Unsupervised Domain Adaptation for Medical Images with an Improved Combination of Losses

Ravi Kant Gupta, Shounak Das and Amit Sethi
Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai, India

Keywords: Adversarial, Deep Learning, Domain Adaptation, Medical Image.

Abstract: This paper presents a novel approach for unsupervised domain adaptation that is tested on H&E stained histology and retinal fundus images. Existing adversarial domain adaptation methods may not effectively align different domains of multimodal distributions associated with classification problems. Since our objective is to enhance domain alignment and reduce domain shifts between these domains by leveraging their unique characteristics, we propose a tailored loss function to address the challenges specific to medical images. This loss combination not only makes the model accurate and robust but also faster in terms of training convergence. We specifically focus on leveraging texture-specific features, such as tissue structure and cell morphology, to enhance adaptation performance in the histology domain. The proposed method – Domain Adaptive Learning (DAL) – was extensively evaluated for accuracy, robustness, and generalization. We conducted experiments on the FHIST and a retina dataset and the results show that DAL significantly surpasses the ViT-based and CNN-based state-of-the-art methods by 1.41% and 6.56% respectively for FHIST dataset while also showing improved results for the retina dataset.

1 INTRODUCTION

In traditional supervised learning, a model is trained using labeled data from the same domain as the test data. Obtaining labels for medical data is challenging due to the intricacies of medical expertise, making it costly and time-consuming. The need for specialized knowledge, meticulous review, and ethical considerations contribute to the difficulty in acquiring accurate and reliable annotations for medical datasets. However, when the distribution of the source and target domains differs significantly, the model's performance may suffer due to the domain shift. This domain shift can be because of color variation, data acquisition bias, distributional differences, domain-specific factors, covariate shift, staining techniques in medical images, etc. Unsupervised domain adaptation (UDA) techniques aim to mitigate this domain shift by aligning the feature distributions or learning domain-invariant representations by using only unlabeled samples from the target domain. By learning domain-invariant representations, adversarial-based UDA models can effectively reduce the domain discrepancy and improve the generalization performance on the target domain. This approach has shown promising results in various domains, such as image classification, object detection, and semantic segmentation. However, while adversarial-based UDA has achieved notable success, challenges still exist. These include addressing the sensitivity to hyper-parameter tuning, handling the high-dimensional feature space, and effectively capturing complex domain shifts.

To address the aforementioned challenge, we develop a UDA approach that surpasses the state-of-the-art performance for medical images. We present our findings from developing convolution neural networks (CNNs) for such tasks based on FHIST dataset (Shakeri et al., 2022), which is composed of several histology datasets, namely CRC-TP (Javed et al., 2020), LC25000 (Borkowski et al.,), BreakHis (Spanhol et al., 2016), and NCT-CRC-HE-100K (Kather et al., 2018). We framed our experiments on CRCTP and NCT with six classes (Benign, Tumor, Muscle, Stroma, Debris, and Inflammatory). The t-distributed stochastic neighbor embedding (tSNE) (van der Maaten and Hinton, 2008) plot in Figure 1 of source data distribution (circle shape) and target data distribution (square shape) while the color of classes differs with light and dark versions of the same color for the FHIST dataset. The sample images of each domain with different classes are shown

Gupta, R., Das, S. and Sethi, A.
Unsupervised Domain Adaptation for Medical Images with an Improved Combination of Losses.
DOI: 10.5220/0012328100003657

In *Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2024) - Volume 1*, pages 205-215
ISBN: 978-989-758-688-0; ISSN: 2184-4305

in Figure 2. To show our model's robustness we also tested it on a retinal dataset from Glaucoma Screening (AIROGS) challenge of ISBI-2021 (De Vente et al., 2023). This retinal dataset is collected from three different centres and consists of two classes (glaucoma positive and glaucoma Negative). Glaucoma is one of the main causes of irreversible blindness and impaired vision in the world. It affects the optic nerve, which connects the eye with the brain, and leads to progressive visual field damage. The sample images of each domain with different classes are shown in Figure 3.

Our research attempts to meet three key objectives: firstly, to reduce the discordance between source and target domains in medical images; secondly, to harness the distinctive attributes of FHIST and Retinal, such as cellular morphology and tissue structure, to elevate adaptation performance specifically within the medical domain; and finally, to transcend the limitations of current UDA techniques, to achieve state-of-the-art accuracy, resilience, and generalization capabilities compared to the previous methods.

Our adoption of deep learning for unsupervised domain adaptation in medical images is driven by its potential to enhance model generalization, extract optimal features, enable versatile cross-domain applications, and achieve field-advancing progress. By tailoring the combination of loss functions which leads to improved convergence and robustness, and with the leverage of deep learning's power, we aim to surpass current methods, benefiting various applications. Inspired by a conditional domain adversarial network (CDAN) (Long et al., 2018a), the core idea is to simultaneously train a feature extractor (typically a deep neural network) and a domain classifier (discriminator) to distinguish between source and target domains. We have examined different CNN-based feature extractor as ResNet-50 (He et al., 2016a), ResNet-101 (He et al., 2016a), ResNet-152 (He et al., 2016a), VIT (Dosovitskiy et al., 2020), and ConvMixer (Trockman and Kolter, 2022) to extract meaningful features. The feature extractor aims to learn domain-invariant representations, while the domain classifier tries to classify the domain of the extracted features correctly. During training, the feature extractor and domain classifier are optimized in an adversarial manner. The feature extractor aims to fool the domain classifier by generating indistinguishable features across domains, while the domain classifier tries to classify the domains correctly. To achieve this, we propose a combination of loss function pseudo label maximum mean discrepancy (PLMMD) along with other losses such as maximum information loss (entropy loss) (Krause et al., 2010), maximum mean discrepancy (MMD) loss (Gretton et al., 2012), minimum class confusion (MCC) loss (Jin et al., 2020), etc. This combination of loss functions has the following specific advantages : Employing MCC loss enhances classification models by minimizing class confusion, particularly in scenarios with imbalanced class distributions. With maximum information loss, our model is encouraged to learn tightly clustered target features with uniform distribution, such that the discriminative information in the target domain is retained, while MDD loss measures the difference between the mean embeddings of two distributions, helping to quantify the dissimilarity between domains and facilitating domain adaptation techniques. The loss PLMMD enhances unsupervised domain adaptation by selectively emphasizing domain-invariant features through weight assignments. The benefit of this loss is, that training convergence is faster as compared to other scenarios. With the help of this novel combination of the loss function our method surpasses not only the CNN-based model state-of-the-art but also the transformer-based model for the medical images. To justify our claims for medical images, we use the FHIST dataset (Shakeri et al., 2022) and Retina dataset.

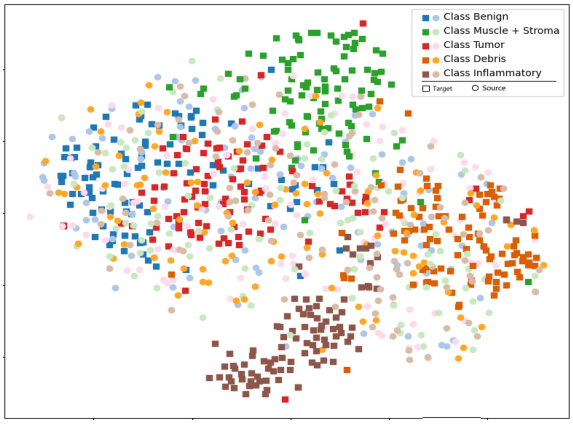

Figure 1: Snapshot of t-SNE plot of source (CRC-TP) (Circle shape) and target (NCT) (Square shape), clearly shows significant difference between source and target data distribution.

Our stated goals were achieved by proposing an improved combination of loss functions tailored to address the unique challenges of H&E stained histology images in FHIST dataset and Retinal dataset. The performance evaluation was focused on accuracy, robustness, and generalization, to surpass state-of-the-art techniques in both domains. Furthermore, the research explored potential cross-domain applications in medical image analysis and computer vision, offering promising advancements in practical unsupervised domain adaptation with the help of various

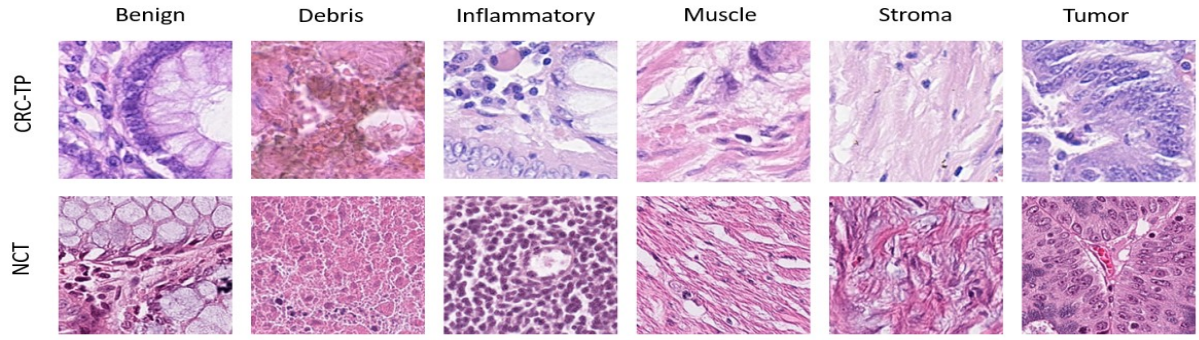

Figure 2: Snapshot of sample images of each class from CRC-TP (top row) and NCT (bottom row) of FHIST dataset.

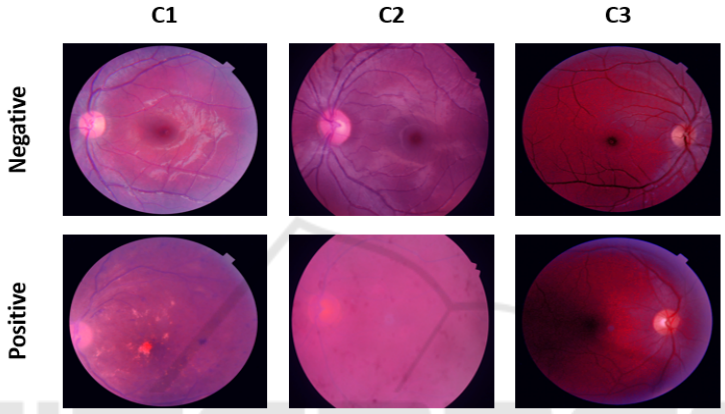

Figure 3: Snapshot of sample images of each class from C1 (row) and NCT (bottom row) of Retinal dataset.

combinations of loss functions with different existing models.

2 BACKGROUND AND RELATED WORK

In unsupervised domain adaptation, we have a source domain $D_s = \{(x_{s_i}, y_{s_i})\}_{i=1}^{n_s}$ of n_s labeled examples and a target domain $D_s = \{(x_{t_i}, y_{t_i})\}_{i=1}^{n_t}$ of n_t unlabeled examples. The source domain and target domain are sampled from joint distributions $P(x_s, y_s)$ and $Q(x_t, y_t)$ respectively. Notably, the two distributions are initially not aligned, that is, $P \neq Q$.

Domain adversarial neural network (DANN) (Ganin et al., 2016) is a framework of choice for UDA. It is a two-player game between domain discriminator D, which is trained to distinguish the source domain from the target domain, and the feature representation F trained to confuse the domain discriminator D as well as classify the source domain samples. The error function of the domain discriminator corresponds well to the discrepancy between the feature distributions $P(f)$ and $Q(f)$ (Ganin and Lempitsky, 2015), a key to bound the target risk in the domain adaptation theory (Ben-David et al., 2010).

Alignment-based domain adaptation is another typical line of work that leverages a domain-adversarial task to align the source and target domains as a whole so that class labels can be transferred from the source domain to the unlabeled target one (Ganin et al., 2016; Pinheiro, 2018; Tzeng et al., 2017; Zhang et al., 2018). Another typical line of work directly minimizes the domain shift measured by various metrics, e.g., maximum mean discrepancy (MMD) (Gretton et al., 2012). These methods are based on domain-level domain alignment. To achieve class-level domain alignment, the works of (Long et al., 2018b; Pei et al., 2018) utilize the multiplicative interaction of feature representations and class predictions so that the domain discriminator can be aware of the classification boundary. Based on the integrated task and domain classifier, (Tang and Jia, 2020) encourages a mutually inhibitory relation between category and domain predictions for any input instance. The works of (Chen et al., 2019a; Xie et al., 2018) align the labeled source centroid and pseudo-labeled target centroid of each shared class in the feature space. Some work uses individual task classifiers for the two domains to de-

tect non-discriminative features and reversely learn a discriminative feature extractor (Lee et al., 2019; Saito et al., 2017a; Saito et al., 2017b). Certain other works focus attention on transferable regions to derive a domain-invariant classification model (Kurmi et al., 2019; Wang et al., 2019; Wen et al., 2019). To help achieve target-discriminative features, (Kang et al., 2018; Sankaranarayanan et al., 2017) generate synthetic images from the raw input data of the two domains via GANs (Goodfellow et al., 2014). The recent work of (Chen et al., 2019b) improves adversarial feature adaptation, where the discriminative structures of target data may be deteriorated (Xu et al., 2019). The work of (Zhao et al., 2019) adapts the feature norms of the two domains to a large range of values so that the learned features are not only task-discriminative but also domain-invariant.

3 PROPOSED METHOD

The challenge of domain shift in a cross-domain classification task using unsupervised domain adaptation leverages the knowledge from a labeled source domain to improve the performance of a classifier on an unlabeled target domain. We propose a tailored loss function that minimizes the domain discrepancy and aligns feature distributions across domains. Our datasets even differ in image sizes for the source domain and the target domain. Before training, the images were subjected to data augmentation such as horizontal flip, vertical flip, and normalization to ensure consistency. To facilitate domain adaptation, we introduce a structure-preserving colour normalization technique to normalize the stain appearance of images across domains (Vahadane et al., 2016). The normalization process aims to preserve the local structure while removing domain-specific variations. Therefore, the images of both dataset were colour normalized (Vahadane et al., 2016).

From the color-normalized images, we extracted features using ResNet-52 trained on ImageNet (He et al., 2016b). Our proposed model architecture is based on a deep neural network with convolutional and fully connected layers, specifically tailored for domain adaptation.

In this work, we design a method to train a deep network $N : x \rightarrow y$ which reduces the shifts in the data distributions across domains, such that the target risk $r_t = E_{(x_t,y_t)\sim Q}[N(x_t) \neq y_t]$ can be bounded by the source risk $r_s = E_{(x_s,y_s)\sim P}[N(x_s) \neq y_s]$ plus the distribution discrepancy disc(P, Q) quantified by a novel conditional domain discriminator. To minimize domain cross-domain discrepancy (Ganin et al., 2016) in

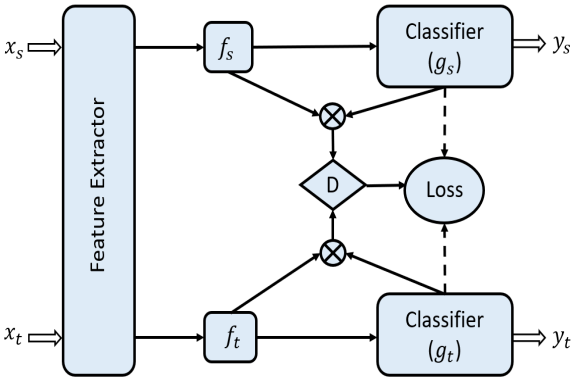

Figure 4: Architecture of the proposed networks, where domain-specific feature representation f and classifier prediction g embody the cross-domain gap to be reduced jointly by the conditional domain discriminator D.

adversarial learning Generative Adversarial Networks (GANs) (Goodfellow et al., 2014) play a vital role. Features are represented by $f = F(x)$ and classifier prediction, $g = N(x)$ generated from deep network N.

We improve existing adversarial domain adaptation methods in two directions. First, when the joint distributions of feature and class, i.e. $P(x_s, y_s)$ and $Q(x_t, y_t)$, are non-identical across domains, adapting only the feature representation f may be insufficient. A quantitative study (Yosinski et al., 2014) shows that deep representations eventually transition from general to specific along deep networks, with transferability decreased remarkably in the domain-specific feature layer f and classifier layer g. Second, due to the nature of multi-class classification, the feature distribution is multimodal, and hence adapting feature distribution may be challenging for adversarial networks.

By conditioning, domain variances in feature representation f and classifier prediction g can be modeled simultaneously. This joint conditioning allows us to bridge the domain gap more effectively, enabling the adapted model to capture and align the underlying data distributions between the source and target domains. Consequently, incorporating classifier prediction as a conditioning factor in domain adaptation holds great potential for achieving improved transferability and generating domain-invariant representations in challenging cross-domain scenarios.

We formulate Conditional Domain Adversarial Network (CDAN) (Long et al., 2018a) as a minimax optimization problem with two competitive error terms: (a) $E(N)$ on the source classifier N, which is minimized to guarantee lower source risk; (b) $E(D,N)$ on the source classifier N and the domain discriminator D across the source and target domains, which is minimized over D but maximized over $f =$

$F(x)$ and $g = N(x)$:

$$L_{clc}(x_{s_i}, y_{s_i}) = \mathbb{E}_{(x_{s_i}, y_{s_i}) \sim D_s} L(N(x_{s_i}), y_{s_i}) \quad (1)$$

$$\begin{aligned} L_{dis}(x_s, x_t) = &- \mathbb{E}_{x_{s_i} \sim D_s} \log[D(f_{s_i}, g_{s_i})] \\ &- \mathbb{E}_{x_{t_j} \sim D_t} \log[1 - D(f_{t_j}, g_{t_j})], \end{aligned} \quad (2)$$

where L is the cross-entropy loss, and $h = (f, g)$ is the joint variable of feature representation f and classifier prediction g. The minimax game of CDAN is

$$\begin{aligned} &\min_N L_{clc}(x_{s_i}, y_{s_i}) - \lambda L_{dis}(x_s, x_t) \\ &\min_D L_{dis}(x_s, x_t), \end{aligned} \quad (3)$$

where λ is a hyper-parameter between the two objectives to trade off source risk and domain adversary.

We condition domain discriminator D on the classifier prediction g through joint variable $h = (f, g)$ to potentially tackle the two aforementioned challenges of adversarial domain adaptation. A simple conditioning of D is $D(f \oplus g)$, where we concatenate the feature representation and classifier prediction in vector $f \oplus g$ and feed it to conditional domain discriminator D. This conditioning strategy is widely adopted by existing conditional GANs (Goodfellow et al., 2014). However, with the concatenation strategy, f and g are independent of each other, thus failing to fully capture multiplicative interactions between feature representation and classifier prediction, which are crucial to domain adaptation. As a result, the multimodal information conveyed in classifier prediction cannot be fully exploited to match the multimodal distributions of complex domains (Song et al., 2009). The multilinear map is defined as the outer product of multiple random vectors. The multilinear map of infinite-dimensional nonlinear feature maps has been successfully applied to embed joint distribution or conditional distribution into reproducing kernel Hilbert spaces (Song et al., 2009; Song and Dai, 2013). Besides the theoretical benefit of the multilinear map $x \otimes y$ over the concatenation $x \oplus y$ (Song et al., 2009; Song et al., 2013). Taking advantage of the multilinear map, in this paper, we condition D on g with the multilinear map. Superior to concatenation, the multilinear map $x \otimes y$ can fully capture the multimodal structures behind complex data distributions. A disadvantage of the multilinear map is dimension explosion.

We enable conditional adversarial domain adaptation over domain-specific feature representation f and classifier prediction g. We jointly minimize with respect to (1) source classifier N and feature extractor F, minimize (2) domain discriminator D, and maximize (2) feature extractor F and source classifier N. This yields the mini-max problem of Domain Adversarial Networks:

$$\begin{aligned} \min_G \quad & \mathbb{E}_{(x_s^i, y_s^i) \sim D_s} L(G(x_s^i), y_s^i) \\ & + \lambda \left(\mathbb{E}_{x_s^i \sim D_s} \log[D(T(h_s^i))] \right. \\ & \left. + \mathbb{E}_{x_t^j \sim D_t} \log[1 - D(T(h_t^j))] \right) \\ \max_D \quad & \mathbb{E}_{x_s^i \sim D_s} \log[D(T(h_s^i))] + \mathbb{E}_{x_t^j \sim D_t} \log[1 - D(T(h_t^j))], \end{aligned} \quad (4)$$

where λ is a hyper-parameter between the source classifier and conditional domain discriminator, and note that $h = (f, g)$ is the joint variable of domain-specific feature representation f and classifier prediction g for adversarial adaptation.

The general problem of adversarial domain adaptation of the proposed model for classification can be formulated as follows:

$$\begin{aligned} L = &\min_N L_{clc}(x_{s_i}, y_{s_i}) - \lambda L_{dis}(x_s, x_t) \\ &+ \beta L_{IM} + \gamma L_{MCC} + \delta L_{MDD} + \eta L_{WMMD} \end{aligned} \quad (5)$$

where λ, β, γ, δ and η are hyper parameters, L_{MCC} is minimum class confusion loss, L_{MDD} is maximum mean discrepancy loss, L_{WMDD} represents weighted maximum mean discrepancy loss and L_{IM} represents information maximization loss. All individual losses have their own specialty and this novel combination of loss significantly surpasses the performance of CNN-based models as well as transformer-based models. A detailed description of all the losses is given below in the losses section.

3.1 Losses

3.1.1 Maximum Mean Discrepancy

Maximum mean discrepancy (MMD) is a kernel-based statistical test used to determine whether given two distributions are the same (Gretton et al., 2012). Given an random variable X , a feature map ϕ maps X to an another space F such that $\phi(X) \in F$. Assuming F satisfies the necessary conditions, we can benefit from the kernel trick to compute the inner product in F:

$$X, Y \text{ such that } k(X, Y) = \langle \phi(X), \phi(Y) \rangle_F, \quad (6)$$

where k is gram matrix produced using the kernel function.

MMD is the distance between feature means. That means for a given probability measure P on X, feature means is an another feature map that takes $\phi(X)$ and maps it to the means of every coordinate of $\phi(X)$:

$$\mu_p(\phi(X)) = [\mathbb{E}[\phi(X_1)], ..., \mathbb{E}[\phi(X_m)]]^T \quad (7)$$

The inner product of feature means of $X \sim P$ and $Y \sim Q$ can be written in terms of kernel function such that:

$$\begin{aligned}\langle \mu_p(\phi(X)), \mu_q(\phi(Y))\rangle_F &= \mathbb{E}_{P,Q}[\langle \phi(X), \phi(Y)\rangle_F] \\ &= \mathbb{E}_{P,Q}[k(X,Y)]\end{aligned} \tag{8}$$

Given X, Y maximum mean discrepancy is the distance between feature means of X, Y :

$$MMD^2(P,Q) = ||\mu_P - \mu_Q||_F^2 \tag{9}$$

$$\begin{aligned}MMD^2(P,Q) &= \langle \mu_P - \mu_Q, \mu_P - \mu_Q\rangle \\ &= \langle \mu_P, \mu_P\rangle - 2\langle \mu_P, \mu_Q\rangle + \langle \mu_Q, \mu_Q\rangle\end{aligned} \tag{10}$$

Using the equation (8), finally above expression becomes

$$\begin{aligned}L_{MMD} &= MMD^2(P,Q) \\ &= \mathbb{E}_P[k(X,X)] - 2\mathbb{E}_{P,Q}[k(X,Y)] + \mathbb{E}_Q[k(Y,Y)]\end{aligned} \tag{11}$$

3.1.2 Pseudo Label Maximum Mean Discrepancy

We calculated the PLMMD using a similar procedure to calculating MMD loss in equation (11). However, our proposed loss differs in terms of weights assigned to each similarity term. Hence we can define PLMMD loss as:

$$\begin{aligned}L_{PLMMD} = w_{XX}\mathbb{E}_P[k(X,X)] - 2w_{XY}\mathbb{E}_{P,Q}[k(X,Y)] \\ + w_{YY}\mathbb{E}_Q[k(Y,Y)],\end{aligned} \tag{12}$$

where, w_{XX} represent weight to get similarity within the source domain, similarly, w_{YY} are weights for similarity within the target domain, and w_{XY} are weights to get similarity within source and target domain. For calculating the weights, first, we generated pseudo labels for the target using a source classifier. After that, the source and target pseudo-label is normalized to account for class imbalances. For each class common to both datasets, dot products of normalized vectors are computed to quantify instance relationships. Calculated dot products are normalized by the count of common classes, ensuring fairness. This returns three weight arrays, representing relationships between instances in the source dataset, target dataset, and source-to-target pairs.

3.1.3 Minimum Class Confusion

The minimum class confusion loss $\mathcal{L}_{MCC}$ (Jin et al., 2020) seeks to minimize confusion terms between classes j and j', such that $j \neq j'$ where the indices are exhaustive over the set of classes. On the target domain, the class confusion term between two classes j and j' is given by:

$$C_{jj'} = \hat{\mathbf{y}}_{\cdot j}^{\mathsf{T}} \hat{\mathbf{y}}_{\cdot j'}^{\mathsf{T}}$$

A much more nuanced and meaningful formulation of the class confusion would be:

$$C_{jj'} = \hat{\mathbf{y}}_{\cdot j}^{\mathsf{T}} \mathbf{W} \hat{\mathbf{y}}_{\cdot j'}^{\mathsf{T}}, \tag{13}$$

where the matrix $\mathbf{W}$ is a diagonal matrix. The diagonal terms W_{ii} are given as the softmax outputs of the entropies in classifying a sample i. $\hat{\mathbf{y}}_{ij}$ is given as:

$$\hat{\mathbf{y}}_{ij} = \frac{\exp(Z_{ij}/T)}{\sum_{j'=1}^{c} \exp(Z_{ij'}/T)}, \tag{14}$$

where c is the number of classes, T is the temperature coefficient, and Z_{ij} is the logistic output of the classifier layer for the class j and the sample i.

After normalizing the class confusion terms, the final MCC Loss function is given as:

$$\mathcal{L}_{MCC} = \frac{1}{c}\sum_{j=1}^{c}\sum_{j'\neq j}^{c} |C_{jj'}|, \tag{15}$$

which is the sum of all the non-diagonal elements of the class confusion matrix. The diagonal terms represent the "certainty" in the classifier, while the non-diagonal terms represent the "uncertainty" in classification. The MCC loss can be added in conjunction with other domain adaptation methods.

3.1.4 Information Maximization Loss

The Information Maximization loss is designed to encourage neural networks to learn more informative representations by maximizing the mutual information between the learned features and the input data (Krause et al., 2010). This type of loss aims to guide the model to capture relevant and distinctive patterns in the data, which can be especially valuable in scenarios where unsupervised learning, domain adaptation, or feature learning are important. The assumptions that $p_t = \text{softmax}(N(f(x_t)))$ are expected to retain as much information about x_t as possible, and decision boundary should not cross high-density regions, but instead lie in low-density regions, which is also known as cluster assumption. These two assumptions can be met by maximizing mutual information between the empirical distribution of the target inputs and the induced target label distribution, which can be formally defined as:

$$\begin{aligned}I(p_t; x_t) &= H(\overline{p}_t) - \frac{1}{n_t}\sum_{j=1}^{n_t} H(p_{tj}) \\ &= -\sum_{k=1}^{K} \overline{p}_{tk}\log(\overline{p}_{tk}) + \frac{1}{n_t}\sum_{j=1}^{n_t}\sum_{k=1}^{K} p_{tkj}\log(p_{tkj}),\end{aligned} \tag{16}$$

where, $p_{tj} = \text{softmax}(G_c(G_f(x_{tj})))$, $\overline{p}_t = \mathbb{E}_{x_t}[p_t]$, and K is the number of classes. Maximizing

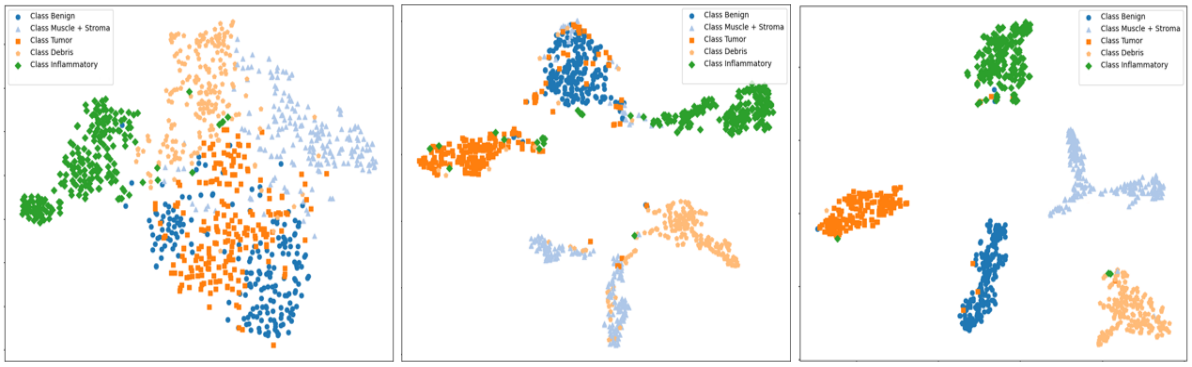

Figure 5: Snapshots of 2D tSNE plots of the target (NCT) domain sample features before training (leftmost), after three epochs (middle), and after six epochs (right).

$-\frac{1}{n_t}\sum_{j=1}^{n_t} H(p_{tj})$ enforces the target predictions close to one-hot encoding, therefore the cluster assumption is guaranteed. To ensure global diversity, we also maximize $H(\overline{p}_t)$ to avoid every target data being assigned to the same class. With $I(p_t;x_t)$, our model is encouraged to learn tightly clustered target features with uniform distribution, such that the discriminative information in the target domain is retained.

4 EXPERIMENTATION AND RESULTS

4.1 Dataset and Implementation

Dataset: To evaluate the proposed method, we introduce the FHIST dataset, a proposed benchmark for the few-shot classification of histological images (Shakeri et al., 2022) and Retinal Dataset from ISBI-2021 Challenge(De Vente et al., 2023). FHIST is composed of several histology datasets, namely CRC-TP (Javed et al., 2020), LC25000 (Borkowski et al.,), BreakHis (Spanhol et al., 2016), and NCT-CRC-HE-100K (Kather et al., 2018). For each class, there are close to 20,000 images in the CRC-TP domain with an image size of 150X150 pixels and around 10,000 images of size 224X224 pixels in the NCT domain. We performed experiments with CRC-TP as the source and NCT as the target and vice versa. The tSNE plots shown in Figure 5 depict the distribution of target (NCT) at different stages of training. Different colors map different class types in the tSNE plot. We have plotted five classes in tSNE which are Benign, Tumor, Debris, Inflammatory, and Muscle + Stroma with 200 sample points from each five classes. We combined the last two classes because of their physiological as well as feature intertwining. The first plot(leftmost) shows the data distribution of NCT(as target) at epoch 0, and the second one shows the data distribution of NCT after three epochs, and the last one (rightmost) shows the target(NCT) data distribution after six epochs of domain adaptation. These histology datasets consist of different tissue types and different organs. We consider each tissue type as a class label with one-hot encoding in the classification task. We framed our experiments on CRC-TP and NCT with six classes (Benign, Tumor, Muscle, Stroma, Debris, and Inflammatory). On the other hand, we demostrated our algorithm on Retinal dataset, in which the images were originally acquired for a diabetic retinopathy screening program. We divided this dataset in three different centres on the basis of center information available in AIROGS challenge of ISBI-2021. This dataset is having 2427 samples in Centre 1 (C1), 497 image samples in Centre 2 (C2) and 276 images in Centre 3 (C3). All centres have images of different sizes.

Implementation: All the experiments were conducted on an NVIDIA A100 in PyTorch, using the CNN-based neural network (ResNet-50) pre-trained on ImageNet (He et al., 2016b) as the backbone for our proposed model. The base learning rate is 0.00001 with a batch size of 32, and we train models by 20 epochs. The hyper-parameters were β=0.05 , γ=1.4 , δ=0.54 and η=0.54 for the experiments of CRC-TP $\rightarrow$ NCT and NCT $\rightarrow$ CRC-TP as well as for the experiments on the Retinal Dataset. We used AdamW (Loshchilov and Hutter, 2019) with a momentum of 0.9, and a weight decay of 0.001 as the optimizer. We adopt the standard protocol for unsupervised domain adaptation (UDA) where all labeled source samples and unlabeled target samples are utilized for training. To report our results for each transfer task, we use center-crop images from the target domain and report the classification performance.

Table 1: Accuracy (%) on the FHIST dataset (Shakeri et al., 2022) with two different UDA tasks and their average, where all methods are fine-tuned on their respective backbone model.

Method	Backbone Feature Extractor	CRC-TP → NCT	NCT → CRC-TP	Average
ResNet-50 (He et al., 2016a)	ResNet-50	40.7	32.9	36.8
DANN (Ganin et al., 2016)		73.5	66.6	70.0
CDAN (Long et al., 2018a)		66.2	61.4	63.8
GVB-GD (Cui et al., 2020)		73.9	66.7	70.3
CHATTY (Wagle et al., 2023)		81.6	67.9	74.7
TVT (Yang et al., 2023)	ViT	86.4	73.2	79.8
Proposed Method	ResNet-50	**87.7**	**74.8**	**81.2**

Table 2: Accuracy (%) on the Retinal dataset (De Vente et al., 2023) with four different UDA tasks.

Experiment	Without Domain Adaptation	Proposed Method
Trained on C1 and tested on C2	62.1	**86.5**
Trained on C1 and tested on C3	75.2	**79.6**
Trained on C1 + C2 and tested on C3	77.8	**80.1**
Trained on C1 + C3 and tested on C2	64.8	**87.1**

For a fair comparison with prior works, we also conduct experiments with the same backbone as ViT-based (Dosovitskiy et al., 2020) as TVT (Yang et al., 2023), ResNet-50 (He et al., 2016a), DANN (Ganin et al., 2016), CDAN (Long et al., 2018a), GVB-GD (Cui et al., 2020), CHATTY+MCC (Wagle et al., 2023) on FHIST dataset.

4.2 Results

Our analysis in Table 1 depicts results with different methods and feature extractors for the FHIST dataset. The top five methods are CNN model using ResNet-50 as a feature extractor trained on ImageNet dataset while TVT uses ViT based model pre-trained on ImageNet-21k dataset. Our proposed method is a CNN-based model that utilizes ResNet-50 as a backbone with a novel combination of loss functions. Our model outperforms CNN-based models such as ResNet-50, DANN, CDAN, GVB-GD, and CHATTY+MCC, and surpasses the state of the Art (SoTA) CNN results by 6.56%. At the same time, our method also surpasses the transformer-based SoTA by 1.41%. We achieved an accuracy of 87.7% for CRC-TP to NCT domain adaptation and 74.8% for NCT to CRC-TP with an average accuracy of 81.26% for both tasks, as mentioned in Table 1 with bold text. Table 2 depicts our model's performance for Retinal dataset by consider permutation combination of centres as source and target respectively. Results in Table 2 shows comparison between with and without domain adaptation. In our experiments, we explored the performance of our model in domain adaptation scenarios, particularly focusing on multiple source domains and a single target domain. When we employed domain adaptation techniques and considered center C1 and C3 as sources, with C2 as the target, we achieved an accuracy of 87.1%. In contrast, without domain adaptation, the accuracy was notably lower at 64.8%. Similarly, when we switched the roles of C1 and C2 as sources and evaluated against C3, we obtained accuracy rates of 80.1% with domain adaptation and 77.8% without it. Furthermore, we conducted experiments to emphasize the significance of domain generalization in the presence of multiple source domains. Specifically, when our model was trained on C1 and tested on C2, it yielded accuracy of 86.5% and 62.1% with and without domain adaptation, respectively. Likewise, training on source C1 and targeting C3 resulted in accuracy rates of 79.6% with domain adaptation and 75.2% without domain adaptation. These findings underscore the importance of domain adaptation and domain generalization techniques in enhancing the robustness and adaptability of our model across diverse source and target domains.

5 DISCUSSION AND CONCLUSION

In this study, we have demonstrated that utilizing different combinations of loss functions with a CNN such as ResNet-50 can lead to significant improvements in unsupervised domain adaptation (UDA) performance that can surpass the performance of ViTs using other UDA methods. By leveraging the strengths of various loss functions tailored to specific domain characteristics, we have surpassed the state-of-the-art (SOTA) performance for histology images. We conducted ablation studies to understand the impact of the different feature extractors such as Con-

vMixer (Trockman and Kolter, 2022) and ResNet-101 (He et al., 2016a). However, the performance in these cases was worse than our reported results. To know the impact of individual loss and a combination of losses, we performed extensive experiments. Through comprehensive experiments, we discovered that Minimum Class Confusion (MCC) loss functions offer an enhancement to classification models by mitigating class confusion, particularly when faced with imbalanced class distributions. In parallel, we observed that information maximization losses aid the classifier in selecting the most certain samples for domain alignment. In our proposed approach, the Pseudo Label Maximum Mean Discrepancy (PLMMD) accelerates training convergence (comparison with CHATTY model) and notably enhances domain alignment by incorporating weighted considerations. Additionally, the Maximum Mean Discrepancy (MMD) loss effectively narrows the gap between the mean embeddings of the two distributions. By artfully combining these distinctive loss functions, we not only surpass the current state-of-the-art but also achieve a comprehensive solution that advances the field of classification models in diverse scenarios.

REFERENCES

Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., and Vaughan, J. (2010). A theory of learning from different domains. *Machine Learning*, 79:151–175.

Borkowski, A. A., Bui, M. M., Thomas, L. B., Wilson, C. P., DeLand, L. A., and Mastorides, S. M. Lc25000 lung and colon histopathological image dataset.

Chen, C., Xie, W., Huang, W., Rong, Y., Ding, X., Huang, Y., Xu, T., and Huang, J. (2019a). Progressive feature alignment for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 627–636.

Chen, X., Wang, S., Long, M., and Wang, J. (2019b). Transferability vs. discriminability: Batch spectral penalization for adversarial domain adaptation. In Chaudhuri, K. and Salakhutdinov, R., editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 1081–1090. PMLR.

Cui, S., Wang, S., Zhuo, J., Su, C., and Tian, Q. (2020). Gradually vanishing bridge for adversarial domain adaptation. pages 12452–12461.

De Vente, C., Vermeer, K., Jaccard, N., Wang, H., Sun, H., Khader, F., Truhn, D., Aimyshev, T., Zhanibekuly, Y., Le, T.-D., Galdran, A., Ballester, M., Carneiro, G., Devika, R., Hrishikesh, P., Puthussery, D., Liu, H., Yang, Z., Kondo, S., Kasai, S., Wang, E., Durvasula, A., Heras, J., Zapata, M., Araujo, T., Aresta, G., Bogunovic, H., Arikan, M., Lee, Y., Cho, H., Choi, Y., Qayyum, A., Razzak, I., Van Ginneken, B., Lemij, H., and Sanchez, C. (2023). Airogs: Artificial intelligence for robust glaucoma screening challenge.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Ganin, Y. and Lempitsky, V. (2015). Unsupervised domain adaptation by backpropagation. In *International conference on machine learning*, pages 1180–1189. PMLR.

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M., and Lempitsky, V. (2016). Domain-adversarial training of neural networks. *Journal of machine learning research*, 17(59):1–35.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., and Weinberger, K., editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc.

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., and Smola, A. (2012). A kernel two-sample test. *Journal of Machine Learning Research*, 13(25):723–773.

He, K., Zhang, X., Ren, S., and Sun, J. (2016a). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

He, K., Zhang, X., Ren, S., and Sun, J. (2016b). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.

Javed, S., Mahmood, A., Werghi, N., Benes, K., and Rajpoot, N. (2020). Multiplex cellular communities in multi-gigapixel colorectal cancer histology images for tissue phenotyping. 29:9204–9219.

Jin, Y., Wang, X., Long, M., and Wang, J. (2020). Minimum class confusion for versatile domain adaptation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXI 16*, pages 464–480. Springer.

Kang, G., Zheng, L., Yan, Y., and Yang, Y. (2018). Deep adversarial attention alignment for unsupervised domain adaptation: The benefit of target expectation maximization. In *Computer Vision – ECCV 2018: 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part XI*, page 420–436, Berlin, Heidelberg. Springer-Verlag.

Kather, J. N., Halama, N., and Marx, A. (2018). 100,000 histological images of human colorectal cancer and healthy tissue.

Krause, A., Perona, P., and Gomes, R. (2010). Discriminative clustering by regularized information maximization. In Lafferty, J., Williams, C., Shawe-Taylor, J., Zemel, R., and Culotta, A., editors, *Advances in Neu-*

ral Information Processing Systems, volume 23. Curran Associates, Inc.

Kurmi, V. K., Kumar, S., and Namboodiri, V. P. (2019). Attending to discriminative certainty for domain adaptation. *CoRR*, abs/1906.03502.

Lee, C., Batra, T., Baig, M. H., and Ulbricht, D. (2019). Sliced wasserstein discrepancy for unsupervised domain adaptation. *CoRR*, abs/1903.04064.

Long, M., Cao, Z., Wang, J., and Jordan, M. I. (2018a). Conditional adversarial domain adaptation. *Advances in neural information processing systems*, 31.

Long, M., Cao, Z., Wang, J., and Jordan, M. I. (2018b). Conditional adversarial domain adaptation. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, page 1647–1657, Red Hook, NY, USA. Curran Associates Inc.

Loshchilov, I. and Hutter, F. (2019). Decoupled weight decay regularization. In *International Conference on Learning Representations*.

Pei, Z., Cao, Z., Long, M., and Wang, J. (2018). Multi-adversarial domain adaptation. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*, AAAI'18/IAAI'18/EAAI'18. AAAI Press.

Pinheiro, P. O. (2018). Unsupervised domain adaptation with similarity learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8004–8013.

Saito, K., Ushiku, Y., Harada, T., and Saenko, K. (2017a). Adversarial dropout regularization. *CoRR*, abs/1711.01575.

Saito, K., Watanabe, K., Ushiku, Y., and Harada, T. (2017b). Maximum classifier discrepancy for unsupervised domain adaptation. *CoRR*, abs/1712.02560.

Sankaranarayanan, S., Balaji, Y., Castillo, C. D., and Chellappa, R. (2017). Generate to adapt: Aligning domains using generative adversarial networks. *CoRR*, abs/1704.01705.

Shakeri, F., Boudiaf, M., Mohammadi, S., Sheth, I., Havaei, M., Ayed, I. B., and Kahou, S. E. (2022). Fhist: a benchmark for few-shot classification of histological images. *arXiv preprint arXiv:2206.00092*.

Song, L. and Dai, B. (2013). Robust low rank kernel embeddings of multivariate distributions. In Burges, C., Bottou, L., Welling, M., Ghahramani, Z., and Weinberger, K., editors, *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc.

Song, L., Fukumizu, K., and Gretton, A. (2013). Kernel embeddings of conditional distributions: A unified kernel framework for nonparametric inference in graphical models. *IEEE Signal Processing Magazine*, 30(4):98–111.

Song, L., Huang, J., Smola, A., and Fukumizu, K. (2009). Hilbert space embeddings of conditional distributions with applications to dynamical systems. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, page 961–968, New York, NY, USA. Association for Computing Machinery.

Spanhol, F. A., Oliveira, L. S., Petitjean, C., and Heutte, L. (2016). A dataset for breast cancer histopathological image classification. *IEEE Transactions on Biomedical Engineering*, 63(7):1455–1462.

Tang, H. and Jia, K. (2020). Discriminative adversarial domain adaptation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(04):5940–5947.

Trockman, A. and Kolter, J. Z. (2022). Patches are all you need? *arXiv preprint arXiv:2201.09792*.

Tzeng, E., Hoffman, J., Saenko, K., and Darrell, T. (2017). Adversarial discriminative domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

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

van der Maaten, L. and Hinton, G. (2008). Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86):2579–2605.

Wagle, M., Gupta, R. K., Sethi, A., et al. (2023). Chatty: Coupled holistic adversarial transport terms with yield for unsupervised domain adaptation. *arXiv preprint arXiv:2304.09623*.

Wang, X., Li, L., Ye, W., Long, M., and Wang, J. (2019). Transferable attention for domain adaptation. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence*, AAAI'19/IAAI'19/EAAI'19. AAAI Press.

Wen, J., Liu, R., Zheng, N., Zheng, Q., Gong, Z., and Yuan, J. (2019). Exploiting local feature patterns for unsupervised domain adaptation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):5401–5408.

Xie, S., Zheng, Z., Chen, L., and Chen, C. (2018). Learning semantic representations for unsupervised domain adaptation. In Dy, J. and Krause, A., editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 5423–5432. PMLR.

Xu, R., Li, G., Yang, J., and Lin, L. (2019). Larger norm more transferable: An adaptive feature norm approach for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1426–1435.

Yang, J., Liu, J., Xu, N., and Huang, J. (2023). Tvt: Transferable vision transformer for unsupervised domain adaptation. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 520–530.

Yosinski, J., Clune, J., Bengio, Y., and Lipson, H. (2014). How transferable are features in deep neural networks? *CoRR*, abs/1411.1792.

Zhang, W., Ouyang, W., Li, W., and Xu, D. (2018). Collaborative and adversarial network for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Zhao, H., Combes, R. T. D., Zhang, K., and Gordon, G. (2019). On learning invariant representations for domain adaptation. In Chaudhuri, K. and Salakhutdinov, R., editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 7523–7532. PMLR.