This ACCV 2022 workshop paper, provided here by the Computer Vision Foundation, is the author-created version. The content of this paper is identical to the content of the officially published ACCV 2022 LNCS version of the paper as available on SpringerLink: https://link.springer.com/conference/accv

Handling Domain Shift for Lesion Detection via Semi-Supervised Domain Adaptation

Manu Sheoran, Monika Sharma, Meghal Dani and Lovekesh Vig
Email: {monika.sharma1@tcs.com}

TCS Research, New Delhi, India

Abstract. As the community progresses towards automated Universal Lesion Detection (ULD), it is vital that the techniques developed are robust and easily adaptable across a variety of datasets coming from different scanners, hospitals, and acquisition protocols. In practice, this remains a challenge due to the complexities of the different types of domain shifts. In this paper, we address the domain-shift by proposing a novel domain adaptation framework for ULD. The proposed model allows for the transfer of lesion knowledge from a large labeled source domain to detect lesions on a new target domain with minimal labeled samples. The proposed method first aligns the feature distribution of the two domains by training a detector on the source domain using a supervised loss, and a discriminator on both source and unlabeled target domains using an adversarial loss. Subsequently, a few labeled samples from the target domain along with labeled source samples are used to adapt the detector using an over-fitting aware and periodic gradient update based joint few-shot fine-tuning technique. Further, we utilize a self-supervision scheme to obtain pseudo-labels having high-confidence on the unlabeled target domain which are used to further train the detector in a semi-supervised manner and improve the detection sensitivity. We evaluate our proposed approach on domain adaptation for lesion detection from CT-scans wherein a ULD network trained on the DeepLesion dataset is adapted to 3 target domain datasets such as LiTS, KiTS and 3Dircadb. By utilizing adversarial, few-shot and incremental semi-supervised training, our method achieves comparable detection sensitivity to the previous methods for few-shot and semi-supervised methods as well as to the Oracle model trained on the labeled target domain.

1 Introduction

Universal Lesion Detection (ULD) aims to assist radiologists by automatically detecting lesions in CT-scans across different organs [1–4]. Although, existing ULD networks perform well over a trained source domain, they are still far from practically deployable for clinical applications due to their limited generalization capabilities across target datasets acquired using different scanners and protocols. This domain shift often degrades the detection performance of ULD by over 30-40% when tested on an unseen but related target domain.

A naive approach to circumvent domain-shift is to fine-tune a ULD network, trained on source domain, over sufficient labeled target domain samples. However, obtaining

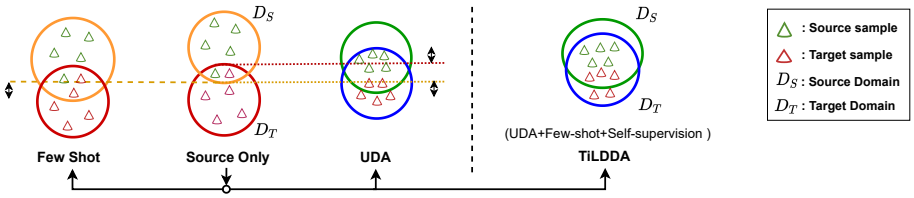

Fig. 1. Visualization of knowledge space of the detector for different adaptation methods across source and target domain. UDA stands for unsupervised domain adaptation. Here, we utilize feature alignment property of unsupervised domain adaptation (UDA) along with few-shot labeled samples from target domain to widen the knowledge space of the detector network for precise lesion detection.

requisite amount of annotations in every new domain is impractical due to the expensive and time-consuming annotation process. Simple fine-tuning may improve sensitivity on the target domain but it suffers from performance drop on the source domain which is not desirable in practical scenarios. For example, when a new CT-machine is added to a facility, then it is expected from a ULD network to maintain its detection sensitivity on new datasets along with the source domain. Therefore, domain adaptation [5–8] is the most effective and widely used technique to easily transfer knowledge from source to new unseen target domains. Widely, there are two approaches to reduce the domain-gap between source and target domain, either by image-to-image translation or by aligning the feature-space. In image-to-image translation techniques [9, 10], researchers have utilized networks such as StyleGAN [11], CycleGAN [10, 12, 13] etc. to generate source images in the style of target images and train a network on the target translated source-images. On the other hand, in feature-space alignment techniques [14–18], authors align the feature-space between source and target domain using either unsupervised adversarial training or prototype alignment. The underlying idea is to generate non-discriminatory features such that the discriminator cannot differentiate between the domains and the task-network trained on a labeled source domain can give similar performance on the new target domain. While large scale annotation of medical scans is expensive, it is often feasible to obtain a few labeled target samples for real world applications. This small amount of annotated data can often provide significant gains for domain-adaptation [19–22].

To learn from few examples of rare objects, two-stage fine-tuning approach (TFA) [23] is proposed where the detector network is first trained with abundant base images and subsequently, only the last layer of trained detectors are finetuned by jointly training few samples from base classes and few samples available for rare/novel classes. However, TFA can help to improve performance on rare classes only if the data for rare classes belong to the same distribution as that of base/source classes.Similarly, in another paper [24], a two-stage semi-supervised object detection method is proposed where detector model is first trained with labeled source data followed by training on unlabelled target data. It utilises an approach called Unbiased Teacher (UBT) where it jointly trains a student and a gradually progressing teacher model by using pseudo labeling technique. The teacher model is provided with weakly augmented inputs and the

generated pseudo-labels are used for the supervision of student model provided with strongly augmented inputs. UBT is utilized to reduce the false positives in the generated pseudo labels, as these false positives can hinder the training process. We avoid the complex student-teacher training by improving the quality of generated pseudo labels by learning better initialization weights via UDA and joint few-shot finetuning.

In this paper, we propose a semi-supervised domain adaptation [25–27] approach which utilizes a combination of unsupervised feature alignment at image as well as instance level similar to Every Pixel Matters (EPM) [28], and few-shot labels from the target domain to further expand the representation-space of the ULD network for adaptation to the target domain, as visualized in Figure 1. Subsequently, we utilize self-supervised learning where we apply the few-shot adapted ULD network on the unlabeled target dataset and obtain pseudo labels using a high confidence threshold. These pseudo labels are used to re-train the ULD network in a semi-supervised manner on the unlabeled target domain. As the combined data for joint few-shot training is dominated by the source domain [29], we train the network via a robust over-fitting aware and periodic gradient update based training scheme which iteratively performs gradient updates on source and target domain samples while accounting for the imbalance in the source and target domain data. The proposed approach can be applied to different convolution based detection backbones and the performance of feature-space alignment based unsupervised domain adaptation techniques can be enhanced and made comparable to that of the Oracle detection network by incorporating few-shot training over target domain labels and semi-supervision using generated pseudo labels on an unlabeled target dataset. To the best of our knowledge, there is very limited research on domain adaptation for lesion detection [30] and we perform transfer of knowledge from a ULD model trained on a large multi-organ dataset to organ-specific target datasets with minimal labeled samples. To summarize, our contributions in this paper are as follows:

- We propose a novel semi-supervised domain adaptation network for ULD via adversarial training, which utilizes few-shot learning for better understanding of the target domain and pseudo-labels based self-supervised learning for more accurate lesion detection on target domain. The network is named *TiLDDA*: **T**owards Universal **L**esion **D**etection via **D**omain **A**daptation.
- A simple anchor-free training scheme is used for lesion detection network which has less design parameters and can handle lesions of multiple sizes from different domains more effectively.
- We evaluate TiLDDA, trained over a source DeepLesion [31] CT dataset, and on 3 target datasets namely, KiTS [32], LiTS [33] and 3Dircadb [34]. The results show consistent improvement in detection sensitivity across all target datasets.
- Owing to the non-availability of lesion detection datasets, we generate bounding box (bbox) annotations of lesions from ground-truth pixel-level segmentation masks on above 3 target domain datasets and release bbox annotations for benchmarking and motivating further research in this area.

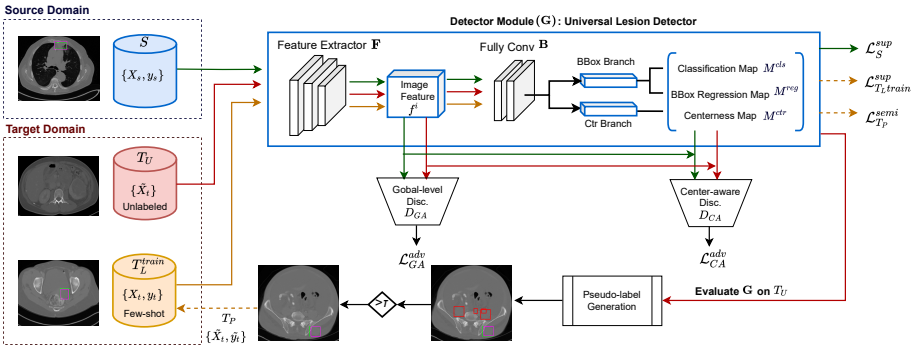

Fig. 2. Overview of our proposed TiLDDA architecture. Source dataset S and unlabeled target dataset T_U are used to train discriminators D_{GA} and D_{CA} using adversarial losses $\mathcal{L}_{GA}^{adv}$ and $\mathcal{L}_{CA}^{adv}$ for domain adaptation. Labeled source domain samples S and few labeled target domain samples T_L^{train} are used to train ULD detector in a few-shot way using supervised losses $\mathcal{L}_S^{sup}$ and $\mathcal{L}_{T_L^{train}}^{sup}$. Further, the few-shot domain adapted ULD is used to generate pseudo-labels on T_U having confidence above a threshold τ. Finally, the pseudo labels are used to re-train the detector in a semi-supervised manner using loss $\mathcal{L}_{T_P}^{semi}$.

2 Methodology

Given a labeled dataset $S = \{(X_s, y_s)\}$ from a source domain D_S, and a dataset T from a different but related target domain D_T split into: an unlabeled set $T_U = \{\tilde{X}_t\}$ and a much smaller labeled set $T_L^{train} = \{(X_t, y_t)\}$, where $T = T_U + T_L^{train}$. Both S and T share the same task, i.e., given an input CT-image X, find the bounding box (Bbox) of the lesion present y. Therefore, the aim of our proposed domain adaptation network is to learn a single set of detector model parameters G_θ such that the model trained on the source domain D_S and few labeled target domain samples T_L^{train} can work efficiently on an unseen target test-set T^{test} without degradation in lesion detection performance. The different components of our proposed domain adaptation pipeline (shown in Figure 2) are as follows:

2.1 Universal Lesion Detection

To cater to the need of detecting multi-sized lesions across different domains, we utilize a robust anchor-free lesion detector (G) based on a fully convolutional one-stage (FCOS) [35, 3] network which performs detection in a per-pixel prediction fashion rather than utilizing the pre-defined anchor-boxes. As shown in Figure 2, for an input image X, we first extract the feature maps (f^i) at i^{th} feature pyramid network (FPN) level using a convolutional feature-extractor F. Next, using a fully-connected detection head B, each pixel location (x, y) of f^i is classified with probability $(p_{x,y})$ as foreground (with class label ${c^*}_{x,y} = 1$) or background (with class label ${c^*}_{x,y} = 0$) and then, for each positive pixel location, a $4D$ vector $u_{x,y}$ is regressed against the corresponding ground-truth bbox annotation ${u^*}_{x,y}$. To further decrease the low-quality bbox

detections, a single centerness layer (Ctr) branch is added in parallel with the regression (Bbox) branch. It is used to give more preference to pixel locations that are present near the center and filter out the pixels that have a skewed feature location inside the ground-truth bbox (y) of the corresponding object. The centerness represents the normalized distance between a particular pixel location and the center of the ground-truth bbox of the corresponding object. The detection loss function, as used in FCOS [35], for ULD baseline is defined as follows:

$$\mathcal{L}^{det}(p_{x,y}, u_{x,y}) = \frac{1}{N_{pos}} \sum_{x,y} \mathcal{L}^{cls}(p_{x,y}, {c^*}_{x,y}) + \frac{\lambda}{N_{pos}} \sum_{x,y} \mathbb{1}_{{c^*}_{x,y}>0} \mathcal{L}^{reg}(u_{x,y}, {u^*}_{x,y}) \tag{1}$$

$$centerness = \sqrt{\frac{min(l^*, r^*)}{max(l^*, r^*)} \times \frac{min(t^*, b^*)}{max(t^*, b^*)}} \tag{2}$$

Here, $\mathcal{L}^{cls}$ and $\mathcal{L}^{reg}$ are the classification focal loss and regression IoU loss for location (x, y), N_{pos} is the no. of positive samples, λ is the balance weight, $\mathbb{1}_{{c^*}_{x,y}>0}$ is an indicator function for every positive location, c^* and u^* are ground-truth labels for classification and regression, respectively. For given regression targets l^*, t^*, r^* & b^* of a location, the term centerness (as defined in Eq. 2) is trained with binary cross entropy (BCE) loss $\mathcal{L}^{ctr}$ and added to the loss function defined in Eq. 1 for the refined results. Finally, the ULD network is trained using a supervised loss $\mathcal{L}^{sup}$ function as defined in Eq. 3.

$$\mathcal{L}^{sup}(X, y) = \mathcal{L}^{det} + \mathcal{L}^{ctr} \tag{3}$$

2.2 Feature Alignment via Adversarial Learning

Here, as inspired by EPM network [28], we utilize unsupervised domain adaptation (UDA) to align the feature distribution for both the domains, source D_S and target D_T, which would result in an increase in the detection sensitivity on the target domain test dataset T^{test} in an unsupervised manner. First, the detector network G is trained on S using a supervised loss-function $\mathcal{L}_S^{sup}$, as defined in Eq.3. To train the discriminators, we first extract source $({f^i}_s)$ and target $({f^i}_t)$ feature maps by applying feature extractor F on S and T_U samples, and perform global feature alignment via a global discriminator D_{GA} which is optimized by minimizing a binary cross-entropy loss $\mathcal{L}_{GA}^{adv}$. This is a domain-prediction loss that aims to identify whether the pixels on i^{th} feature map (f^i) belong to the source / target domain. For a location (x, y) on f^i, $\mathcal{L}_{GA}^{adv}$ can be defined as below:

$$\mathcal{L}_{GA}^{adv}(X_s, \tilde{X}_t) = -\sum_{x,y} z \log\left(D_{GA}(f_s^i)^{(x,y)}\right) + (1-z) \log\left(1 - D_{GA}(f_t^i)^{(x,y)}\right) \tag{4}$$

We set the domain label z of source and target as 1 and 0, respectively. Next, the detection head B predicts pixel-wise objectness maps M^{obj} and centerness maps M^{cls}

Algorithm 1: Proposed Joint Few-shot Learning

Data: Source dataset S and few-shot labeled target dataset T_L^{train}, detector model G_θ, and Hyper-parameters: α, β, and κ.
$n(S)$, $n(T_L^{train}) \leftarrow$ Total source and labeled target samples
for iterations $= 1, 2, 3, ...$ **do**
  Train-source: Gradients $\nabla_\theta = G_\theta^{'}(S; \theta)$;
  Updated parameters: $\theta^{'} \leftarrow \theta - \alpha\nabla_\theta$;
  $\eta = \frac{n(S)}{n(T_L^{train})*\kappa}$;
  if (iterations $\mathbf{mod}$ $\eta) = 0$ **do**
    Train-target: Gradients $\nabla_{\theta'} = G_{\theta'}^{'}(T_L^{train}; \theta^{'})$;
    Updated parameters: $\theta \leftarrow \theta^{'} - \beta\nabla_{\theta'}$;
  else
    $\theta \leftarrow \theta^{'}$;

which are combined to generate a centre-aware map M^{CA} [28]. The extracted feature maps f^i along with M^{CA} are utilized to train another center-aware discriminator D_{CA} with domain-prediction loss $\mathcal{L}_{CA}^{adv}$, as given in Eq. 5 in order to perform center-aware alignment at the pixel level.

$$\begin{aligned}\mathcal{L}_{CA}^{adv}(X_s, \tilde{X}_t) = -\sum_{x,y} z \log\left(D_{CA}(M_s^{CA} \odot f_s^i)^{(x,y)}\right) \\ +(1-z)\log\left(1 - D_{CA}(M_t^{CA} \odot f_t^i)^{(x,y)}\right)\end{aligned} \tag{5}$$

We apply the gradient reversal layer (GRL) [36] before each discriminator for adversarial learning, which reverses the sign of the gradient while optimizing the detector. The loss for the discriminators is minimized via Eq. 4 and Eq.5, while the detector is optimized by maximizing these loss functions, in order to deceive the discriminator. Hence, the overall loss function for UDA using δ and γ as balancing weights, can be expressed as follows:

$$\mathcal{L}^{UDA}(S, T_U) = \mathcal{L}_S^{sup}(X_s, y_s) + \delta\mathcal{L}_{GA}^{adv}(X_s, \tilde{X}_t) + \gamma\mathcal{L}_{CA}^{adv}(X_s, \tilde{X}_t) \tag{6}$$

2.3 Proposed Joint Few-shot Learning (FSL)

Different from standard few-shot learning, where the tasks for target domain are different from the source domain and hence, we either train the network on available source samples first and then use the trained model weights as initialization or, in case of no source domain, we can use weights from already trained models such as ImageNet weights to fine-tune the task network on few target samples separately. Here in this paper, we are trying to solve the domain shift problem where we have the same task, i.e. lesion detection from CT-scans, for both the source (D_S) and target (D_T) domains.

Therefore, we can jointly fine-tune the ULD baseline G on a small labeled target domain dataset T_L^{train} combined with the larger labeled source dataset S. However, this setting suffers from data imbalance as the combined data is dominated by the source domain and hence, the training will be biased towards the source domain. To mitigate this issue, we propose a modified version of the few-shot training paradigm as given by Algorithm 1, which aims to regularize the ULD network and enable it to focus more on target domain samples without over-fitting on one particular domain. The idea is to train the detector G on both domains by alternatively updating their weights so as to ensure balanced updation across source and target samples. This is achieved by finding the best possible gradient direction due to the shared parameter optimization of the two losses. The loss on source train set S is computed using model parameter θ. The loss on the target train set T_L^{train} is computed using shared updated parameter $\theta^{'} = \theta - \alpha\nabla_\theta$ after each η iterations. To avoid over-fitting on target domain, we compute η such that κ epochs of target are trained when 1 epoch of source is trained. We empirically determined the optimal value of $\kappa = 3$. The supervised loss function for the proposed FSL is defined in Eq. 7, where $\mathbb{1}_\eta$ is an indicator function that takes a value of 1 after each η iteration.

$$\mathcal{L}^{few}(S, T_L^{train}) = \mathcal{L}_S^{sup}(X_s, y_s, \theta) + \mathbb{1}_\eta \mathcal{L}_{T_L^{train}}^{sup}(X_t, y_t, \theta^{'}) \tag{7}$$

2.4 Few-shot Domain Adaptation (FDA)

Next, we apply the adversarial learning ($\mathcal{L}_{GA}^{adv}$ and $\mathcal{L}_{CA}^{adv}$) over source and target domain for feature alignment with the proposed FSL ($\mathcal{L}^{few}$) on the combined domain. This helps in increasing the similarity between the two domains via feature-alignment and also widens the knowledge space of ULD by incorporating information from the target domain in the form of few-shot labeled samples. The loss function for FDA is defined as follows:

$$\mathcal{L}^{FDA}(S, T_U, T_L^{train}) = \mathcal{L}^{few}(X_s, X_t, y_s, y_t) + \delta\mathcal{L}_{GA}^{adv}(X_s, \tilde{X}_t) + \gamma\mathcal{L}_{CA}^{adv}(X_s, \tilde{X}_t) \tag{8}$$

2.5 Self-supervision

As unlabeled samples T_U of target domain are available in abundance, we utilize a self-supervised learning mechanism to further improve the ULD performance on T by expanding the few-shot labeled sample space for T. Here, we obtain bbox predictions ($\tilde{y_t}$), having confidence-score above a detection threshold (τ), on unlabeled target samples $\tilde{X}_t$ by applying the few-shot adapted UDA network. Hence, we generate pseudo samples ($T_P = \{\tilde{X}_t, \tilde{y_t}\}$) to further fine-tune the FDA network in a semi-supervised manner using ($\mathcal{L}_{T_P}^{semi}$) (defined in Eq. 9) on target domain.

$$\mathcal{L}_{T_P}^{semi} = \mathcal{L}^{sup}(\tilde{X}_t, \tilde{y_t}) \tag{9}$$

	Data split	No. of Patients	No. of Images	No. of Lesions
KiTS (230 Patients)	T_U^{train}	180	3914	4083
	T_L^{train}	10	919	1305
	T^{test}	40	923	949
	Data split	**No. of Patients**	**No. of Images**	**No. of Lesions**
LiTS (130 Patients)	T_U^{train}	80	4270	11932
	T_L^{train}	10	847	2342
	T^{test}	40	2073	4571
	Data split	**No. of Patients**	**No. of Images**	**No. of Lesions**
3Dircadb (15 Patients)	T_U^{train}	4	144	430
	T_L^{train}	3	113	163
	T^{test}	8	311	676

Table 1. Data distribution of Target Domain Datasets T. Here, T_U^{train}, T_L^{train} and T^{test} represent the unlabeled train data, labeled few-shot train data and test-data.

3 Experiments and Results

3.1 Overall Training Scheme of TiLDDA

We train the ULD network G on source samples (S) and use the source domain weights for initializing our proposed TiLDAA network. For domain adaptation on T, we first train the detector G and discriminators D_{GA} & D_{CA} via the FDA training method using loss defined in Eq. 8. Subsequently, we apply the adapted detector G on unlabeled target images $\tilde{X}_t$ and generate pseudo-labels ($T_P = \{\tilde{X}_t, \tilde{y}_t\}$). Next, we re-train the ULD network using the semi-supervised loss defined in Eq. 9. Hence, the final objective loss-function of our proposed TiLDDA network using hyper-parameters $\delta, \gamma, \eta,$ & λ is as follows:

$$\mathcal{L}^{TiLDDA}(S, T_U, T_L^{train}, T_P) = \mathcal{L}_S^{sup}(S, \theta) + \delta\mathcal{L}_{GA}^{adv}(X_s, \tilde{X}_t) + \gamma\mathcal{L}_{CA}^{adv}(X_s, \tilde{X}_t) + \mathbb{1}_\eta(\mathcal{L}_{T^{train}}^{sup}(T_L^{train}, \theta^{'}) + \lambda\mathcal{L}_{T_P}^{semi}(T_P, \theta^{'})) \tag{10}$$

3.2 Data and Evaluation metric

We evaluate our TiLDDA network on lesion-detection from CT-scans by adapting the ULD model trained on DeepLesion [31] as source domain dataset S to 3 different target domain datasets T such as KiTS [32], LiTS [33] and 3Dircadb [34]. We provide details for different Source and Target domain datasets as follows:

– **Source Domain Database** S: DeepLesion* is the largest publicly available multi-organ lesion detection dataset, released by National Institutes of Health (NIH) Clinical Center. It consists of approximately $32,000$ annotated lesions from $10,594$

*DeepLesion: https://nihcc.app.box.com/v/DeepLesion

CT-scans of $4,427$ unique patients having 1-3 lesions bounding boxes annotated for each CT scan by radiologists.

– **Target Domain Database** T: Since, we were unable to find relevant CT datasets for lesion detection, we utilized the ground-truth segmentation masks for lesions provided in the following target datasets to generate the bounding box-annotations. To introduce domain shift properly, we have selected target datasets that are collected across different geographical locations. The details of these datasets are given as below:
  1. *KiTS*[†]: This cohort includes 230 CT-scans of patients who underwent partial or radical nephrectomy for suspected renal malignancy between 2010 and 2018 at University of Minnesota Medical Center, US. The kidney-region and the kidney-tumors in this dataset are annotated by experts and segmentation masks are released publicly.
  2. *LiTS*[‡]: This dataset consists of 130 pre- and post-therapy CT-scans released by Technical University of Munich, Germany. The image data is also very diverse with respect to resolution and image quality. The manual segmentations of tumors present in liver region are provided in the dataset.
  3. *3Dircadb*[§]: 3D Image Reconstruction for Comparison of Algorithm Database (3Dircadb) released by Research Institute against Digestive Cancer, Strasbourg Cedex, France. It consists of 15 CT-scans of patients with manual segmentation of liver tumors performed by clinical experts.

Please refer Table 1 for data-split used for training and testing. For all the experiments, we have used labeled data of 10 patients from LiTS and KiTS dataset. But due to the very small size of 3Dircadb, we utilize labeled data of 3 patients only. The idea behind using very small-sized 3Dircadb dataset is to evaluate how effectively the proposed TiLDDA network can adapt with minimal target domain samples. As part of pre-processing the CT-images, we include black-border clipping, re-sampling voxel space to $0.8 \times 0.8 \times 2$ mm^3 and HU-windowing with a range of $[-1024, 3072]$. We also perform data augmentations such as horizontal and vertical flipping, resizing and pixel translations along x- and y-axis. For evaluation, average of detection sensitivities over four false positive rates (FPs $= \{0.5, 1, 2, 4\}$) is computed and for all future references in the paper, detection sensitivity means average detection sensitivity.

3.3 Experimental Setup

The feature extractor F is composed of ResNet-101 backbone along with 5 FPN levels and the fully-convolutional block B consists of 3 branches for classification, regression and centerness computations. For robust DA, feature alignment is done across all FPN levels and the architectures of detector G and discriminators D_{GA} & D_{CA} are similar to that used in EPM [28]. We implement TiLDDA in PyTorch-1.4 and train it on a

[†]KiTS: https://kits19.grand-challenge.org/data

[‡]LiTS: https://competitions.codalab.org/competitions/17094

[§]3Dircadb: https://www.ircad.fr/research/3d-ircadb-01

Training scheme	KiTS [32]	LiTS [33]	3Dircadb [34]
Source Only (DeepLesion) (FCOS) [35]	34.2	36.7	37.3
Vanilla Few-shot	44.8	40.8	20.7
UBT (few-shot) [24]	36.4	40.2	18.7
UBT (few-shot + semi-supervision) [24]	44.1	47.3	27.2
TFA (joint few-shot) [23]	54.1	51.2	42.2
Proposed joint few-shot (FSL)	56.6	53.1	45.6
UDA (EPM) [28]	39.4	44.6	42.1
Fewshot DA (FDA)	58.6	53.8	47.1
Fewshot DA + Self-supervision (TiLDDA)	*71.6*	*55.2*	*49.5*
Oracle (Target only)	**77**	**57.6**	**61.1**

Table 2. Average sensitivity (%) on target datasets using different training schemes.

NVIDIA $V100$ 16GB GPU using a batch-size of 4. For all our experiments, we set the values of κ, δ, γ, λ, and τ to 3, 0.01, 0.1, 0.5, and 0.7, respectively. The weights used in GRL for adversarial training are set to 0.01 and 0.02 for D_{GA}& D_{CA}, respectively. The detector network G for FDA is initialized using weights learned via pre-training on source S. An SGD optimizer is used to train FDA network for $65,000$ iterations with a learning rate of e^{-3} and decay-factor of 10 after $32,000$ and 52000 iterations. For overall training of TiLDAA, FDA model is further fine-tuned on S and updated T samples using a learning rate of e^{-4} for $25,000$ iterations.

3.4 Result and Ablation Study

The lesion detection sensitivity on S using the ULD baseline with ResNet-101 backbone is 80% and the aim of our proposed TiLDDA network is to perform well on target domain dataset T as well while maintaining the performance on source domain. Table 2 presents the performance of different training schemes on test-split T^{test} of target domain. The upper bound of detection sensitivity on T^{test} is determined by supervised training of the ULD baseline G directly on target samples (T_U^{train}) only in a supervised manner and this setting is referred to as Oracle setting. The lower bound is computed by evaluating the target T test-set T^{test} directly using the ULD model trained on source (S) only.

It is evident from Table 2 that there is a drop of about 30% to 40% in the detection sensitivity compared to that of Oracle setting due to the domain-shift issue. To circumvent this issue, we begin by training the network using different few-shot techniques without any domain adaptation. First, we use the vanilla few-shot finetuning, where the ULD network is initialized with ImageNet weights and trained only using few labeled target samples T_L^{train}. As expected, the ULD network performs better as compared to Source only training scheme on the test set T^{test}, except for 3Dircadb target domain where training samples are low. Next, we use a semi-supervised method UBT proposed by Yen et. al [24] which utilizes pseudo labels along with few-shot finetuning, there is a further improvement (7% to 8%) in detection sensitivity but it's still limited as lesion detection knowledge from source domain is not being utilized till now by these methods. Hence, we also experimented with a two-step joint few-shot

Training Scheme	**UaDAN** [37]	**EPM** [28]
Source Only (DeepLesion)	35.6	36.7
UDA	40.1	44.6
FDA	50.7	53.8
TiLDDA	*52.8*	*55.2*
Oracle	**56.5**	**57.6**

Table 3. Experiment to show that our proposed training scheme can be used with feature-space alignment based unsupervised domain adaptation methods to further enhance their detection performance. Average sensitivity (%) for LiTS test dataset using different training schemes applied on UaDAN [37] method and EPM [28] method.

finetuning approach (TFA) [23], where we used the entire source domain data in second step of fine-tuning resulting in a steep increase in the detection sensitivity value, especially for the 3Dircadb target domain. This confirms that training target samples with source samples can help to improve the detection sensitivity for the target dataset. However, as mentioned in Section 2.3 simple joint few-shot training suffers from data-imbalance issue, hence we train the network with our proposed joint few-shot training scheme (FSL) described in Algorithm 1 and demonstrate an increase in sensitivity (2% to 3%). Further, utilizing source domain data alone for handling domain shift issue is not enough, hence we apply a UDA method which utilizes adversarial training to align cross-domain features, similar to EPM [28], and observe that even without using any data from target domain, there is a small but significant improvement (5% to 7%) in sensitivity as compared to Source only training scheme. Subsequent to this, we combine the UDA and proposed joint few-shot method for few-shot adaptation (FDA) to train the ULD network and obtain an enhanced performance. At last, we generate psuedo labels (T_P) using the FDA model and further, fine-tune it via TiLDDA model. It can be seen clearly that we obtain a remarkable improvement (12% to 35%) in lesion detection as compared to source only training using our proposed TiLDDA network with very few labeled target samples.

Next, in order to support our claim that our proposed training scheme can be used with different convolution-based detection backbones and the performance of feature-space alignment based unsupervised DA methods can be improved, we utilize a UDA method proposed in [37] and apply the proposed incremental training schema (joint few-shot + pseudo label based self-supervision) on LiTS as target dataset. We observe a similar trend for improvement of the lesion detection sensitivity as obtained with EPM baseline used for TiLDAA, as shown in Table 3.

Further, we present the ablation-study in Table 4 on the number of few-shot labeled samples (T_L^{train}) of different target domain datasets and hyper-parameter κ used in Algorithm 1. We observe that 10 is the optimal number of few-shot labeled samples to obtain best performance. But due to the very small size of 3Dircadb dataset, we utilize labeled data of 3 patients only from 3Dircadb. As the combined data in few-shot learning is dominated by source samples, so we train the network on target samples for more number of epochs as compared to source domain using different values of κ and

Target domain	No. of patients	$n(T_L^{train})$	κ	Sensitivity (%)
LiTS	1	81	1	46.3
	5	428	1	50.3
	10	**847**	1	51.4
			3	**53.8**
			5	53.3
KiTS	**10**	**919**	1	56.1
			3	58.6
			5	57.1
3Dircadb	**3**	**113**	1	45.4
			3	**47.1**
			5	46.2

Table 4. Average sensitivity (%) for different number of labeled target samples $(n(T_L^{train}))$ and hyper-parameter κ for different target domains using FDA training scheme.

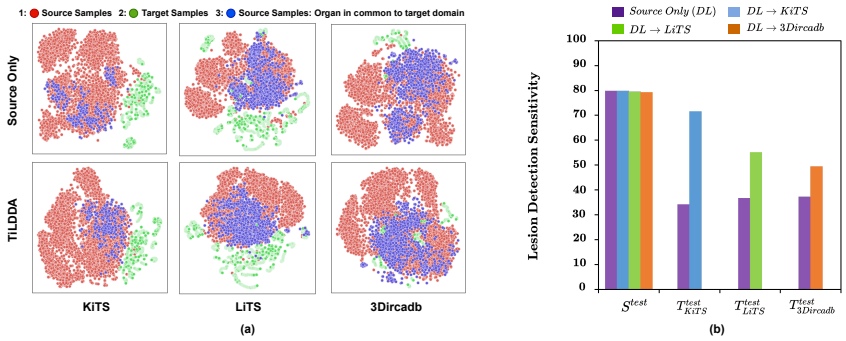

Fig. 3. (a) The t-SNE visualization of source and target sample distributions before and after using TiLDDA.(b) Effect of domain-adaptation on lesion-detection sensitivity of test-set of S and T domain datasets. Here, DL refers to DeepLesion dataset.

found that a value of 3 is optimal that avoids the model from over-fitting over target domain.

Additionally, we present a qualitative comparison using t-SNE [38] plots in Figure 3(a) to visualize the distribution from test-split of source D_S and target D_T domain samples using Source-Only and TiLDDA training schemes. We extract embeddings of the test-samples using feature extractor F. The labels 1, 2, and 3 correspond to samples from source, target and samples of source domain organ in common to the target domain, respectively. The common organ of DeepLesion and, LiTS and 3Dircadb is liver. However, the common organ of DeepLesion and KiTS is kidney. We can clearly observe that after adaptation, the embeddings of target domain organs are now aligned better with the common organ of source domain resulting in an enhanced detection sensitivity for the target domain. It validates our claim that the detection knowledge from source domain can be transferred to the target domain to improve the lesion detection

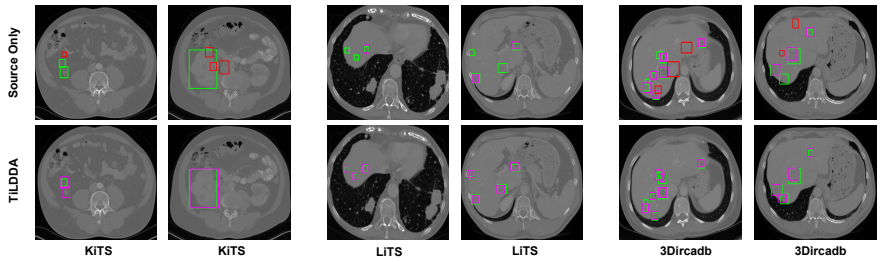

Fig. 4. Qualitative comparison of Lesion Detection before and after using TiLDDA. Here green, magenta, and red color boxes represent ground-truth, true-positive (TP), and false-positive (FP) lesion detection, respectively.

performance. Further, we present the comparison of detection-sensitivity on test-set of S and T datasets before and after applying TiLDAA in Figure 3(b). It can be seen that the performance on source domain is maintained during domain adaptation and our proposed method TiLDAA gives better detection sensitivity on target domain as compared to the Source only trained model. Also, it is clearly evident in Figure 4 that TiLDDA is able to reduce false positives and detect lesions which were missed using source-only trained lesion detector.

4 Conclusion and Future Work

In this paper, we present a simple but effective self-supervision based few-shot domain adaptation technique for ULD which can be used to enhance performance of existing detection methods. We utilize multi-organ lesion detection knowledge from a larger universal lesion detection source domain dataset to efficiently detect lesions on three organ-specific target domains, and achieve comparable performance to the Oracle training scheme by utilizing only a few labeled target samples. We first adversarially align the representation space of the two domains via unsupervised domain adaptation and with a few labeled target samples, further fine-tune the detector in a semi-supervised way using the self-generated pseudo labels. We experimentally show the efficacy of our method by reducing the performance drop on unseen target domains compared to an Oracle model trained on a fully labeled target dataset. In the current setup, both source and target domains have a common task of detecting lesions from CT images across a common set of organs. Going forward, we would like to propose a network that can adapt to out-of-distribution organs and work across cross-modality domains.

References

1. Yan, K., et al.: MULAN: multitask universal lesion analysis network for joint lesion detection, tagging, and segmentation. In: MICCAI, Springer (2019) 194–202
2. Yan, K., et al.: Universal Lesion Detection by learning from multiple heterogeneously labeled datasets. arXiv preprint arXiv:2005.13753 (2020)

3. Sheoran, M., Dani, M., Sharma, M., Vig, L.: An efficient anchor-free universal lesion detection in ct-scans. In: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), IEEE (2022) 1–4
4. Sheoran, M., Dani, M., Sharma, M., Vig, L.: Dkma-uld: Domain knowledge augmented multi-head attention based robust universal lesion detection. arXiv preprint arXiv:2203.06886 (2022)
5. Gopalan, R., Li, R., Chellappa, R.: Domain adaptation for object recognition: An unsupervised approach. In: 2011 international conference on computer vision, IEEE (2011) 999–1006
6. Long, M., Zhu, H., Wang, J., Jordan, M.I.: Unsupervised domain adaptation with residual transfer networks. Advances in neural information processing systems **29** (2016)
7. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering **22** (2010) 1345–1359
8. Saito, K., Ushiku, Y., Harada, T., Saenko, K.: Strong-weak distribution alignment for adaptive object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2019) 6956–6965
9. Saxena, S., Teli, M.N.: Comparison and analysis of image-to-image generative adversarial networks: A survey. CoRR **abs/2112.12625** (2021)
10. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. (2017) 2223–2232
11. Rojtberg, P., Pollabauer, T., Kuijper, A.: Style-transfer gans for bridging the domain gap in synthetic pose estimator training. 2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR) (2020) 188–195
12. Yang, J., Dvornek, N.C., Zhang, F., Chapiro, J., Lin, M., Duncan, J.S.: Unsupervised domain adaptation via disentangled representations: Application to cross-modality liver segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer (2019) 255–263
13. Dou, Q., Ouyang, C., Chen, C., Chen, H., Glocker, B., Zhuang, X., Heng, P.A.: Pnp-adanet: Plug-and-play adversarial domain adaptation network at unpaired cross-modality cardiac segmentation. IEEE Access **7** (2019) 99065–99076
14. Li, H., Pan, S.J., Wang, S., Kot, A.C.: Domain generalization with adversarial feature learning. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2018) 5400–5409
15. Lee, S.M., Kim, D., Kim, N., Jeong, S.G.: Drop to adapt: Learning discriminative features for unsupervised domain adaptation. 2019 IEEE/CVF International Conference on Computer Vision (ICCV) (2019) 91–100
16. Tanwisuth, K., Fan, X., Zheng, H., Zhang, S., Zhang, H., Chen, B., Zhou, M.: A prototype-oriented framework for unsupervised domain adaptation. CoRR **abs/2110.12024** (2021)
17. Kamnitsas, K., Baumgartner, C., Ledig, C., Newcombe, V., Simpson, J., Kane, A., Menon, D., Nori, A., Criminisi, A., Rueckert, D., et al.: Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. In: International conference on information processing in medical imaging, Springer (2017) 597–609
18. Shin, S.Y., Lee, S., Summers, R.M.: Unsupervised domain adaptation for small bowel segmentation using disentangled representation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer (2021) 282–292
19. Zhao, A., Ding, M., Lu, Z., Xiang, T., Niu, Y., Guan, J., Wen, J.R.: Domain-adaptive few-shot learning. In: 2021 IEEE Winter Conference on Applications of Computer Vision (WACV). (2021) 1389–1398
20. Teshima, T., Sato, I., Sugiyama, M.: Few-shot domain adaptation by causal mechanism transfer. CoRR **abs/2002.03497** (2020)

21. Wang, T., Zhang, X., Yuan, L., Feng, J.: Few-shot adaptive faster r-cnn. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019) 7166–7175
22. Li, S., Sui, X., Fu, J., Fu, H., Luo, X., Feng, Y., Xu, X., Liu, Y., Ting, D.S., Goh, R.S.M.: Few-shot domain adaptation with polymorphic transformers. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer (2021) 330–340
23. Wang, X., Huang, T.E., Darrell, T., Gonzalez, J.E., Yu, F.: Frustratingly simple few-shot object detection. arXiv preprint arXiv:2003.06957 (2020)
24. Liu, Y.C., Ma, C.Y., He, Z., Kuo, C.W., Chen, K., Zhang, P., Wu, B., Kira, Z., Vajda, P.: Unbiased teacher for semi-supervised object detection. arXiv preprint arXiv:2102.09480 (2021)
25. Pan, F., Shin, I., Rameau, F., Lee, S., Kweon, I.S.: Unsupervised intra-domain adaptation for semantic segmentation through self-supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2020) 3764–3773
26. Li, J., Li, G., Shi, Y., Yu, Y.: Cross-domain adaptive clustering for semi-supervised domain adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2021) 2505–2514
27. RoyChowdhury, A., Chakrabarty, P., Singh, A., Jin, S., Jiang, H., Cao, L., Learned-Miller, E.: Automatic adaptation of object detectors to new domains using self-training. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2019) 780–790
28. Hsu, C.C., Tsai, Y.H., Lin, Y.Y., Yang, M.H.: Every pixel matters: Center-aware feature alignment for domain adaptive object detector. ArXiv **abs/2008.08574** (2020)
29. Li, Z., Hoiem, D.: Learning without forgetting. IEEE transactions on pattern analysis and machine intelligence **40** (2017) 2935–2947
30. Wang, J., He, Y., Fang, W., Chen, Y., Li, W., Shi, G.: Unsupervised domain adaptation model for lesion detection in retinal oct images. Physics in Medicine & Biology **66** (2021) 215006
31. Yan, K., et al.: DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. J. Med. Imaging (2018)
32. Heller, N., Sathianathen, N., Kalapara, A., Walczak, E., Moore, K., Kaluzniak, H., Rosenberg, J., Blake, P., Rengel, Z., Oestreich, M., et al.: The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes. arXiv preprint arXiv:1904.00445 (2019)
33. Bilic, P., et al.: The liver tumor segmentation benchmark (lits). arXiv preprint arXiv:1901.04056 (2019)
34. Huang, Q., Sun, J., Ding, H., Wang, X., Wang, G.: Robust liver vessel extraction using 3d u-net with variant dice loss function. Computers in biology and medicine **101** (2018) 153–162
35. Tian, Z., et al.: FCOS: Fully convolutional one-stage object detection. In: ICCV. (2019) 9627–9636
36. Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation. In: International conference on machine learning, PMLR (2015) 1180–1189
37. Guan, D., Huang, J., Xiao, A., Lu, S., Cao, Y.: Uncertainty-aware unsupervised domain adaptation in object detection. IEEE Transactions on Multimedia (2021)
38. Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research **9** (2008)