Unsupervised Cross-domain Image Classification by Distance Metric Guided Feature Alignment

Qingjie Meng[(✉)], Daniel Rueckert, and Bernhard Kainz

Department of Computing, BioMedIA, Imperial College, London, UK
q.meng16@imperial.ac.uk

Abstract. Learning deep neural networks that are generalizable across different domains remains a challenge due to the problem of domain shift. Unsupervised domain adaptation is a promising avenue which transfers knowledge from a source domain to a target domain without using any labels in the target domain. Contemporary techniques focus on extracting domain-invariant features using domain adversarial training. However, these techniques neglect to learn discriminative class boundaries in the latent representation space on a target domain and yield limited adaptation performance. To address this problem, we propose distance metric guided feature alignment (MetFA) to extract discriminative as well as domain-invariant features on both source and target domains. The proposed MetFA method explicitly and directly learns the latent representation without using domain adversarial training. Our model integrates class distribution alignment to transfer semantic knowledge from a source domain to a target domain. We evaluate the proposed method on fetal ultrasound datasets for cross-device image classification. Experimental results demonstrate that the proposed method outperforms the state-of-the-art and enables model generalization.

1 Introduction

Despite the success of deep neural networks (DNNs) for medical imaging applications [4,11,21,26,27], learning a task-specific model which generalizes to various medical datasets remains a challenge. This is due to the difference of feature distributions between different datasets, which is known as domain shift [29]. In medical imaging, domain shift can result from different imaging modalities (*e.g.*, magnetic resonance imaging and ultrasound) or different image acquisition devices. In this paper, we focus on model generalization between different image acquisition devices, transferring knowledge from a source device domain to a target device domain.

Fine-tuning DNNs on labelled data from the target domain is a possible solution but is often infeasible due to the need for sufficient manual annotations. More importantly, fine-tuned models remain domain specific because performance gains do not propagate back to the source domain. Deep domain

adaptation has been widely studied for tackling the problem of domain shift by extracting domain-invariant features [13,15,22]. Such approaches allow porting DNNs to the target domain without extensive annotation as well as preserving performance in both source and target domains. Unsupervised domain adaptation aims at transferring knowledge from a labeled source domain to an unlabeled target domain where both domains share a common label space [13,20,25]. This setting is important for real-world medical imaging scenarios, where data annotation is laborious, time consuming and requires rare expertise is available.

In this work, we propose distance metric guided feature alignment (MetFA) to learn a domain-invariant latent representations for model generalization in an unsupervised domain adaptation setting. We evaluate the proposed method on a challenging medical application, the classification of standardized diagnostic fetal ultrasound (US) view planes during prenatal screening. In many countries, fetal US is clinical routine for early detection of pathological development and informs subsequent decisions about treatment and delivery options [31]. However, domain shift caused by different acquisition devices and prohibitively expensive data annotation restricts the generalization of vanilla DNN classifiers. We show that MetFA enables unsupervised cross-device classification in fetal US.

Contribution. The main contributions of this paper are: (1) We propose distance metric guided feature alignment (MetFA), which learns a shared latent representation space between a labeled source domain and an unlabeled target domain; (2) we develop a framework that jointly learns class distribution alignment and MetFA, which further transfers semantic knowledge from a source domain to a target domain for model generalization; (3) we utilize the proposed method for cross-device anatomical classification in fetal US, which is an important medical imaging application that inherently requires knowledge transfer between different device domains to facilitate the use of DNNs for large scale population screening (Codes in https://github.com/qingjie99/MetFA).

Related Work. *Unsupervised domain adaptation (UDA)* mainly focuses on feature distribution alignment. Most UDA approaches explore an appropriate metric to measure the distance of feature distributions between two domains and subsequently train DNNs to minimize this distance [33]. Previous work such as Maximum Mean Discrepancy [22,35] utilizes kernels to measure the discrepancy between representations. Recent research explores domain adversarial training, where a domain discriminator is used to estimate this discrepancy while a feature extractor tries to deceive the discriminator by learning domain-invariant representations [2,23,24]. UDA has been applied to various medical imaging applications such as anatomical segmentation [3,6,9,17,28] and diagnostic classification [1,16]. Most of these works utilize domain adversarial training for feature alignment. In contrast to these works, we explicitly manipulate the latent space to learn discriminative features. Our work is inspired by MiniMax Entropy (MME) proposed in [30], which estimates domain-invariant prototypes and clusters target domain features around these prototypes in a *semi-supervised domain adaptation setting*. In contrast to [30], our method (1) embeds extracted features into a shared latent space with a fixed prior distribution before prototypes are

estimated, and (2) simultaneously reduces intra-class variance while increasing inter-class variance across domains via cross-domain metric learning.

Metric learning aims at learning embedded representations that cluster similar samples while separating dissimilar samples in latent space [37]. Previous metric learning methods measure feature similarity by learning a linear Mahalanobis distance [19,36]. More recent works focus on deep metric learning, which learns non-linear relationships of data using DNNs with different losses, such as contrastive loss [8,14], triplet loss [5,36] and N-pair loss [32]. Deep metric learning has shown great benefits for domain adaptation. For example, Sohn et al. [33] proposed a deep metric learning method for unsupervised domain adaptation in disjoint label space. Dou et al. [10] introduced deep metric learning for domain generalization. Most existing metric-learning-based domain adaptation methods only utilize metric learning on the labeled source domain and neglect the relationship between intra-class samples. In contrast to these methods, we introduce cross-domain metric learning to (1) jointly measure the similarity between samples in a labeled source domain and an unlabeled target domain and (2) learn metrics between different groups of intra-class samples.
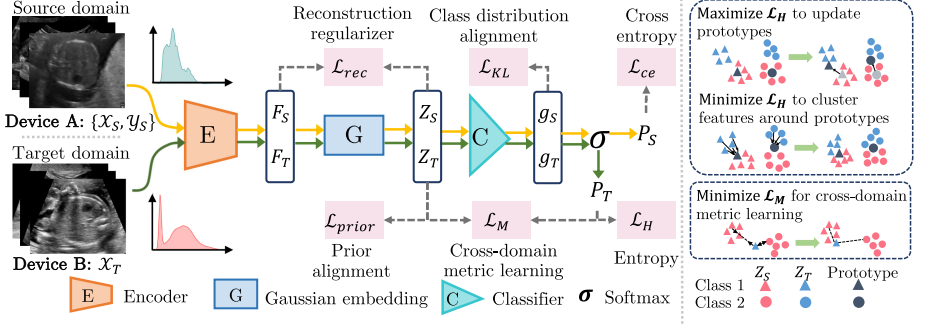
2   Method

We are given images and the corresponding labels from a source domain $\mathcal{D}_S = \{\mathcal{X}_S, \mathcal{Y}_S\}$ as well as unlabeled images from a target domain $\mathcal{D}_T = \{\mathcal{X}_T\}$. Both domains share a common label space and contain M classes. Our goal is to classify unlabeled target domain data by aligning latent features of both domains. The proposed method contains three main parts (see Fig. 1): (1) supervised classification on the labeled source domain, (2) distance metric guided feature alignment (MetFA) to transfer knowledge from the source domain to the target domain, and (3) class distribution alignment to preserve source domain class relationships in the target domain.

Classification. Classification in the unlabeled target domain is guided by the labeled source domain by sharing whole networks including an encoder E, a Gaussian embedding G and a classifier C. The cross-entropy loss is

$$\mathcal{L}_{ce} = -\mathbb{E}_{\{\mathbf{x},y\} \sim \{\mathcal{X}_S, \mathcal{Y}_S\}} \sum_{t=1}^{M} \mathbb{1}[y = t] log(C(G(E(\mathbf{x})))). \tag{1}$$

Classifier C simultaneously predicts class distributions for the target domain as $P_T(\hat{y}|\mathbf{x})|_{\mathbf{x} \in \mathcal{X}_T}$ (abbreviated as P_T). This prediction will be utilized in MetFA.

MetFA: Distance Metric Guided Feature Alignment. Feature embedding is used to constrain features from both domains to lie in a shared latent space. In this latent space, class representations (prototypes) are estimated to extract domain-invariant features in each class, while cross-domain metric learning is introduced to further separate clusters of different classes in both domains.

Fig. 1. Left: An overview of the proposed method. Our method consists of (1) supervised classification on the labeled source domain (optimize $\mathcal{L}_{ce}$), (2) distance metric guided feature alignment (MetFA), which aligns features between both domains (optimize $\mathcal{L}_{prior}$, $\mathcal{L}_H$, $\mathcal{L}_M$, $\mathcal{L}_{rec}$), and (3) class distribution alignment, which preserves class relationships in both domains (optimize $\mathcal{L}_{KL}$). Right: Schematic illustration of $\mathcal{L}_H$ and $\mathcal{L}_M$ optimization.

Feature embedding encourages features (F_S, F_T) extracted by an encoder E to share the same fixed prior distribution in a latent space $\mathcal{Z}$, which is similar to distribution matching in a variational autoencoder [18]. In our method, a Gaussian embedding G is built to model F_S and F_T by a standard Gaussian distribution $\mathcal{N}(0, I)$. Specifically, $Z_i \sim q(\mathcal{Z}|\mathcal{X}_i)|i \in \{S,T\}$ is sampled from $\mathcal{N}(\mu_i, \Sigma_i)|i \in \{S,T\}$ with the reparameterization trick [18], where $\{\mu_i, \Sigma_i\} = G(F_i)|i \in \{S,T\}$. The prior alignment loss is the Kullback-Leibler (KL) divergence between $\mathcal{N}(0, I)$ and $\mathcal{N}(\mu_i, \Sigma_i)|i \in \{S,T\}$, which is

$$\mathcal{L}_{prior} = D_{KL}(\mathcal{N}(\mu_S, \Sigma_S) \parallel \mathcal{N}(0,I)) + D_{KL}(\mathcal{N}(\mu_T, \Sigma_T)) \parallel \mathcal{N}(0,I)). \qquad (2)$$

In order to guarantee that embedded features are representative of the extracted features, we add a feature reconstruction loss $\mathcal{L}_{rec}$ as a regularizer:

$$\mathcal{L}_{rec} = \|F_S - Z_S\|_2^2 + \|F_T - Z_T\|_2^2. \qquad (3)$$

Feature embedding constrains distribution matching. In the absence of target domain labels, it is essential for subsequent feature alignment. However, feature embedding itself is unlikely to ensure that features are domain-invariant and discriminative between different classes. The rest of MetFA tackles this problem.

Domain-invariant feature extraction is motivated by Minimax Entropy (MME), proposed by Saito et al. [30]. Using unlabeled data in the target domain, MME learns a single domain-invariant prototype (a representation point) for each class in both domains and clusters target domain samples around these prototypes (see Fig. 1 upper right). We implement prototypes as the weights $\mathbf{W}$ of the last dense layer in the classifier C.

Training MME contains two iterative steps. The first step is to move prototypes from source domain to target domain, which is maximizing the similarity

between $\mathbf{W}$ and its input features (H_T). This similarity maximization is equivalent to maximizing the entropy of $\mathcal{X}_T$ with respect to $\mathbf{W}$, using

$$\mathcal{L}_H = -\mathbb{E}_{\mathbf{x} \sim \mathcal{X}_T} \sum_{i=1}^{M} p_T(\hat{y} = i|\mathbf{x}) \log p_T(\hat{y} = i|\mathbf{x}), \ p_T \in P_T = \sigma(\frac{1}{\tau_0} \frac{\mathbf{W}^T H_T}{\|H_T\|}), \ (4)$$

where σ is a softmax function and τ_0 is a temperature parameter. The second step is to assign target domain features to the domain-invariant prototypes. To achieve this, $\mathcal{L}_H$ is minimized with respect to E, G and $C \setminus \mathbf{W}$ (C without $\mathbf{W}$).

Cross-domain metric learning is proposed to maximize the margin between different classes across domains. We define latent features of $\mathcal{X}_S$ and $\mathcal{X}_T$ (which are Z_S and Z_T) respectively as support samples and query samples. In cross-domain metric learning the distance between query and support samples is minimized when they are from the same class and simultaneously maximized when they are from different classes (see Fig. 1 lower right). The metric loss is

$$\mathcal{L}_M = \frac{1}{N} \sum_{i=1}^{M} \sum_{j=1}^{c_i^T} \log(1 + \sum_{k \neq i}^{k \in [1,M]} e^{d_j^i - d_j^k}) = -\frac{1}{N} \sum_{i=1}^{M} \sum_{j=1}^{c_i^T} \log \frac{e^{d_j^i}}{e^{d_j^i} + \sum_{k \neq i}^{k \in [1,M]} e^{d_j^k}}, \tag{5}$$

where N and c_i^T are the number of all query samples and query samples from class i. Note that the labels of query samples are P_T in Eq. 4. d_j^i is the distance between a query sample q_j^i and a same class support sample s_t^i. d_j^k is the distance between q_j^i and s_t^k from different classes. Considering the relationship between intra-class samples and using a hard mining strategy [7], we define d_j^i and d_j^k as

$$d_j^i = \max_t d(q_j^i, s_t^i), \ t \in [1, c_i^S], \ q_j^i \sim Z_T, \ s_t^i \sim Z_S,$$
$$d_j^k = \min_t d(q_j^i, s_t^k), \ t \in [1, c_k^S], \ q_j^i \sim Z_T, \ s_t^k \sim Z_S, \tag{6}$$

where c_i^S and c_k^S are the number of support samples from class i and class k. We use the squared Euclidean distance for $d(\cdot, \cdot)$ in Eq. 6.

Class Distribution Alignment. Apart from structuring a feature space for better class predictions, we want to further transfer semantic knowledge which is preserving class relationships between domains. Class distribution alignment is used for class relationship preservation between multiple labeled source domains in a domain generalization task [10]. In our method, we align class distributions between a labeled source domain and an unlabeled target domain. We utilize the symmetrized KL-divergence to define the class distribution alignment loss

$$\mathcal{L}_{KL} = \frac{1}{M} \sum_{i=1}^{M} \Lambda[D_{KL}(\bar{p}_i^S \| \bar{p}_i^T) + D_{KL}(\bar{p}_i^T \| \bar{p}_i^S)],$$
$$\bar{p}_i^S = \sigma(\frac{1}{\tau_1} \frac{1}{c_i^S} \sum_{y=i} g_{\mathbf{x}}^S)|_{(\mathbf{x},y) \sim \{\mathcal{X}_S, \mathcal{Y}_S\}}, \ \bar{p}_i^T = \sigma(\frac{1}{\tau_1} \frac{1}{c_i^T} \sum_{\hat{y}=i} g_{\mathbf{x}}^T)|_{(\mathbf{x},\hat{y}) \sim \{\mathcal{X}_T, P_T(\mathbf{x})\}}. \tag{7}$$

Here, $\Lambda = [c_1^T, c_2^T, ..., c_M^T]$ contains the number of target domain samples predicted for each class. $\bar{p}_i^S$ and $\bar{p}_i^T$ are the ith class distributions in source and target domain. $g_\mathbf{x}^S$ and $g_\mathbf{x}^T$ are the pre-softmax activations from classifier C and τ_1 is a temperature parameter.

Optimization. The overall objective function of the proposed method is:

$$\min_{E,G,C \backslash \mathbf{W}} \{\mathcal{L} + \lambda_6 \mathcal{L}_H\}, \quad \min_{\mathbf{W}} \{\mathcal{L} - \lambda_6 \mathcal{L}_H\},$$
$$\text{with} \quad \mathcal{L} = \lambda_1 \mathcal{L}_{ce} + \lambda_2 \mathcal{L}_{prior} + \lambda_3 \mathcal{L}_M + \lambda_4 \mathcal{L}_{rec} + \lambda_5 \mathcal{L}_{KL}. \tag{8}$$

Here λ_1 to λ_6 are hyper-parameters chosen experimentally depending on the application. Our model is end-to-end trainable, with $\mathbf{W}$ and the rest of the networks are trained in an alternating fashion according to Eq. 8. We apply L2 regularization (scale = 10^{-5}) to all weights during training to prevent overfitting and apply random image flipping as data augmentation. Our model is trained on a Nvidia Titan X GPU.

3   Evaluation and Results

We evaluate the proposed method on 2D fetal US images acquired during routine prenatal screening. This US data is obtained by different imaging devices: Device A (GE Voluson E8) acquires $\sim 12k$ images and device B (Philips EPIQ V7 G) acquires unpaired $\sim 5.5k$ images. In both datasets, six anatomical standard planes have been selected by expert sonographers, including Four Chamber View (4CH), Abdominal, Left Ventricular Outflow Tract (LVOT), Right Ventricular Outflow Tract (RVOT), Femur and Lips. We evaluate our method in two scenarios where device A is the source domain while device B is the target domain, and vice versa. During training, the source domain is fully labeled and the target domain is unlabeled. In both scenarios, hyper-parameters λ_1 to λ_6 in Eq. 8 are $\lambda_1 = 10, \lambda_2 = 10^{-2}, \lambda_3 = 10^{-1}, \lambda_4 = 1, \lambda_5 = 10, \lambda_6 = 5$. τ_0 in Eq. 4 is 0.05 (same to [30]) and τ_1 in Eq. 7 is 2 (same to [10]). We use Stochastic Gradient Descent (SGD) with momentum optimizer to update our model.

Comparison Methods. We evaluate a VGG network which contains an encoder E and a classifier C from the proposed method as a baseline. This baseline is trained on data only from the source domain (*Source only*) to demonstrate the existence of domain shift. We compare the proposed method with the state-of-the-art domain-adaptation algorithms, including domain-adversarial training of neural networks (DANN) [13], adversarial discriminative domain adaptation (ADDA) [34] and semi-supervised domain adaptation via minimax entropy (MME) [30]. Note that for fair comparison, we use the MME model in an unsupervised learning paradigm. Additionally, given target domain labels, we show fine-tuned and fully-supervised classification on the target domain as references. Fine-tuned classification is pre-trained on the labeled source domain and fine-tuned on the labeled target domain. This fine-tuned model is evaluated on both

source and target domains. Fully-supervised classification is trained from scratch on the labeled target domain and evaluated on the target domain.

Ablation Study. We further explore the effectiveness of different components in the proposed method by removing different loss components: UDA-MetFA-I: only contains $\mathcal{L}_{ce}$, $\mathcal{L}_{prior}$ and $\mathcal{L}_H$; UDA-MetFA-II: UDA-MetFA-I plus $\mathcal{L}_M$; UDA-MetFA-III: UDA-MetFA-II plus $\mathcal{L}_{KL}$; UDA-MetFA-IV: UDA-MetFA-II plus $\mathcal{L}_{rec}$; UDA-MetFA-V: contains all components.

Table 1. Comparison of *Source only*, the state-of-the-art and ablation study (UDA-MetFA- I to V) for fetal US anatomical classification with **device A as source domain and device B as target domain**. Fine-tuned and Fully-supervised are reference results given target domain labels. Best results in bold.

Methods	S: device A			T: device B		
	F1-score	Recall	Precision	F1-score	Recall	Precision
Source only	0.8782	0.8800	0.8786	0.2455	0.3400	0.3070
ADDA [34]	0.8841	0.8850	0.8860	0.1377	0.2050	0.1623
DANN [13]	0.8321	0.8350	0.8564	0.3390	0.3650	0.3756
MME [30]	0.8876	0.8900	0.8914	0.4398	0.5133	0.4565
UDA-MetFA-I	0.8894	0.8900	0.8911	0.5255	0.5550	0.5599
UDA-MetFA-II	0.8951	0.8967	0.8997	0.5959	0.6400	0.6359
UDA-MetFA-III	**0.9202**	**0.9200**	**0.9207**	0.6301	0.6850	0.6143
UDA-MetFA-IV	0.8970	0.8967	0.8986	0.6930	0.7067	0.7011
UDA-MetFA-V	0.8990	0.9000	0.9027	**0.7713**	**0.7717**	**0.7874**
Fine-tuned	0.7987	0.8050	0.8140	0.7114	0.7150	0.7373
Fully-supervised	–	–	–	0.5919	0.6100	0.6576

Results. Table 1 shows the experimental results of baselines and the ablation study where device A is the source domain and device B is the target domain. From this table, we observe that the UDA-MetFA-V model outperforms other baselines. In the target domain, UDA-MetFA-V achieves an average F1-score of 0.7713 while the highest average F1-score of other baselines is 0.4398 (MME [30]). UDA-MetFA-I greatly outperforms MME [30] in the target domain, demonstrating the importance of feature embedding in the proposed method. UDA-MetFA-V performs better than other ablation models in the target domain, illustrating the effectiveness of all components in the proposed method. Furthermore, the results of Fine-tuned and *source only* in the source domain indicate that the fine-tuned model remains less generalizable, whereas the proposed method (UDA-MetFA-V) enables model generalization with improved classification performance in both source and target domains.

We further compare MME (best baseline in Table 1) with the proposed method (UDA-MetFA-V) in confusion matrices and t-SNE plots. Figure 2(a) demonstrates that our method extracts more discriminative features for better classification, especially on easily confused anatomies (*e.g.*, LVOT vs. RVOT). Figure 2(b) shows that for UDA-MetFA-V, target features are closer to source features while features of different classes are more separated. This indicates that the proposed MetFA benefits the extraction of discriminative and domain-invariant features.

Table 2 shows the results of comparison methods and the proposed method (UDA-MetFA-V) on switched domains, where device B is the source domain and device A is the target domain. We observe that UDA-MetFA-V outperforms the state-of-the-art in both source and target domains, demonstrating that our method is capable of successfully transferring knowledge from source domain to target domain as well as improving model generalization.

MME [30]	UDA-MetFA-V	MME [30]	UDA-MetFA-V

(a) Confusion Matrix   (b) t-SNE visualization
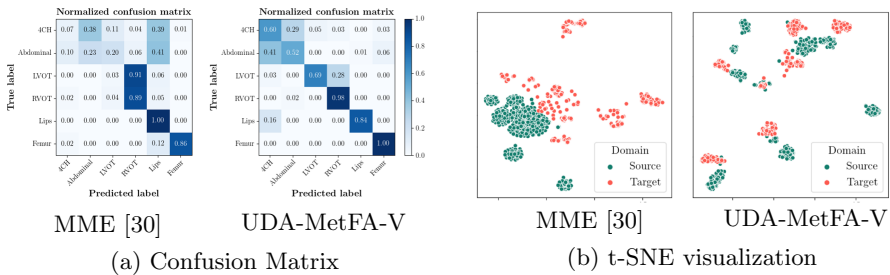
Fig. 2. Comparison of MME [30] and UDA-MetFA-V on (a) confusion matrix of target domain (device B) and (b) t-SNE plot of extracted test data features.

Table 2. Comparison of baselines and UDA-MetFA-V with **device B as source domain and device A as target domain**. Best results in bold.

Methods	S: device B			T: device A		
	F1-score	Recall	Precision	F1-score	Recall	Precision
Source only	0.5919	0.6100	0.6576	0.2854	0.3300	0.3555
DANN [13]	0.5198	0.5450	0.5451	0.3318	0.3500	0.3450
MME [30]	0.3776	0.4183	0.4500	0.1520	0.1883	0.2101
UDA-MetFA-V	**0.7101**	**0.7150**	**0.7441**	**0.5776**	**0.5550**	**0.6303**
Fully-supervised	–	–	–	0.8782	0.8800	0.8786

Discussion. Domain adaptation is commonly used to transfer a performant, task-specific model from a source domain to a target domain. However, the DNNs learning ability in a source domain can limit this ability in a target domain. This may explain the lower classification performance of the proposed method compared with a fully-supervised method in the target domain in Table 2. Current UDA methods rarely discuss the performance of DNNs in the source domain. From Table 2, we observe that tracking the source domain performance can be potentially used for data selection during model improvement in the source domain. A limitation of our method is the empirical hyper-parameters selection. For a specific application, we adjust hyper-parameters according to their importance and select the best combination with grid search. Meta-learning [12] will be explored in future work to allow automatic hyper-parameter selection.

4   Conclusion

In this paper, we discuss the problem of model generalization for unsupervised domain adaption. We propose metric learning for improved feature alignment (MetFA) to extract discriminative and domain-invariant features across domains. MetFA explicitly structures latent representations without using domain adversarial training. Our model integrates class distribution alignment for transferring semantic knowledge from a source domain to a target domain. Experiments on cross-device fetal US screening images demonstrate the effectiveness and practical applicability of our method compared with the state-of-the-art.

Appendices

A   Examples of Ultrasound Images

We show more examples of ultrasound images acquired from different image acquisition devices (Fig. 3).

(a) 4CH

(b) Abdominal

(c) Femur

(d) Lips

(e) LVOT

(f) RVOT
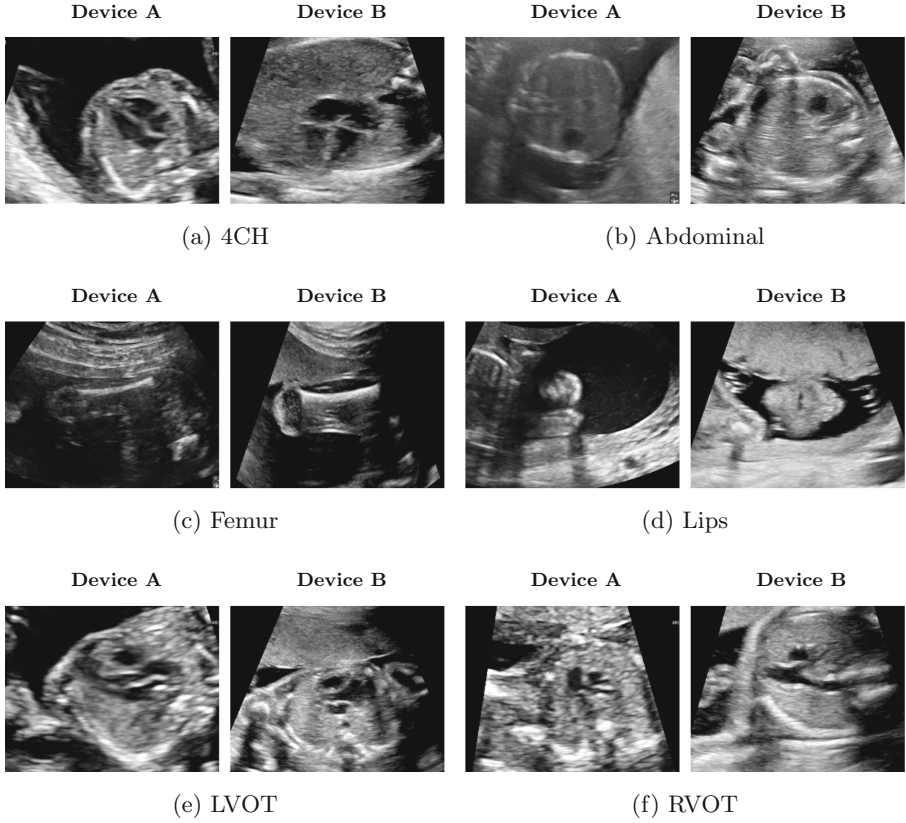
Fig. 3. Examples of ultrasound images acquired by different image acquisition devices. Device A is *GE Voluson E8* and device B is *Philips EPIQV7 G*.

B   Split of Training Data

See Table 3.

Table 3. The number of images in each class for training. In the first scenario (S: device A, T: device B), images in device A are used as labeled data and images in device B are unlabeled. In the second scenario (S: device B, T: device A), images in device B are labeled and images in device A are unlabeled.

	4CH	Abdominal	Femur	Lips	LVOT	RVOT
Device A	700	700	700	700	700	700
Device B	828	728	815	600	328	559

References

1. Bernard, O., et al.: Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE Trans. Med. Imaging **37**, 2514–2525 (2018)
2. Bousmalis, K., Trigeorgis, G., Silberman, N., Krishnan, D., Erhan, D.: Domain separation networks. In: NeurIPS, pp. 343–351 (2016)
3. Cai, J., Zhang, Z., Cui, L., Zheng, Y., Yang, L.: Towards cross-modal organ translation and segmentation: a cycle- and shape-consistent generative adversarial network. Med. Image Anal. **52**, 174–184 (2019)
4. Chartsias, A., Joyce, T., Giuffrida, M.V., Tsaftaris, S.A.: Multimodal MR synthesis via modality-invariant latent representation. IEEE Trans. Med. Imaging **37**, 803–814 (2017)
5. Chechik, G., Sharma, V., Shalit, U., Bengio, S.: Large scale online learning of image similarity through ranking. J. Mach. Learn. Res. **11**, 1109–1135 (2010)
6. Chen, C., Dou, Q., Chen, H., Qin, J., Heng, P.: Synergistic image and feature adaptation: towards cross-modality domain adaptation for medical image segmentation. In: AAAI, pp. 865–872 (2019)
7. Chen, G., Zhang, T., Lu, J., Zhou, J.: Deep meta metric learning. In: ICCV (2019)
8. Chopra, S., Hadsell, R., LeCun, Y.: Learning a similarity metric discriminatively, with application to face verification. In: CVPR, pp. 539–546 (2005)
9. Dong, N., Kampffmeyer, M., Liang, X., Wang, Z., Dai, W., Xing, E.: Unsupervised domain adaptation for automatic estimation of cardiothoracic ratio. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11071, pp. 544–552. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00934-2_61
10. Dou, Q., de Castro, D.C., Kamnitsas, K., Glocker, B.: Domain generalization via model-agnostic learning of semantic features. In: NeurIPS (2019)
11. Dou, Q., Liu, Q., Heng, P.A., Glocker, B.: Unpaired multi-modal segmentation via knowledge distillation. IEEE Trans. Med. Imaging **38** (2019)
12. Feurer, M., Klein, A., Eggensperger, K., Springenberg, J., Blum, M., Hutter, F.: Efficient and robust automated machine learning. In: NeurIPS, pp. 2962–2970 (2015)
13. Ganin, Y., et al.: Domain-adversarial training of neural networks. J. Mach. Learn. Res. **17**(1), 2030–2096 (2016)
14. Hadsell, R., Chopra, S., LeCun, Y.: Dimensionality reduction by learning an invariant mapping. In: CVPR, pp. 1735–1742 (2006)
15. Häusser, P., Frerix, T., Mordvintsev, A., Cremers, D.: Associative domain adaptation. In: ICCV, pp. 2784–2792 (2017)
16. Huang, Y., Zheng, H., Liu, C., Ding, X., Gustavo, R.K.: Epithelium-stroma classification via convolutional neural networks and unsupervised domain adaptation in histopathological images. IEEE J. Biomed. Health Inform. **21**, 1625–1632 (2017)
17. Kamnitsas, K., et al.: Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. In: Niethammer, M., et al. (eds.) IPMI 2017. LNCS, vol. 10265, pp. 597–609. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59050-9_47
18. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. In: Bengio, Y., LeCun, Y. (eds.) ICLR (2014)
19. Köstinger, M., Hirzer, M., Wohlhart, P., Roth, P.M., Bischof, H.: Large scale metric learning from equivalence constraints. In: CVPR, pp. 2288–2295 (2012)

20. Lee, C.Y., Batra, T., Baig, M.H., Ulbricht, D.: Sliced Wasserstein discrepancy for unsupervised domain adaptation. In: CVPR, pp, 10285–10295 (2019)
21. Liu, M., Zhang, J., Adeli, E., Shen, D.: Landmark-based deep multi-instance learning for brain disease diagnosis. Med. Image Anal. **43**, 157–168 (2018)
22. Long, M., Cao, Y., Wang, J., Jordan, M.I.: Learning transferable features with deep adaptation networks. In: ICML, pp. 97–105 (2015)
23. Long, M., CAO, Z., Wang, J., Jordan, M.I.: Conditional adversarial domain adaptation. In: NeurIPS, pp. 1640–1650 (2018)
24. Luo, Z., Zou, Y., Hoffman, J., Fei-Fei, L.F.: Label efficient learning of transferable representations across domains and tasks. In: NeurIPS, pp. 165–177 (2017)
25. Meng, Q., Rueckert, D., Kainz, B.: Learning cross-domain generalizable features by representation disentanglement. arXiv: 2003.00321 (2020)
26. Meng, Q., Sinclair, M., Zimmer, V., et al.: Weakly supervised estimation of shadow confidence maps in fetal ultrasound imaging. IEEE Trans. Med. Imaging **38**, 2755–2767 (2019)
27. Nie, D., Zhang, H., Adeli, E., Liu, L., Shen, D.: 3D deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 212–220. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_25
28. Ouyang, C., Kamnitsas, K., Biffi, C., Duan, J., Rueckert, D.: Data efficient unsupervised domain adaptation for cross-modality image segmentation. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11765, pp. 669–677. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32245-8_74
29. Quinonero-Candela, J., Sugiyama, M., Schwaighofer, A., Lawrence, N.D.: Dataset Shift in Machine Learning. Neural Information Processing. MIT Press, Cambridge (2008)
30. Saito, K., Kim, D., Sclaroff, S., Darrell, T., Saenko, K.: Semi-supervised domain adaptation via minimax entropy. In: ICCV (2019)
31. Salomon, L.J., et al.: Practice guidelines for performance of the routine midtrimester fetal ultrasound scan. Ultrasound Obstet. Gynecol. **37**, 116–126 (2011)
32. Sohn, K.: Improved deep metric learning with multi-class N-pair loss objective. In: Lee, D.D., Sugiyama, M., von Luxburg, U., Guyon, I., Garnett, R. (eds.) NeurIPS, pp. 1849–1857 (2016)
33. Sohn, K., Shang, W., Yu, X., Chandraker, M.: Unsupervised domain adaptation for distance metric learning. In: ICLR (2019)
34. Tzeng, E., Hoffman, J., Saenko, K., Darrell, T.: Adversarial discriminative domain adaptation. In: CVPR, pp. 2962–2971 (2017)
35. Tzeng, E., Hoffman, J., Zhang, N., Saenko, K., Darrell, T.: Deep domain confusion: maximizing for domain invariance. arXiv:1412.3474, December 2014
36. Weinberger, K.Q., Saul, L.K.: Distance metric learning for large margin nearest neighbor classification. JMLR, 207–244 (2009)
37. Xing, E.P., Jordan, M.I., Russell, S.J., Ng, A.Y.: Distance metric learning with application to clustering with side-information. In: NeurIPS, pp. 521–528 (2003)