GCN Based Unsupervised Domain Adaptation With Feature Disentanglement For Medical Image Classification

Dwarikanath Mahapatra[1] DWARIKANATH.MAHAPATRA@INCEPTIONIAI.ORG
[1] *Inception Institute of AI, Abu Dhabi, UAE*

Steven Korevaar[2] STEVEN.KOREVAAR@STUDENT.RMIT.EDU.AU
[2] *School of Engineering, RMIT University, Melbourne, Australia. Australia.*

Ruwan Tennakoon[3] RUWAN.TENNAKOON@RMIT.EDU.AU
[2] *School of Computing Technologies, RMIT University, Melbourne, Australia.*

Editors: Under Review for MIDL 2022

Abstract

The success of deep learning has set new benchmarks for many medical image tasks. However, deep models often fail to generalize in the presence of distribution shifts between training (source) data and test (target) data. One method commonly employed to counter distribution shifts is domain adaptation: using samples from the target domain to learn to account for shifted distributions. In this work we propose an unsupervised domain adaptation approach that uses graph neural networks to learn semantic and structural features that are invariant across domains allowing for better performance across distribution shifts. We test the proposed method for classification on two challenging medical image datsets with distribution shifts - multi center chest xray images and histopathology images. Experiments show our method achieves state-of-the-art results on those data sets.

Keywords: unsupervised domain adaptation, Graph convolution networks, Camelyon17, CheXpert, NIH Xray.

1. Introduction

With the success of convolutional neural networks (CNNs) new benchmarks have been set for many medical image classification tasks such as diabetic retinopathy grading(Gulshan et al., 2016), digital pathology image classification (Liu et al., 2017) and chest X-ray images (Irvin et al., 2017; Wang et al., 2017), as well as segmentation tasks such as (Li et al., 2021; Painchaud et al., 2020). However, adopting such algorithms in clinical practice poses challenges due to the domain shift problem where the target dataset has different characteristics than the source dataset on which the model has been trained. These differences are most commonly seen in the numerous image capturing protocols, parameters, scanner manufacturers, and many other factors. This problem is particularly acute when the dataset has images from multiple facilities where these factors are not controlled. Since annotating samples from hospitals and domains is challenging due to scarcity of experts, it is essential to design models that perform consistently on images acquired from multiple sources.

Semi-supervised and Unsupervised Domain Adaptation (UDA) methods have been proposed to address the domain shift problem. Semi-supervised approaches assume the availability of few labeled instances from the target domain for training the model along with the source data (Puybareau et al., 2019). On the other hand, UDA techniques (Chen et al., 2020; Ouyang et al., 2019) do not rely on the availability of labels from the target domain. Generally, the goal of UDA methods is to learn a domain-invariant representation by enforcing some constraint (e.g. Maximum Mean Discrepancy (Kumagai and Iwata, 2019)) that brings the latent space, z, of the two domains closer and thus more alike in distribution allowing for more comparable performance in classification/segmentation.

While state-of-the-art (SOTA) UDA methods have been able to use convolutional neural network (CNN) based architectures to obtain impressive results; these methods often only enforce alignment of global domain statistics (Xie et al., 2018) which leads to the loss of important semantic class label information. Semantic transfer methods (Luo et al., 2017; Motiian et al., 2017) address this by propagating the class label information into deep adversarial adaptation networks. Unfortunately, it is difficult to model and integrate semantic label transfer into existing deep networks. To deal with the above limitations (Ma et al., 2019) propose an end-to-end Graph Convolutional Adversarial Network (GCAN) for unsupervised domain adaptation. Graph based methods have another advantage over normal CNNs as they better exploit the global relationship between different nodes (or samples), and can effectively learn both global as well as local information. In this work we combine feature disentanglement with graph convolutional networks (GCN) for unsupervised domain adaptation and apply it to two different standard medical imaging datasets for classification and compare it to the current SOTA methods.

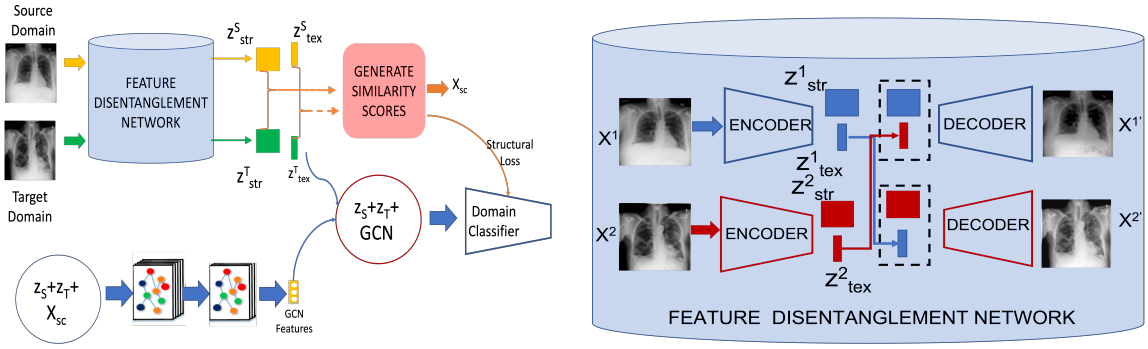
Related Work: Prior works on UDA focused on medical image classification (Bermúdez-Chacón et al., 2016; Ahn et al., 2020), object localisation and lesion segmentation (Heimann et al., 2013; Kamnitsas et al., 2017), and histopathology stain normalization (Chang et al., 2021).Graph networks for UDA (Ma et al., 2019; Wu et al., 2020) have found applications for medical imaging (Ahmedt-Aristizabal et al., 2021) such as brain surface segmentation (Gopinath et al., 2020) and brain image classification (Hong et al., 2019a,b).

1.1. Our Contributions:

Although previous works used feature disentanglement and graph based domain adaptation separately, they did not combine them for medical image classification. In this work: 1) we propose an end-end Graph Convolutional Adversarial Network (GCAN) for unsupervised domain adaptation in medical images. We seek to align features for domain and structure alignment. 2) We perform feature disentanglement using swapped autoencoders to obtain texture and structural features, which are used for graph construction and defining generator losses; 3) We demonstrate our method's effectiveness on multiple medical image datasets.

2. Method

Given a source data set, $\mathcal{D}_S = \{(x_S^i, y_S^i)\}_{i=1}^{n_s}$, consisting of n_s labeled samples, and a target data set, $\mathcal{D}_T = \{(x_T^i)\}_{i=1}^{n_t}$ consisting of n_t unlabeled target samples, unsupervised domain adaptation aims to learn a classifier that can reliably classify unseen target samples. Here, $x_S^i \sim p^S$ is a source data point sampled from source distribution p^S, $y_S^i \in \mathcal{Y}_S$ is the label,



(a) (b)

Figure 1: (a) Workflow of our proposed method. Given images from source and target domain we disentangle features into texture and structure features, and generate similarity scores to construct an adjacency matrix. Generated features by the GCN are domain invariant and along with the latent features are used to train a domain classifier. (b) Architecture of feature disentanglement network with swapped structure features

and $x_T^i \sim p^T$ is a target data point sampled from target distribution p^T. As per the covariate shift assumption, we assume that $p^S(y|x) = p^T(y|x)\ \forall x$. Thus the only thing that changes between source and target is the distribution of the input samples, x.

2.1. Graph Convolutional Adversarial Network

The core component of our method is a an adversarial generator based on graphs that generates features which are domain invariant. Our proposed approach consists of a feature disentanglement module which separates semantic (textural) and structural features (often called style and content respectively). The output of the feature disentanglement module is constructed into a graph and then fed into a graph neural network which will learn more global relationships between samples, which in turn leads to more discriminative and domain invariant feature learning. The divergence of domain statistics measured by the adversarial loss and other loss terms guides the feature extractor to learn domain-invariant representations. The architecture of our proposed approach is shown in Figure 1. Our network is trained by minimizing the following objective function:

$$\mathcal{L}(\mathcal{X}_S, \mathcal{Y}_S, \mathcal{X}_T) = \mathcal{L}_C(\mathcal{X}_S, \mathcal{Y}_S) + \lambda_1 \mathcal{L}_{Adv}(\mathcal{X}_S, \mathcal{X}_T) + \lambda_2 \mathcal{L}_{Str}(\mathcal{X}_S, \mathcal{X}_T) \tag{1}$$

Classification loss $\mathcal{L}_C$ is a cross entropy loss $\mathcal{L}_C(\mathcal{X}_S, \mathcal{Y}_S) = -\sum_{c=1}^{n_s} y_c \log(p_c)$, y_c is the indicator variable and p_c is the class probability. The other loss terms are described below.

2.1.1. Adversarial Loss

We use an adversarial loss function defined in Eq. 2. A domain classifier D identifies whether features from feature generator G are from the source or target domain. Conversely, the

generator G is being trained to produce samples that can fool D. This two-player minimax game will reach an equilibrium state when features from G are domain-invariant.

$$\mathcal{L}_{Adv}(\mathcal{X}_S, \mathcal{X}_T) = \mathbb{E}_{x \in D_S}[\log(1 - D(G(x)))] + \mathbb{E}_{x \in D_T}[\log(D(G(x)))] \tag{2}$$

2.1.2. Structure Feature Alignment

While the adversarial loss enforces alignment of global domain statistics we also want that the structural information of samples be preserved. Structural information is the local information that describes an image's underlying structure and visible organs or parts, and is expected to be invariant across similar images from different domains. We propose a feature disentanglement network that generates texture and structure features and we enforce that the structure feature representations be similar across different domains.

Feature Disentanglement: Figure 1 (b) shows the architecture of our network for feature disentanglement. Note that the feature disentanglement network is pre-trained on a separate set of source domain images with known label. Similar to a classic autoencoder, the encoder E produces a latent code $z \sim Z$ for image $x \sim X$. The G reconstructs the original image from z using an image reconstruction loss that is defined as:

$$\mathcal{L}_{Rec}(E, G) = \mathbb{E}_{x \sim X}\left[\|x - G(E(x))\|\right] \tag{3}$$

Additionally, the generated image should be realistic as determined by the Discriminator D and is enforced using the adversarial loss defined as:

$$\mathcal{L}_{Adv}(E, G, D) = \mathbb{E}_{x \sim X}\left[-\log(D(G(E(x))))\right] \tag{4}$$

Furthermore, as part of our objective to achieve feature disentanglement we decompose the latent code z into two components $[z_{str}, z_{tex}]$ corresponding to the structure and texture components. We enforce that keeping the structure component and swapping the texture component with a similar image still produces realistic images. This is achieved by using a modified version of the adversarial loss, which we term as the swapped GAN loss, and is defined as :

$$\mathcal{L}_{swap}(E, G, D) = \mathbb{E}_{x^1, x^2 \sim X, x^1 \neq x^2}\left[-\log(D(G(z^1_{tex}, z^2_{str})))\right] \tag{5}$$

Here z^1_{tex}, z^2_{str} are the first and second components of images X^1, X^2's latent representations, and X^1, X^2 from the same dataset in a minibatch. The component z_{str} is a tensor with spatial dimensions, while z_{tex} is a vector that encode structure and texture information. $\mathcal{L}_{Rec}$ and $\mathcal{L}_{Adv}$, are applied to image X^1 while $\mathcal{L}_{swap}$ is applied to the latent components from X^1, X^2. The final loss function for feature disentanglement is defined as

$$\mathcal{L}_{Disent} = \mathcal{L}_{Rec} + 0.7\mathcal{L}_{Adv} + 0.7\mathcal{L}_{swap} \tag{6}$$

After training is complete the network can take an input image and output its disentangled feature representations $[z^S_{str}, z^S_{tex}]$, $[z^T_{str}, z^T_{tex}]$ for source and target domain.To preserve similarity of structure features, we calculate the cosine similarity loss as

$$\mathcal{L}_{Str} = 1 - \langle z^S_{str}, z^T_{str} \rangle, \tag{7}$$

where $\langle . \rangle$ denotes cosine similarity. Note that the texture features need not be similar since they are domain specific features, and hence we don't define a texture loss.

The cosine similarity scores and the latent vectors are used to construct densely-connected instance graphs. The Graph Convolutional Network (GCN) (Kipf and Welling, 2017) is then applied to the instance graphs to learn GCN features encoded in the latent representations. The GCN aims to learn the layerwise propagation operations that can be applied directly on graphs. Given an undirected graph with m nodes, the set of edges between nodes, and an adjacency matrix $\mathbf{A} \in R^{m \times m}$, a linear transformation of graph convolution is defined as the multiplication of a graph signal $\mathbf{X} \in R^{k \times m}$ with a filter $\mathbf{W} \in R^{k \times c}$:

$$\mathbf{Z} = \hat{\mathbf{D}}^{-\frac{1}{2}} \hat{\mathbf{A}} \hat{\mathbf{D}}^{-\frac{1}{2}} \mathbf{X}^T \mathbf{W}, \tag{8}$$

where $\hat{\mathbf{A}} = \mathbf{A} + \mathbf{I}$, $\mathbf{I}$ being the identity matrix, and $\hat{\mathbf{D}}_{ii} = \sum_j \hat{\mathbf{A}}_{ij}$. The output is a $c \times m$ matrix $\mathbf{Z}$. The GCN can be constructed by stacking multiple graph convolutional layers followed by a non-linear operation (such as ReLU). Each node in the instance graph represents the latent feature of a sample and is represented as $[z_{tex}, z_{str}]$, and the adjacency matrix is constructed using the cosine similarity score as $\hat{\mathbf{A}} = \mathbf{X}_{sc} \mathbf{X}_{sc}^T$. Note that the cosine similarity score quantifies the semantic similarity between two samples while the cosine similarity loss defined in Eqn 7 is to be minimized (thus maximizing the cosine similarity). Here $\mathbf{X}_{sc} \in R^{w \times h}$ is the matrix of cosine similarity scores, w is the batch size, and $h = 2$ is the dimension of the similarity scores of each sample obtained from corresponding $[z_{tex}, z_{str}]$.

Training And Implementation: Given $\mathbf{X}$ and $\mathbf{A}$ the GCN features are obtained according to Eq.8. Source and target domain graphs are individually constructed and fed into the parameters-shared GCNs to learn representations. The dimension of z_{tex} is 256, while z_{str} is 64×64. **VAE Network:** The encoder consists of 3 convolution blocks followed by max pooling after each step. The decoder is also symmetrically designed. 3×3 convolution filters are used and $64, 32, 32$ filters are used in each conv layer. The input to the VAE is 256×256. For the domain classifier we use DenseNet-121 network with pre-trained weights from ImageNet and finetuned using self supervised learning. As a pre-text task we use the classifier to predict the intensity values of masked regions.

3. Experiments And Results

3.1. Results For CAMELYON17 Dataset

Dataset Description: We use the CAMELYON17 dataset (Bándi et al., 2019) to evaluate the performance of the proposed method on tumor/normal classification. In this dataset, a total of 500 $H\&E$ stained WSIs are collected from five medical centers (denoted as by $C1_{17}, C2_{17}, C3_{17}, C4_{17}, C5_{17}$ respectively). 50 of these WSIs include lesion-level annotations. All positive and negative WSIs are randomly split into training/validation/test sets and provided by the organizers in a 50/30/20% split for the individual medical centers to obtain the following split: $C1_{17}$:37/22/15, $C2_{17}$: 34/20/14, $C3_{17}$: 43/24/18, $C4_{17}$: 35/20/15, $C5_{17}$: 36/20/15. 256×256 image patches are extracted from the annotated tumors for positive patches and from tissue regions of WSIs without tumors for negative patches. We use $\lambda_1 = 0.9$ and $\lambda_2 = 1.1$.

Method	*Center* 1	*Center* 2	*Center* 3	*Center* 4	*Center* 5	*Average*	p
No UDA	0.8068	0.7203	0.7027	0.8289	0.8203	0.7758	0.0001
MMD	0.8742	0.6926	0.8711	0.8578	0.7854	0.8162	0.0001
CycleGAN	0.9010	0.7173	0.8914	0.8811	0.8102	0.8402	0.002
(Vahadane et al., 2016)	0.9123	0.7347	0.9063	0.8949	0.8223	0.8541	0.003
(Gadermayr et al., 2018)	0.9487	0.8115	0.8727	0.9235	0.9351	0.8983	0.013
(Mahapatra et al., 2020)	0.9668	0.8537	0.9385	0.9548	0.9462	0.9320	0.024
(Chang et al., 2021)	0.979	0.948	0.946	0.965	0.942	0.956	0.017
Proposed	**0.988**	**0.963**	**0.958**	**0.979**	**0.949**	**0.969**	-
Ablation Study Results							
FSL-Same Domain	0.991	0.972	0.965	0.986	0.957	0.974	0.07
w/o $\mathcal{L}_{Str}$	0.929	0.928	0.921	0.932	0.929	0.936	0.004

Table 1: Classification results in terms of AUC measures for different domain adaptation methods on the CAMELYON17 dataset. Note: $FSL-SD$ is a fully-supervised model trained on target domain data.

Implementation Details Since the images have been taken from different medical centers their appearance varies despite sharing the same disease labels. This is due to slightly different protocols of $H\&E$ staining. Stain normalization has been a widely explored topic which aims to standardize the appearance of images across all centers, which is equivalent to domain adaptation. Recent approaches to stain normalization/domain adaptation favour use of GANs and other deep learning methods. We compare our approach to recent approaches and also with (Chang et al., 2021) which explicitly performs unsupervised domain adaptation using MixUp.

Evaluation of our method's performance consist of the following steps: 1) We use $C1_{17}$ as the source dataset and train a ResNet-50 classifier (He et al., 2016) (ResNet_{C1}). Each of the remaining datasets from the other centers are, separately, taken as the target dataset, the corresponding domain adapted images are generated, and classified using ResNet_{C1}. As a baseline we also perform the experiment without domain adaptation denoted as $No-UDA$ where ResNet_{C1} is used to classify images from other centers. We also report results for a network trained in a fully-supervised manner on the training set from the same domain ($FSL-SameDomain$) to give the strongest upper-bound expectation for a UDA model trained on other domain's data. , where a ResNet-50 is trained on the training images and used to classify test images, all from the same hospital. This approach will give the best results for a given classifier (in our case ResNet-50). All the above experiments are repeated using each of $C2_{17}, C3_{17}, C4_{17}, C5_{17}$ as the source dataset. Table 1 summarizes our results.

The results in Table 1 show that UDA methods are definitely better than conventional stain normalization approaches as evidenced by the superior performance of our proposed method and (Chang et al., 2021). Our method performs the best amongst all the methods, and is very close to $FSL-SameDomain$. This clearly shows that our proposed GCN based approach performs better than other UDA methods. The ablation studies also show that our proposed individual loss term $\mathcal{L}_{Str}$ has a significant contribution to the overall performance of our method and excluding it significantly degrades the performance.

3.2. Results on Chest Xray Dataset

There are two publicly available chest Xray datasets - the NIH Xray (Wang et al., 2017) and the CheXpert (Irvin et al., 2017) - which have a large image collection and, most importantly, have the same set of disease labels. A brief description is given as: **NIH Chest Xray** Dataset: For lung disease classification we adopted the NIH ChestXray14 dataset (Wang et al., 2017) having 112,120 expert-annotated frontal-view X-rays from 30,805 unique patients and has 14 disease labels. Original images were resized to 256×256, and $\lambda_1 = 0.9, \lambda_2 = 1.2$. **CheXpert** Dataset: We used the CheXpert dataset (Irvin et al., 2017) consisting of 224,316 chest radiographs of 65,240 patients labeled for the presence of 14 common chest conditions. The validation ground-truth is obtained using majority voting from annotations of 3 board-certified radiologists. Original images were resized to 256×256, and $\lambda_1 = 0.95, \lambda_2 = 1.1$.

We first divide both datasets into train/validation/test splits on the patient level at 70/10/20 ratio, such that images from one patient are in only one of the splits. Then we train a DenseNet-121 (Rajpurkar et al., 2017) classifier on one of the datasets (say NIH's train split). Here the NIH dataset serves as the source data and CheXpert is the target dataset. We then apply the trained model on the test split of the NIH dataset and the results are denoted as $FSL - SameDomain$. When we apply this model to the test split of the CheXpert data without domain adaptation the results are reported under $No - UDA$.

Classification results for different domain adaptation techniques are reported in Table 2 where the NIH dataset was the source domain and the performance metrics are for the CheXpert dataset's *test split*. Table 3 summarizes the performance using the CheXpert dataset as the source dataset, and applied to the NIH Xray dataset's test split. We observe that UDA methods perform worse than $FSL - SameDomain$. This is expected since it is very challenging to perfectly account for domain shift. However all UDA methods perform much better than fully supervised methods trained on one domain and applied on another domain without domain adaptation.

Amongst the different UDA methods we find our method performs the best, including outperforming convectional approaches such as those based on Maximum Mean Discrepancy (MMD) and cycleGANs. The DANN architecture (Ganin et al., 2016) outperforms MMD and cycleGANs, and is on par with graph convolutional methods GCAN (Ma et al., 2019) and GCN2 (Hong et al., 2019b). However our method outperforms all compared methods which can be attributed to the combination of GCNs, which learn more useful global relationships between different samples, and feature disentanglement which in turn leads to more discriminative feature learning.

4. Conclusion

Our graph convolutional network based unsupervised domain adaptation method outperforms conventional CNN methods as graphs better learn the interaction between samples by focusing on more global interactions while CNNs focus on the local neighborhood. This enables GCN to perform better UDA as demonstrated by results on multiple datasets. While feature disentanglement also contributes to improved performance, there is scope for improvement. In future work we wish to explore better disentanglement techniques starting with improving our current approach, and aim to extend this approach for segmentation.

	No UDA	MMD	CycleGANs	DANN	$FSL-SD$	Proposed	w/o $\mathcal{L}_{Str}$	GCAN	GCN2
Atel.	0.697	0.741	0.765	0.792	*0.849*	**0.825**	0.782	0.798	0.809
Card.	0.814	0.851	0.874	0.902	*0.954*	**0.931**	0.897	0.908	0.919
Eff.	0.761	0.801	0.824	0.851	*0.903*	**0.884**	0.859	0.862	0.870
Infil.	0.652	0.699	0.736	0.761	*0.814*	**0.788**	0.764	0.757	0.765
Mass	0.739	0.785	0.817	0.849	*0.907*	**0.890**	0.852	0.858	0.871
Nodule	0.694	0.738	0.758	0.791	*0.825*	**0.818**	0.795	0.803	0.807
Pneu.	0.703	0.748	0.769	0.802	*0.844*	**0.828**	0.798	0.800	0.810
Pneumot.	0.781	0.807	0.832	0.869	*0.928*	**0.903**	0.873	0.867	0.882
Consol.	0.704	0.724	0.742	0.783	*0.835*	**0.818**	0.789	0.776	0.792
Edema	0.792	0.816	0.838	0.862	*0.928*	**0.910**	0.865	0.865	0.883
Emphy.	0.815	0.831	0.865	0.894	*0.951*	**0.934**	0.901	0.908	0.921
Fibr.	0.719	0.745	0.762	0.797	*0.847*	**0.828**	0.799	0.811	0.817
PT	0.728	0.754	0.773	0.804	*0.842*	**0.830**	0.798	0.799	0.812
Hernia	0.811	0.846	0.864	0.892	*0.941*	**0.923**	0.898	0.904	0.914

Table 2: Classification results on the CheXpert dataset's test split using NIH data as the source domain. Note: $FSL-SD$ is a fully-supervised model trained on target domain data.

	No UDA	MMD	CycleGANs	DANN	$FSL-SD$	Proposed	w/o $\mathcal{L}_{Str}$	GCAN	GCN2
Atel.	0.718	0.734	0.751	0.773	*0.814*	**0.798**	0.771	0.78	0.786
Card.	0.823	0.846	0.861	0.882	*0.929*	**0.931**	0.897	0.895	0.906
Eff.	0.744	0.762	0.785	0.819	*0.863*	**0.884**	0.859	0.811	0.833
Infil.	0.730	0.741	0.761	0.785	*0.821*	**0.799**	0.764	0.777	0.789
Mass	0.739	0.785	0.817	0.837	*0.869*	**0.843**	0.832	0.828	0.831
Nodule	0.694	0.738	0.758	0.791	*0.825*	**0.818**	0.795	0.782	0.802
Pneu.	0.683	0.709	0.726	0.759	*0.798*	**0.773**	0.762	0.751	0.763
Pneumot.	0.771	0.793	0.814	0.838	*0.863*	**0.847**	0.822	0.832	0.835
Consol.	0.712	0.731	0.746	0.770	*0.805*	**0.784**	0.761	0.765	0.774
Edema	0.783	0.801	0.818	0.836	*0.872*	**0.849**	0.835	0.828	0.837
Emphy.	0.803	0.821	0.837	0.863	*0.904*	**0.879**	0.863	0.857	0.868
Fibr.	0.711	0.726	0.741	0.766	*0.802*	**0.779**	0.762	0.761	0.768
PT	0.710	0.721	0.737	0.762	*0.798*	**0.774**	0.761	0.756	0.763
Hernia	0.785	0.816	0.836	0.861	*0.892*	**0.868**	0.853	0.851	0.860

Table 3: Classification results on the NIH Xray dataset's test split using CheXpert data as the source domain. Note: $FSL-SD$ is a fully-supervised model trained on target domain data.

References

David Ahmedt-Aristizabal, Mohammad Ali Armin, Simon Denman, Clinton Fookes, and Lars Petersson. Graph-based deep learning for medical diagnosis and analysis: Past, present and future. *Sensors*, 21(14):4758, Jul 2021. ISSN 1424-8220. doi: 10.3390/s21144758. URL http://dx.doi.org/10.3390/s21144758.

Euijoon Ahn, Ashnil Kumar, Michael Fulham, Dagan Feng, and Jinman Kim. Unsupervised domain adaptation to classify medical images using zero-bias convolutional auto-encoders and context-based feature augmentation. *IEEE transactions on medical imaging*, 39(7): 2385–2394, 2020.

Róger Bermúdez-Chacón, Carlos Becker, Mathieu Salzmann, and Pascal Fua. Scalable unsupervised domain adaptation for electron microscopy. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI*, pages 597–609, 2016.

P. Bándi, , and et al. From Detection of Individual Metastases to Classification of Lymph Node Status at the Patient Level: The CAMELYON17 Challenge. *IEEE Trans. Med. Imag.*, 38(2):550–560, 2019.

Jia-Ren Chang, Min-Sheng Wu, Wei-Hsiang Yu, Chi-Chung Chen, Cheng-Kung Yang, Yen-Yu Lin, and Chao-Yuan Yeh. Stain mix-up: Unsupervised domain generalization for histopathology images. In Marleen de Bruijne, Philippe C. Cattin, Stéphane Cotin, Nicolas Padoy, Stefanie Speidel, Yefeng Zheng, and Caroline Essert, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, pages 117–126, Cham, 2021. Springer International Publishing.

Cheng Chen, Qi Dou, Hao Chen, Jing Qin, and Pheng Ann Heng. Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation. *IEEE Transactions on Medical Imaging*, 39(7):2494–2505, 2020. doi: 10.1109/TMI.2020.2972701.

M. Gadermayr, V. Appel, B. M. Klinkhammer, P. Boor, and D. Merhof. Which way round?: A study on the performance of stain-translation for segmenting arbitrarily dyed histological images. In *MICCAI (I)*, pages 165–173, 2018.

Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks, 2016.

Karthik Gopinath, Christian Desrosiers, and Herve Lombaert. Graph domain adaptation for alignment-invariant brain surface segmentation, 2020.

Varun Gulshan, Lily Peng, Marc Coram, Martin C. Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, Jorge Cuadros, Ramasamy Kim, Rajiv Raman, Philip C. Nelson, Jessica L. Mega, and Dale R. Webster. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*, 316(22):2402–2410, 12 2016. ISSN 0098-7484. doi: 10.1001/jama.2016.17216. URL https://doi.org/10.1001/jama.2016.17216.

K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *In Proc. CVPR*, 2016.

Tobias Heimann, Peter Mountney, Matthias John, and Razvan Ionasec. Learning without labeling: Domain adaptation for ultrasound transducer localization. In Kensaku Mori,

Ichiro Sakuma, Yoshinobu Sato, Christian Barillot, and Nassir Navab, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, pages 49–56, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

Yoonmi Hong, Geng Chen, Pew-Thian Yap, and Dinggang Shen. Multifold acceleration of diffusion mri via deep learning reconstruction from slice-undersampled data. In Albert C. S. Chung, James C. Gee, Paul A. Yushkevich, and Siqi Bao, editors, *Information Processing in Medical Imaging*, pages 530–541, Cham, 2019a. Springer International Publishing.

Yoonmi Hong, Jaeil Kim, Geng Chen, Weili Lin, Pew-Thian Yap, and Dinggang Shen. Longitudinal prediction of infant diffusion mri data via graph convolutional adversarial networks. *IEEE Transactions on Medical Imaging*, 38(12):2717–2725, 2019b. doi: 10.1109/TMI.2019.2911203.

Jeremy Irvin, Pranav Rajpurkar, Michael Ko, Yifan Yu, Silviana Ciurea-Ilcus, Chris Chute, Henrik Marklund, Behzad Haghgoo, Robyn Ball, Katie Shpanskaya, Jayne Seekins, David A. Mong, Safwan S. Halabi, Jesse K. Sandberg, Ricky Jones, David B. Larson, Curtis P. Langlotz, Bhavik N. Patel, Matthew P. Lungren, and Andrew Y. Ng. CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In *arXiv preprint arXiv:1901.07031*, 2017.

Konstantinos Kamnitsas, Christian Baumgartner, Christian Ledig, Virginia Newcombe, Joanna Simpson, Andrew Kane, David Menon, Aditya Nori, Antonio Criminisi, Daniel Rueckert, and Ben Glocker. Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. In Marc Niethammer, Martin Styner, Stephen Aylward, Hongtu Zhu, Ipek Oguz, Pew-Thian Yap, and Dinggang Shen, editors, *Information Processing in Medical Imaging*, pages 597–609, Cham, 2017. Springer International Publishing.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks, 2017.

Atsutoshi Kumagai and Tomoharu Iwata. Unsupervised domain adaptation by matching distributions based on the maximum mean discrepancy via unilateral transformations. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):4106–4113, Jul. 2019. doi: 10.1609/aaai.v33i01.33014106. URL https://ojs.aaai.org/index.php/AAAI/article/view/4309.

Feiyan Li, Weisheng Li, Sheng Qin, and Linhong Wang. Mdfa-net: Multiscale dual-path feature aggregation network for cardiac segmentation on multi-sequence cardiac mr. *Knowledge-Based Systems*, 215:106776, 2021. ISSN 0950-7051. doi: https://doi.org/10.1016/j.knosys.2021.106776. URL https://www.sciencedirect.com/science/article/pii/S0950705121000393.

Yun Liu, Krishna Gadepalli, Mohammad Norouzi, George E. Dahl, Timo Kohlberger, Aleksey Boyko, Subhashini Venugopalan, Aleksei Timofeev, Philip Q. Nelson, Greg S. Corrado, Jason D. Hipp, Lily Peng, and Martin C. Stumpe. Detecting cancer metastases on gigapixel pathology images. In *arXiv preprint arXiv:1703.02442*, 2017.

Zelun Luo, Yuliang Zou, Judy Hoffman, and Li Fei-Fei. Label efficient learning of transferable representations across domains and tasks. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 164–176, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

Xinhong Ma, Tianzhu Zhang, and Changsheng Xu. Gcan: Graph convolutional adversarial network for unsupervised domain adaptation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8258–8268, 2019. doi: 10.1109/CVPR.2019.00846.

D. Mahapatra, B. Bozorgtabar, J.-P. Thiran, and L. Shao. Structure preserving stain normalization of histopathology images using self supervised semantic guidance. In *In Proc. MICCAI*, pages 309–319, 2020.

Saeid Motiian, Quinn Jones, Seyed Mehdi Iranmanesh, and Gianfranco Doretto. Few-shot adversarial domain adaptation. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 6673–6683, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

Cheng Ouyang, Konstantinos Kamnitsas, Carlo Biffi, Jinming Duan, and Daniel Rueckert. Data efficient unsupervised domain adaptation for cross-modality image segmentation. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, page 669–677, Berlin, Heidelberg, 2019. Springer-Verlag. ISBN 978-3-030-32244-1. doi: 10.1007/978-3-030-32245-8_74. URL https://doi.org/10.1007/978-3-030-32245-8_74.

Nathan Painchaud, Youssef Skandarani, Thierry Judge, Olivier Bernard, Alain Lalande, and Pierre-Marc Jodoin. Cardiac segmentation with strong anatomical guarantees. *IEEE Transactions on Medical Imaging*, 39(11):3703–3713, 2020. doi: 10.1109/TMI.2020.3003240.

Élodie Puybareau, Zhou Zhao, Younes Khoudli, Edwin Carlinet, Yongchao Xu, Jérôme Lacotte, and Thierry Géraud. Left atrial segmentation in a few seconds using fully convolutional network and transfer learning. In Mihaela Pop, Maxime Sermesant, Jichao Zhao, Shuo Li, Kristin McLeod, Alistair Young, Kawal Rhode, and Tommaso Mansi, editors, *Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges*, pages 339–347, Cham, 2019. Springer International Publishing.

P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P Lungren, and A.Y Ng. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. In *arXiv preprint arXiv:1711.05225,*, 2017.

A. Vahadane, T. Peng, A. Sethi, S. Albarqouni, L. Wang, M. Baust, K. Steiger, A. M. Schlitter, I. Esposito, and N. Navab. Structure-preserving color normalization and sparse stain separation for histological images. *IEEE IEEE Trans. Med. Imag.*, 35(8):1962–1971, 2016.

X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R.M. Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *In Proc. CVPR*, 2017.

Man Wu, Shirui Pan, Chuan Zhou, Xiaojun Chang, and Xingquan Zhu. Unsupervised domain adaptive graph convolutional networks. In *Proceedings of The Web Conference 2020*, pages 1457–1467, 2020.

Shaoan Xie, Zibin Zheng, Liang Chen, and Chuan Chen. Learning semantic representations for unsupervised domain adaptation. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 5423–5432. PMLR, 10–15 Jul 2018.

Appendix A. T-SNE Visualizations

In Figure 2 we show the T-sne plots for our Proposed method, for GCN2 (Hong et al., 2019b), DANN (Ganin et al., 2016) and our method without L_{Str}. The features are obtained from the final layer, and for all methods the output feature vector is 512 dimensional, and the corresponding parameter values are $\lambda_1 = 0.95, \lambda_2 = 1.1$. The plots show visualizations for 5 disease labels for the CheXpert dataset. We show data from the source (marked with dots) and target domain (marked with '+') for these labels. For our proposed method (Figure 2 (a)), the source and target domain data of the same label map to nearby areas, which indicates that the domain adaptation step successfully generates domain invariant features. Additionally, the clusters for different labels is fairly well separated. These two characteristics are not observed for the other methods. GCN2 has some level of separability but there is undesirable overlap across different different classes. The overlap across classes is even worse when we exclude the feature disentanglement components.

The t-sne visualizations clearly indicates that our proposed approach using feature disentanglement is highly effective in learning domain invariant representations for different disease labels. It also highlights the importance of feature disentanglement step. By separating the images into structure and texture features we are able to learn representations that have better discriminative power.

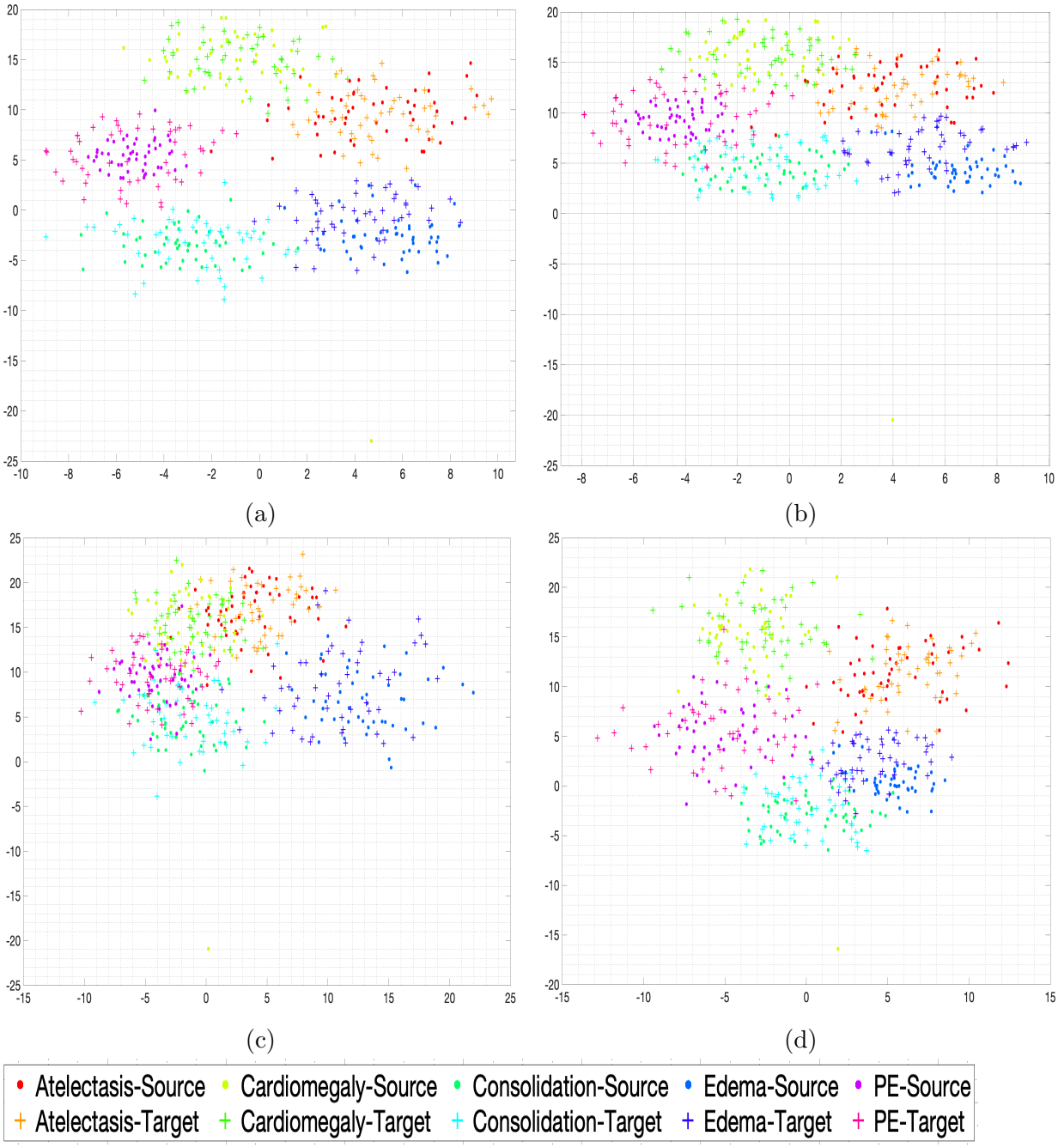


Figure 2: Visualization of tsne plots for different methods using 5 disease labelsfrom the CheXpert Dataset: (a) Our proposed method; (b) Our method without using structure loss; (c) using DANN; (d) the GCN2 method of (Hong et al., 2019b).