Compared with vanilla GAN, cGAN conditions the training of the generator and discriminator on extra information such as the class label. Yoo et al., [44] proposed pixel-level domain transfer using cGAN with a domain discriminator. Liu and Tuzel [45] utilized GANs coupled with shared weights to generate paired synthetic source and target images sharing high-level abstraction. Boussaliss et al., [40] leveraged cGAN with the content-similarity loss to generate realistic target images and jointly trained the GAN discriminator with the task network. Unidirectional translation has been applied to remove dataset variations. For example, Bentaieb et al., [29] designed a stain normalization approach, using a task conditional GAN to translate H&E images to a reference stain. Madani et al., [46] proposed a semi-supervised approach for cardiac abnormality classification using a GAN discriminator for abnormality classification in minimally labeled X-ray images, and showed that the adversarial loss could reduce domain overfitting. Mahmood et al., [39] translated real endoscopy images to graphically-rendered synthetic colon images with ground-truth annotations, for depth-estimation during surgical navigation. Unidirectional translation has also been applied to cross-modality scenarios. For instance, Zhao et al., [47] proposed a modified U-Net to translate paired brain CT to MRI.

3.1.2 Bidirectional Translation

Bidirectional image translation (also known as reconstruction-based DT) leverages two GANs, constraining the mapping space by enforcing semantic-consistency between the original and reconstructed images. CycleGAN, by Zhu et al., [48], is one of the most popular architectures for bidirectional translation. CycleGAN utilizes cycle-consistency to constrain the translation mapping and improve the quality of generated images. CycleGAN has been expanded to handle larger domain shifts with semantic-consistency loss functions (CyCADA [35]), multi-domain translation (StarGAN [49]), and translation between two domains with multi-modal conditional distributions (MUNIT [50]). In supervised learning, bidirectional translation expands the training data to make the segmentation task model...
robust. The translation and segmentation network can be trained either independently (two stages) or jointly. Zhang et al., [51] presented a one-stage framework with an additional shape-consistency loss in CycleGAN to achieve better segmentation masks and lower failures. Chartsias et al., [11] used a two-stage framework to segment MRI images using CT images. Cai et al., [52] combined segmentation loss on generated images as an additional shape constraint for 3D translation and leveraged MRI for pancreas segmentation in CT images. In the unsupervised setting, image translation is used to create labeled data for the target domain. Huo et al., [12] proposed a joint optimization approach for the synthesis and segmentation of CT images using labeled MRI. Their framework achieved comparable performance in comparison to the fully labeled case.

There are a few observations about GANs: (a) CycleGAN does not guarantee consistent translation of minor anatomical structures and boundaries [53], and thus needs additional constraints like gradient [53] and shape consistency [51]. For instance, Jiang et al., [54] incorporated tumor-shape and feature-based losses to preserve tumors while translating CT data to MRI data; (b) Attention networks can account for varying transferability of different image regions [55]. For instance, Liu et al., [56] proposed a novel attention-based U-Net [57] as a GAN generator to translate hard-to-generate textured regions from MRI to CT. For alternate scenarios such as 3D-2D, paired images, or semi-supervised DA-DT, Zhang et al., [51] segmented X-ray images by using synthetic X-ray images created from accessible 3D CT annotations. Nguyen et al., [58] used semi-supervised DA with paired CT images to constrain CycleGAN to generate more realistic images. Pan et al., [30] leveraged MRI to generate missing PET images for patients for Alzheimer’s disease diagnosis. Chen et al., [59] proposed state-of-the-art unsupervised segmentation method using bidirectional DA-DT between MRI and CT, combining CycleGAN with shared feature encoder layers between domains. Their method resembled CyCADA [41] and showed the efficacy of combining DT with feature-based alignment; (c) DA-DT can be used for single-modality medical imaging. Chen et al., [28] leveraged a CycleGAN with semantic-aware adversarial loss to perform lung segmentation across different chest X-ray datasets.

### 3.2 Latent Feature Space Transformation in Domain Adaptation

Unlike the image-to-image translation in DT-DA, the LFST-DA transforms the source domain and target domain images to a shared latent feature space to learn a domain-invariant feature representation. The goal is to minimize domain-specific information while preserving the task-related information. The LFST-DA can be trained in an unsupervised fashion to obtain a domain-invariant representation, or in a concurrent manner where the representation network and the task network (e.g., image classification network) are trained simultaneously to improve the performance. LFST-DA is used in three basic implementations: divergence minimization [60-64], adversarial training [65-68], and cross-domain reconstruction [69, 70]. Compared to DT-DA, LSFT-DA is more computationally efficient because it focuses on translating relevant information only instead of the complete image [34]. Also, feature-based domain alignment outperforms DT-DA by preserving task-critical features [35].

#### 3.2.1 Divergence Minimization

A simple approach to learn domain-invariant features and remove distribution-shift is to minimize some divergence criterion between source and target data distributions. Common choices include maximum mean discrepancy (MMD) [60], correlation alignment (CORAL) [61, 63], contrastive domain discrepancy (CDD) [64], and Wasserstein distance [62]. MMD, CORAL, and Wasserstein distances are class-agnostic divergence metrics and do not discard class labels when aligning samples. CDD-based DA aligns samples based on their labels, by minimizing the intrain-class discrepancy and maximizing the inter-class discrepancy. MMD and CORAL are two of the most utilized divergence metrics that match the first-order moment (mean) and the second-order moment (covariance) of distributions. However, the represented hidden features can be complicated in the real world and may not be fully characterized by mean and covariance. Wasserstein distance aligns feature distributions between domains via optimal transport theory. Compared to the adversarial-based approaches, divergence-based DA has not been as widely explored in medical imaging. For cross-modality DA, Zhu et al., [71] utilized maximum mean discrepancy to map MR and PET images to a common space to mitigate missing data. Several works have used same-modality DA to mitigate dataset variations in X-ray [72], retinal fundus [73], and electron microscopy images [74].

#### 3.2.2 Adversarial Training

Instead of minimizing a divergence metric, adversarial methods train a discriminator, typically a separate network, in an adversarial fashion against the feature encoder network. The goal of the feature network is to learn a latent representation such that the discriminator is unable to identify the input sample domain from the representation. For medical imaging, feature-based adversarial domain adaptation has been widely utilized for various applications. For example, in cross-modality adaptation, Zhang et al., [75] applied a domain discriminator to adapt models trained for pathology images to microscopy images. LSFT-DA is also used for single-modality adaptation to overcome dataset variations in pathology images, MR images, and ultrasound images. For example, Lafarge et al., [76] have utilized a domain discriminator to mitigate the color variations of histopathology images for mitosis detection in breast cancer. Kamnitsas et al., [77] have applied a domain discriminator to MR images from different scanners and imaging protocols to improve the brain lesion segmentation performance.

#### 3.2.3 Reconstruction-based Adaptation

The reconstruction-based adaptation maximizes the inter-domain similarity by encoding images from each domain to reconstruct images in the other domain. The reconstruction network (decoder) performs feature alignment by recreating the feature extractor’s input while the feature extractor (encoder) transforms input image into latent representation. Ghifary et al., [70] proposed
DRCN for object detection, using only target domain data reconstruction while Bousmalis et al., [69] proposed a domain separation network that extracts image representations in two subspaces: the private domain features and the shared-domain features, the latter being used to reconstruct input image. For medical imaging, reconstruction-based methods are less developed and are usually combined with adversarial learning. For same-modality adaptation, Oliveira et al., [80] have combined image-to-image translation with a feature-based discriminator to mitigate the variations in X-ray images and improve segmentation performance. For cross-modality adaptation, Ouyang et al., [81] combined variational autoencoder (VAE) with adversarial training to adapt MR to CT scans.

4 Challenges and Opportunities

4.1 Domain Selection and Direction of Domain Adaptation

Selecting related domains for effective knowledge transfer is an open-research area in ML. In medical imaging, domains are often selected based on the type of imaging technique (e.g., radiology), anatomy, availability of labeled data, and whether the modalities are complementary for the underlying task [30]. Regarding whether DA could be performed symmetrically across domains, the potential information loss in a particular direction is critical for assessing task performance. For example, for unsupervised DA from CT to MRI, reverse DA may sometimes be needed to preserve tumors [54]. For supervised DA between multiple H&E stained images, Tellez et al., [82] showed that mitosis-detection and cancer tissue classification in a particular color space leads to higher accuracy. Typically, to assess domain relationship and DA direction, it is necessary to use (a) large-scale empirical studies such as [6,58] exploring bi-directional DA across multiple datasets, (b) a representation-shift metric [24] to roughly quantify the risk of applying learned-representations from a particular domain to a new domain, or (c) multi-source DA [83], which automatically explores latent source domains in multi-source datasets and quantifies the membership of each target sample. However, such experimentation requires extensive benchmarking studies that are lacking in medical imaging.

4.2 Transferability of Individual Samples

Most DA studies for medical imaging assume that all samples are equally transferable across two domains. Thus, they focus on globally aligning domain distributions. However, the ability to transfer (or align) varies across clinical samples because of: (a) intra-domain variations (e.g., in multi-modal DA between MRI and CT, each modality can have contrast variations) [75]; (b) noisy annotations due to human subjectivity; (c) target label space being a subset of source label space [84]; and (d) varying transferability among different image regions [55] (e.g., tumors are difficult to translate and could be missed during CT to MRI image-translation [54]). Some samples in the source domain may be less useful and can lead to negative transferring [84], which adversely impacts DA. Selecting relevant samples or reducing the impact of outlier samples using transferability frameworks is a potential solution. Some strategies include weighting samples based on classifier discrepancy [85], down-weighting outlier classes using the classification probability for target data [86], leveraging open-set based optimization [87], and leveraging an attention mechanism [55] to focus on hard-to-transfer samples or using a noise co-adaptation layer [88]. Recent medical imaging studies have explored sample selection and transferability assessment using reverse classification accuracy [89], attention-based U-Net [56], and transferable semantic representations [84].

4.3 Limitations of Domain Adaptation in Medical Imaging

For medical imaging, most DL-based DA uses adversarial methods, primarily GAN for unsupervised DA. Adversarial methods are prone to errors because the discriminator can be confused, and there is no guarantee that the domain distributions are sufficiently similar [90]. Moreover, the generator in GAN is prone to “hallucinating” content to convince the discriminator that data belongs to the target distribution [91]. As such, CycleGAN could be trained to synthesize tumors in images of healthy patients. Beyond applying consistency constraints during image translation, artifacts which are not directly visible in synthesized images, are also important for consideration. For example, CycleGANs incorporate high-frequency information in the intermediate representation used by the second generator to translate the image back to the source domain [92]. This high frequency information can interfere with downstream tasks.

DT-DA approaches require translating the entire image, increasing the complexity of the models for large-sized medical images like whole slide images. Few studies [28, 36] have compared adversarial DA methods for MRI-CT translation. However, a comprehensive comparison of various feature-based DA approaches is lacking. Future studies could explore combining DT-DA and LFST-DA approaches [59]. Moreover, current frameworks typically focus on source-target domain pair, while many tasks, such as stain normalization in histopathology images, can be multi-domain [93].

4.4 Leveraging Synthetic Data

DA for medical imaging can be applied in relatively under-explored applications such as single-view 3D reconstruction [94] or temporal disease analysis [95]. This could benefit image-guided surgery, in which training data is very scarce and difficult to obtain [96]. One way is to leverage synthetic data with ground truth information, adapting it to the real data. This approach has been successfully applied in natural images [97]. Reverse domain adaptation (i.e., translating real data to synthetic data) is also a promising solution. Mahmood et al., [39] generated synthetic endoscopy data with known depth information by using an anatomical colon model and a virtual endoscope. This simulated data was used for 3D reconstruction of real endoscopic images. Pan et al., [30] translated MR data to generate synthetic PET images to infer missing patient scans for temporal analysis of Alzheimer’s disease [30]. Another area that could benefit from synthetic data is skin lesion detection [98].
<table>
<thead>
<tr>
<th>DA Method</th>
<th>Framework</th>
<th>Task</th>
<th>Source &amp; Target Domains</th>
<th>Anatomy</th>
<th>Learning Scenario</th>
<th>Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Single Modality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DT</td>
<td>Reconstruction</td>
<td>S</td>
<td>X-ray (different demographics)</td>
<td>Lung</td>
<td>Unsupervised</td>
<td>Chen. et. al., [28]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>X-ray (different demographics)</td>
<td>Lung</td>
<td>Unsupervised</td>
<td>Tang et. al., [100]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S</td>
<td>Synthetic data → MRI</td>
<td>Brain</td>
<td>Unsupervised</td>
<td>Gholami et. al., [99]</td>
</tr>
<tr>
<td></td>
<td>Unidirectional</td>
<td>C</td>
<td>X-ray (different demographics)</td>
<td>Lung</td>
<td>Semi-supervised</td>
<td>Madani et. al., [46]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>Ultrasound (different sources)</td>
<td>Fetal Head</td>
<td></td>
<td>Yang et. al., [25]</td>
</tr>
<tr>
<td>LFST</td>
<td>Adversarial</td>
<td>S</td>
<td>X-ray (different disease states)</td>
<td>Lung</td>
<td>Unsupervised</td>
<td>Dong et. al., [104]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S</td>
<td>Ultrasound (different sources)</td>
<td>Heart</td>
<td>Unsupervised</td>
<td>Degel et. al., [26]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S</td>
<td>MRI (different FLAIR sequences)</td>
<td>Brain</td>
<td>Unsupervised</td>
<td>Orbes-Arteaga et. al., [78]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S</td>
<td>MRI (moderate to severe TBI)</td>
<td>Brain</td>
<td>Unsupervised</td>
<td>Kamnitsas et. al., [77]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S</td>
<td>MRI (different disease types, demographics)</td>
<td>Brain</td>
<td>Unsupervised</td>
<td>Novosad et. al., [106]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S</td>
<td>MRI (different vendors)</td>
<td>Brain</td>
<td>Unsupervised</td>
<td>Yan et. al., [109]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S</td>
<td>MRI (cross-institutional)</td>
<td>Brain, Spine</td>
<td>Unsupervised</td>
<td>Shanis et. al., [79]</td>
</tr>
<tr>
<td></td>
<td>Divergence</td>
<td>S</td>
<td>X-ray (different disease types, demographics, contrast)</td>
<td>Lung</td>
<td>Unsupervised</td>
<td>Venkataramani et. al., [72]</td>
</tr>
<tr>
<td></td>
<td>Reconstruction</td>
<td>S</td>
<td>X-ray (different datasets)</td>
<td>Breast</td>
<td>Semi-supervised, Unsupervised</td>
<td>Oliveira et. al., [80]</td>
</tr>
<tr>
<td><strong>Cross-Modality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DT</td>
<td>Reconstruction</td>
<td>S</td>
<td>CT → MRI</td>
<td>Lung</td>
<td>Unsupervised</td>
<td>Jiang et. al., [54]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>CT → X-Ray</td>
<td>Lung</td>
<td>Unsupervised</td>
<td>Zhang et. al., [51]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S</td>
<td>MRI ↔ CT</td>
<td>Heart</td>
<td>Unsupervised</td>
<td>Chen et. al., [59]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S</td>
<td>MRI → CT</td>
<td>Heart</td>
<td>Unsupervised</td>
<td>Chartsias et. al., [11]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S</td>
<td>CT ↔ MRI</td>
<td>Heart</td>
<td>Supervised</td>
<td>Zhang et. al., [41]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S (3D)</td>
<td>CT ↔ MRI</td>
<td>Heart, Breast, Pancreas</td>
<td>Unsupervised</td>
<td>Cai et. al., [52]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S</td>
<td>MRI → CT</td>
<td>Abdomen</td>
<td>Unsupervised</td>
<td>Huo et. al., [12]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S</td>
<td>MRI → CT, MRI (T1 → T2)</td>
<td>Abdomen, Brain</td>
<td>Unsupervised</td>
<td>Liu et. al., [56]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S</td>
<td>CT → MRI</td>
<td>Hip, Thigh, Pelvis</td>
<td>Unsupervised</td>
<td>Hiasa et. al., [53]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>MRI → PET</td>
<td>Brain</td>
<td>Supervised</td>
<td>Pan et. al., [30]</td>
</tr>
<tr>
<td></td>
<td>Unidirectional</td>
<td>S</td>
<td>MRI → CT</td>
<td>Brain</td>
<td></td>
<td>Zhao et. al., [47]</td>
</tr>
<tr>
<td></td>
<td>3DR</td>
<td>CT → 3D rendered depth-map</td>
<td>Colon</td>
<td></td>
<td>Unsupervised</td>
<td>Mahmood et. al., [39]</td>
</tr>
<tr>
<td>LFST</td>
<td>Divergence</td>
<td>C</td>
<td>MRI → PET</td>
<td>Brain</td>
<td>Unsupervised</td>
<td>Zhu et. al., [71]</td>
</tr>
<tr>
<td></td>
<td>Adversarial</td>
<td>S</td>
<td>CT → MRI</td>
<td>Abdomen</td>
<td>Unsupervised</td>
<td>Yang et. al., [102]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S</td>
<td>MRI → CT</td>
<td>Heart</td>
<td>Unsupervised</td>
<td>Dou et. al., [103]</td>
</tr>
<tr>
<td></td>
<td>Reconstruction</td>
<td>S (3D)</td>
<td>MRI → CT</td>
<td>Heart</td>
<td></td>
<td>Ouyang et. al., [81]</td>
</tr>
<tr>
<td><strong>Pathology</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single Modality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DT</td>
<td>Unidirectional</td>
<td>C</td>
<td>WSI (different H&amp;E stains)</td>
<td>Breast, Colon, Ovary</td>
<td>Unsupervised</td>
<td>Bentaieb et. al., [29]</td>
</tr>
<tr>
<td>LFST</td>
<td>Adversarial</td>
<td>C</td>
<td>WSI (different stains)</td>
<td>Breast</td>
<td>Unsupervised</td>
<td>Lafarge et. al., [76]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>WSI (cross-institutional)</td>
<td>Prostate</td>
<td>Unsupervised</td>
<td>Ren et. al., [107]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S</td>
<td>WSI (different stains)</td>
<td>Colon</td>
<td>Supervised</td>
<td>Hou et. al., [110]</td>
</tr>
<tr>
<td></td>
<td>Divergence</td>
<td>S</td>
<td>Microscopy (different specimens)</td>
<td>Brain</td>
<td>Unsupervised</td>
<td>Chacon et. al., [74]</td>
</tr>
<tr>
<td><strong>Cross-Modality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DT</td>
<td>Reconstruction</td>
<td>S</td>
<td>WSI (CK → PD-L1)</td>
<td>Lung</td>
<td>Unsupervised</td>
<td>Kapil et. al., [101]</td>
</tr>
<tr>
<td></td>
<td>Unidirectional</td>
<td>S</td>
<td>WSI (H&amp;E ↔ IF)</td>
<td>Breast, Bladder, Lung</td>
<td>Unsupervised</td>
<td>Brieu et. al., [16]</td>
</tr>
<tr>
<td>LFST</td>
<td>Adversarial</td>
<td>C</td>
<td>WSI → Microscopy</td>
<td>Colon</td>
<td></td>
<td>Zhang et. al., [5]</td>
</tr>
<tr>
<td><strong>Ophthalmology</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single Modality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DT</td>
<td>Unidirectional</td>
<td>S</td>
<td>Retinal fundus (multiple datasets)</td>
<td>Retina</td>
<td>Unsupervised</td>
<td>Zhao et. al., [111]</td>
</tr>
<tr>
<td>LFST</td>
<td>Adversarial</td>
<td>S</td>
<td>Retinal fundus (multiple datasets)</td>
<td>Retina</td>
<td>Unsupervised</td>
<td>Javanmardi et. al., [105]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S</td>
<td>Retinal fundus (multiple datasets)</td>
<td>Retina</td>
<td></td>
<td>Wang et. al., [108]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Divergence</td>
<td>S</td>
<td>Retinal fundus (multiple datasets)</td>
<td>Retina</td>
<td></td>
</tr>
</tbody>
</table>
5 Conclusions and Future Directions

Deep learning has been widely applied to medical imaging data analysis. However the lack of well-annotated images and the heterogeneity of multi-center medical imaging datasets are two key challenges for DL performance. DA has emerged as an effective approach for minimizing domain-shift and leveraging labeled data from distinct but related domains. FST-DA and DT-DA are two popular approaches to minimize the distribution divergence in multiple medical imaging studies exploring same-modality or cross-modality scenarios. They have proven to achieve good performance, particularly in unsupervised DA settings and organ segmentation tasks. Current approaches are primarily adversarial with domains being selected based on certain heuristics and underlying tasks. Extensive benchmarking studies are needed to quantify the domain relationship for different imaging modalities and to compare adversarial and non-adversarial approaches. Varying sample transferability and multi-modal domains for medical imaging are two other major issues. One strategy is to explore down-weighting or attention-based networks. Also alternative multi-modal frameworks such as MUNIT [30] can be explored. Finally, for certain application areas in medical imaging such as 3D reconstruction and temporal disease analysis where DA is relatively unexplored, synthetic data can be used.

Acknowledgments

The work was supported in part by grants from the National Science Foundation EAGER Award NSF1651360, Children’s Healthcare of Atlanta and Georgia Tech Partnership Grant, Giglio Breast Cancer Research Fund, Petit Institute Faculty Fellow Award, and Carol Ann and David D. Flanagan Faculty Fellow Research Fund for Professor May D Wang. This work was also supported in part by the scholarship from China Scholarship Council (CSC) under the Grant CSC NO. 201406010343. The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References


