MetaCorrection: Domain-aware Meta Loss Correction for Unsupervised Domain Adaptation in Semantic Segmentation

Xiaoqing Guo[1*]   Chen Yang[1*]   Baopu Li[2]   Yixuan Yuan[1†]

[1]City University of Hong Kong   [2]Baidu USA

{xqguo.ee, cyang.ee}@my.cityu.edu.hk   baopuli@baidu.com   yxyuan.ee@cityu.edu.hk

Abstract

Unsupervised domain adaptation (UDA) aims to transfer the knowledge from the labeled source domain to the unlabeled target domain. Existing self-training based UDA approaches assign pseudo labels for target data and treat them as ground truth labels to fully leverage unlabeled target data for model adaptation. However, the generated pseudo labels from the model optimized on the source domain inevitably contain noise due to the domain gap. To tackle this issue, we advance a MetaCorrection framework, where a Domain-aware Meta-learning strategy is devised to benefit Loss Correction (DMLC) for UDA semantic segmentation. In particular, we model the noise distribution of pseudo labels in target domain by introducing a noise transition matrix (NTM) and construct meta data set with domain-invariant source data to guide the estimation of NTM. Through the risk minimization on the meta data set, the optimized NTM thus can correct the noisy issues in pseudo labels and enhance the generalization ability of the model on the target data. Considering the capacity gap between shallow and deep features, we further employ the proposed DMLC strategy to provide matched and compatible supervision signals for different level features, thereby ensuring deep adaptation. Extensive experimental results highlight the effectiveness of our method[a] against existing state-of-the-art methods on three benchmarks.

1. Introduction

Unsupervised domain adaptation (UDA) aims to adapt a model for the unlabeled target domain through transferring the knowledge from a labeled source domain with the same label space. UDA for semantic segmentation is a crucial practical problem since it may be beneficial for various real-world applications, such as simulation for robots [18] and autonomous driving [44]. The main challenge of UDA
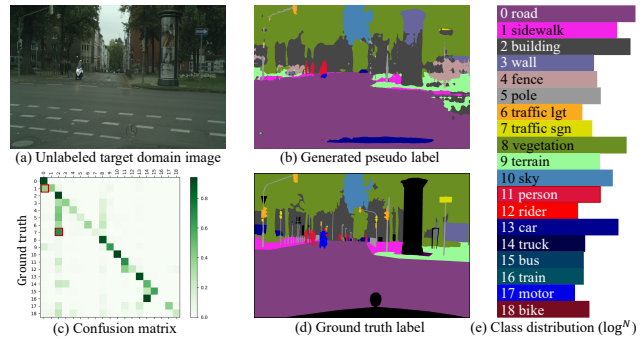

Figure 1. Sample of the noisy pseudo labels on Cityscapes [10]. The generated pseudo labels suffer from the data distribution biases in comparison to the ground truth.

semantic segmentation lies in the divergence of data distribution between two domains [5, 48]. Such domain gap often results in significant performance degradation if the model learned on the labeled source data is directly applied to the target samples [46, 51].

There exist two major lines of approaches to tackle the domain gap problem. On one hand, adversarial learning based UDA methods as a dominant stream have been devised to bridge the domain gap by aligning the distributions of two domains in the appearance [6, 7, 9, 22], feature [2, 11, 39, 45] or output spaces [20, 26, 40, 41]. Despite the significant progress of domain alignment, these works ignored the domain-specific knowledge and could not guarantee the sufficient discriminative capability of the classifier for the specific task. On the other hand, self-training based methods [4, 5, 17, 23, 24, 28, 30, 35, 42, 46, 48, 49, 50, 51] have emerged as promising alternatives towards UDA, which enhanced the discrimination property of target features and implicitly encouraged cross-domain alignment by simultaneously training with pseudo-labeled target data and labeled source data. Specifically, self-training methods assign pixel-wise pseudo labels according to confidence score [17, 23, 30, 51] or uncertainty [24, 48], providing extra supervision for target data to optimize the model.

However, one issue with self-training based UDA methods is that the generated pseudo labels usually suffer from

the noise problem, as illustrated in Figure 1 (b). The presence of noisy pseudo labels may severely hamper the generalization ability of the adapted models, because deep neural networks (DNN) may overfit due to these noisy labeled data. Although some existing works [14, 23, 51] manually define a threshold to eliminate the low-confidence pseudo-labeled samples, it is still challenging in several aspects. First, the threshold value is hard to predefine manually. It may depend on many factors such as the stage of training procedure, the degree of discrepancy between two domains, the number of pixels in each class, the location of the pixel, etc. Secondly, those selected training samples may be misclassified with high confidence, leading to accumulated errors. In fact, the noisy pseudo labels tend to appear in underrepresented minor classes or ambiguous classes. For instance, the minor category 'traffic sign' is overwhelmed by major category 'building', and 'road' is usually connected to 'sidewalk', yielding noisy labels, as in Figure 1 (c). In this scenario, noisy pseudo labels can be theoretically converted from the ground truth labels via a NTM [37, 47], which encodes the inter-class misclassification relationship.

To heuristically discover intrinsic inter-class noise transition probabilities underlying target data, we model the noise distribution of pseudo labels by a NTM and devise a domain-aware meta-learning strategy to estimate the NTM in a learning-to-learn fashion. The key idea of domain-aware meta-learning is to obtain the meta-knowledge of underlying label distribution of clean data in the target domain, and we introduce a domain predictor to adaptively select domain-invariant source data with ground truth labels as meta data set, so as to guide the derivation of NTM. The domain-aware meta-learning strategy enables the gradient of empirical risk measured on meta data to update the NTM, thereby boosting the generalization capacity. Then the approximated noise distribution can be utilized to explicitly correct the supervision signal for target data, aiming at solving the noisy pseudo label problem in a self-training based UDA method. An alternating optimization approach is further adopted to mutually improve the estimation of NTM and the UDA segmentation model. For simplicity, we refer to the whole Domain-aware Meta-learning strategy for Loss Correction in the above process as DMLC. Moreover, we devise a MetaCorrection framework, which incorporates DMLC to provide matched supervision signals for outputs of different levels, thereby enhancing the deep adaptation of model. In particular, we introduce the learnable NTMs for different layers, and adopt the proposed domain-aware meta-learning to estimate the corresponding noise distributions and benefit the loss correction.

We summarize our contributions in four aspects.

- We present a MetaCorrection framework, which incorporates the proposed DMLC strategy for UDA semantic segmentation. To our best knowledge, it represents the first effort to formally model the noise distribution of pseudo labels in target domain by a learnable NTM and further solve it in a meta-learning strategy.

- In the DMLC strategy, we formulate the misclassification probability of inter-classes to model noise distribution in target domain and devise a domain-aware meta-learning algorithm to estimate NTM for loss correction in a data driven manner.

- Our MetaCorrection framework aims to provide matched and compatible supervision signals for different layers with the proposed DMLC strategy, boosting the adaptation performance of model.

- We conduct extensive comparison experiments and ablation studies to thoroughly verify the impact and effect of the proposed MetaCorrection framework.

2. Related Work
2.1. UDA in Semantic Segmentation

Unsupervised domain adaptation (UDA) aims to bridge the distribution gap between the labeled source domain and unlabeled target domain, thus improving the generalization capability of the learned models on the target data. The general idea of UDA semantic segmentation methods is to perform domain alignment through adversarial learning [2, 6, 7, 9, 11, 20, 22, 26, 40, 41] or utilize self-training strategy on target samples to adapt the segmentation models [4, 17, 23, 28, 30, 35, 46, 48, 49, 50, 51]. We briefly review some typical works in the following parts.

Adversarial Learning based UDA semantic segmentation models usually contain two networks [40]. One network behaves as a generator to obtain the segmentation maps for source and target inputs, while the other network serves as a discriminator to derive domain predictions. The generator intends to fool the discriminator to ensure the cross-domain alignment of feature [2, 11] or output [20, 26, 40, 41] levels. Other methods tried to narrow down the domain gap at the input level via image style transformation [6, 7, 9, 22]. However, these domain alignment methods induced by adversarial learning ignored the domain-specific information and could not guarantee the discriminative ability for semantic segmentation [42, 48].

Self-training. Another line of work for UDA semantic segmentation leverages the idea of self-training to adapt the segmentation model and learn the domain specific information [28, 46, 49]. Previous methods [4, 50] introduced entropy minimization to fully leverage the unlabeled target data for model training and encouraged the model to predict with high confidence score. Recently, increasing researchers investigated the problem of pseudo label noise in target domain by filtering out noisy samples with respect to confidence score or uncertainty [17, 23, 30, 35, 48, 51]. Zou et al. [51] proposed to threshold the argmax values of predictions and selected high-confidence pseudo-labeled samples. Zheng et al. [48] utilized uncertainty estimation and
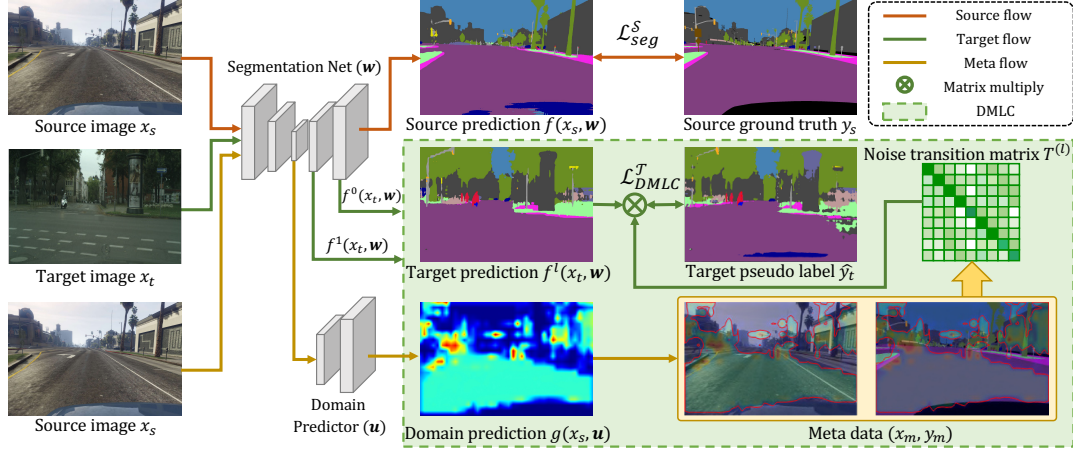
Figure 2. The proposed MetaCorrection framework contains a segmentation net and a domain predictor. Both source and target images are passed through the segmentation net to perform semantic segmentation. The source data is supervised by the loss between prediction and the corresponding ground truth label, while the supervision signals of noisy pseudo-labeled target data are corrected by the learnable NTMs. Domain predictor is introduced to select domain-invariant source pixels for the guidance of NTM estimation.

enabled the dynamic threshold to obtain rectified pseudo labels. However, these methods only involved confident samples for training, which may result in biased prediction in minor classes and cannot distinguish confused categories. Moreover, when the noise ratio is high (at the early stage of training procedure), these models filtered out a large number of target samples, which may lose useful information in omitted samples. In this paper, we model the noise distribution of pseudo labels in target domain with a learnable matrix that encodes inter-class noise transition relationship, and propose a DMLC strategy to adaptively distill knowledge from all samples.

2.2. Deep Learning with Noisy Labels

Many efforts have been devoted to tackling the noisy label problem in DNN training and can be roughly categorized into three typical strategies: label correction [1, 47], sample reweighting [19, 21, 36], and loss correction [13, 16, 31, 37, 43]. Zhang *et al.* [47] introduced the meta-learning algorithm to conducted a dynamic linear combination of noisy label and prediction from DNN, thereby refurbishing noisy labels. Li *et al.* [37] progressively incorporated increasing samples in an easy-to-hard manner to enable a robust model trained with noisy samples. Authors in [19, 37] utilized multiple layer perception network to automatically assign large weighting factor for easy samples. Goldberger *et al.* [13] estimated the noise pattern through embedding a noise adaptation layer in DNN model. Wang *et al.* [43] leveraged a small set of trusted clean-labeled samples to estimate the NTM for loss correction. Nonetheless, these strategies were designed for fully-supervised whole-image classification, and could not be directly incorporated into UDA semantic segmentation. Our work represents the first effort to exploit the loss correction with an effective meta-learning strategy for self-training based UDA.

3. Methodology

3.1. Preliminaries

We focus on the problem of UDA in semantic segmentation. In the source domain, we have access to source images $X_{\mathcal{S}} = \{x_s \in \mathbb{R}^{H \times W \times 3}\}_{s \in \mathcal{S}}$ and the corresponding pixel-wise one-hot labels $Y_{\mathcal{S}} = \{y_s \in \{0, 1\}^{H \times W \times C}\}_{s \in \mathcal{S}}$, while only target images $X_{\mathcal{T}} = \{x_t \in \mathbb{R}^{H \times W \times 3}\}_{t \in \mathcal{T}}$ are available in the target domain. Note that H, W, C denote the height, width of images and the number of classes, respectively. The goal is to learn a segmentation net $f(\cdot)_{\mathbf{w}}$ that can correctly categorize pixels for target data $X_{\mathcal{T}}$. Self-training based methods [17, 23, 30, 35, 46, 49, 50] regarded pseudo labels of target data as learnable hidden variables, $\widehat{Y}_{\mathcal{T}} = \{\widehat{y}_t\}_{t \in \mathcal{T}} = \{\arg\max f(x_t)_{\mathbf{w}}\}_{t \in \mathcal{T}}$, and utilized them as approximate ground truth labels for model training. Then cross-entropy loss over the source and target dataset for self-training can be defined as

$$
\begin{aligned}
\mathcal{L}_{ST} &= \mathcal{L}_{seg}^{\mathcal{S}}(X_{\mathcal{S}}, Y_{\mathcal{S}}) + \mathcal{L}_{seg}^{\mathcal{T}}(X_{\mathcal{T}}, \widehat{Y}_{\mathcal{T}}) \\
&= -\sum_{s \in \mathcal{S}} y_s \log f(x_s, \mathbf{w}) - \sum_{t \in \mathcal{T}} \widehat{y}_t \log f(x_t, \mathbf{w}).
\end{aligned}
\tag{1}
$$

By minimizing the empirical risk of target data with respect to the estimated pseudo label $\widehat{y}_t$, the optimized model thus can be discriminatively adapted to the target domain.

3.2. Self-training with Loss Correction

Jointly optimizing the model and estimating pseudo labels for target data is difficult as the accuracy of generated pseudo labels cannot be guaranteed. The noise in pseudo labels could deteriorate the performance of existing self-training based methods, and result in unstable training and biased predictions. A feasible way is to enhance the noise tolerance property of target domain risk minimization via

loss correction. To incorporate loss correction, we assume that the generated pseudo labels $\widehat{Y}_t$ can be bridged to the ground truth labels Y_t via an underlying noise transition matrix (NTM) $T \in [0,1]^{C \times C}$, which specifies the probability of ground truth label j flipping to noisy label k by $T_{jk} = p(\widehat{y}_t = k \mid y_t = j)$.

If we directly optimize the segmentation net $f(\cdot)_{\mathbf{w}}$ on the noisy pseudo-labeled taget data, we would obtain a class posterior probability for noisy label $p(\widehat{y}_t = k \mid x_t)$. NTM bridges the posterior for noisy label $p(\widehat{y}_t = k \mid x_t)$ and the class probability for ground truth label via:

$$p(\widehat{y}_t = k \mid x_t, \mathbf{w}) = \sum_{j=1}^{C} T_{jk} p(y_t = j \mid x_t, \mathbf{w}), \quad (2)$$
$$\Rightarrow p(\widehat{y}_t \mid x_t, \mathbf{w}) = p(y_t \mid x_t, \mathbf{w}) T.$$

Given the NTM (T), we modify and correct the self-training loss $\mathcal{L}_{seg}^{\mathcal{T}}(X_{\mathcal{T}})$ in Eq. (1) with respect to noisy pseudo-labeled target data as

$$\mathcal{L}_{LC}^{\mathcal{T}}(X_{\mathcal{T}}, \widehat{Y}_{\mathcal{T}}) = -\sum_{t \in \mathcal{T}} \widehat{y}_t \log[f(x_t, \mathbf{w}) T]. \quad (3)$$

This corrected loss function encourages the similarity between noise adapted posterior class probabilities and the noisy pseudo labels. It is obvious that once the NTM is obtained, we can recover the desired estimation of class posterior probability $p(y_t | x_t, \mathbf{w})$ by the softmax output $f(x_t, \mathbf{w})$ even training the segmentation model with noisy data.

3.3. Domain-aware Meta Loss Correction (DMLC)

The effectiveness of loss correction methods highly depends on the estimation of NTM (T). Some previous attempts constructed T with a strong assumption on the noise type [16, 31], which impeded the generalization capability of model to complicated label noise. Other works required a set of clean labeled data to guide the estimation of T [16, 37, 43]. For example, Gold Loss Correction [16] utilizes the mean probability of all samples in the clean set categorized to class i to approximate T. This requirement makes the loss correction algorithm infeasible to be applied to unsupervised learning task directly.

To heuristically explore the inter-class noise transition probabilities, we devise a Domain-aware Meta-learning strategy to enable Loss Correction (DMLC) for UDA task. DMLC alternatively estimates T by minimizing the empirical risk on the domain-invariant meta data with clean labels and optimizes the segmentation net with supervision signal corrected by previously approximated T on the unlabeled target data. The key idea of DMLC lies in the estimation of T, and we first construct a set of meta data set $\{X_{\mathcal{M}}, M_{\mathcal{M}}\} = \{x_m, y_m\}_{m \in \mathcal{M}}$ with clean labels, representing the meta-knowledge of underlying label distribution of clean samples. Due to the lack of annotation in target domain, we attempt to select domain-invariant source data to

construct such a meta data set. In particular, an additional pre-trained domain predictor $g(\cdot)_{\mathbf{u}}$ and a threshold coefficient τ are introduced to sample target-like source pixels as the meta data, as illustrated in Figure 2. Only those samples with domain predictions $g(x_s, \mathbf{u})$ larger than τ are involved in meta-learning procedure. With the constructed meta data set, NTM can be updated to T^* via:

$$T^* = \underset{T \in [0,1]^{c \times c}}{\arg\min} - \sum_{m \in \mathcal{M}} y_m \log f(x_m, \mathbf{w}(T)^*),$$
$$where \ \mathbf{w}(T)^* = \underset{\mathbf{w}}{\arg\min} - \sum_{t \in \mathcal{T}} \widehat{y}_t \log[f(x_t, \mathbf{w}) T], \quad (4)$$

where $\mathbf{w}(T)^*$ represents the optimal segmentation net with the minimal corrected loss on the noisy pseudo-labeled target data, and the updated T^* should minimize the empirical loss on meta data with the optimal segmentation net [36, 37, 43]. Intuitively, during the optimization procedure of segmentation net, it is difficult to distinguish hard samples and noisy samples since both of them can produce large loss values, leading to the overfitting of noise. Guiding the estimation of T via risk minimization on meta data set, our method can avoid the wrong supervision signals and boost the generalization ability of adapted model.

With the estimated T^*, the noisy pseudo-labeled target data can be effectively utilized to optimize the segmentation net. Jointly optimizing the segmentation net on source and target data, the proposed DMLC model can be extended from self-training in Eq. (1) and our objective function for UDA can be formulated as:

$$\mathcal{L}_{DMLC} = \mathcal{L}_{seg}^{\mathcal{S}}(X_{\mathcal{S}}, Y_{\mathcal{S}}) + \mathcal{L}_{DMLC}^{\mathcal{T}}(X_{\mathcal{T}}, \widehat{Y}_{\mathcal{T}})$$
$$= -\sum_{s \in \mathcal{S}} y_s \log f(x_s, \mathbf{w}) - \sum_{t \in \mathcal{T}} \widehat{y}_t \log[f(x_t, \mathbf{w}) T^*]. \quad (5)$$

Jointly optimizing NTM associated with the segmentation net, the proposed DMLC can simultaneously estimate the noise distribution in pseudo labels and perform loss correction to target low segmentation error on target data.

3.4. Alternating Optimization for DMLC

To synergically optimize the NTM (T) and the segmentation net $f(\cdot)\mathbf{w}$, we adopt an alternating optimization strategy, and the training procedure consists of three steps: *virtual optimization, meta optimization, actual optimization*, as in Figure 3. The virtual and meta optimizations aim to optimize T, and the actual optimization is to update parameters in the segmentation net with fixed T.

During virtual optimization step (Figure 3 (a)), a meta net is copied from segmentation net with parameters $\mathbf{w}^i$, and a mini-batch of target images is forwarded through the meta net. Then the parameters in the meta net are updated by moving the current $\mathbf{w}^i$ along the gradient descent direction of corrected loss function with learning rate γ_v:

$$\hat{\mathbf{w}}^{i+1}(T^i) = \mathbf{w}^i + \gamma_v \nabla_{\mathbf{w}} \sum_{t \in \mathcal{T}} \widehat{y}_t \log[f(x_t, \mathbf{w}^i) T^i]. \quad (6)$$
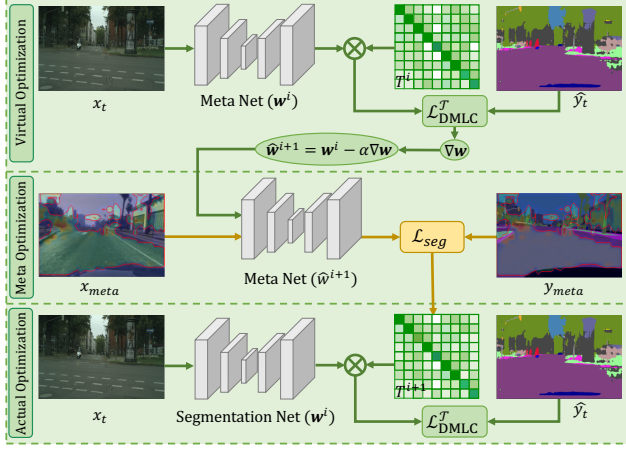
Figure 3. Illustration of alternating optimization strategy for DMLC, including three steps: virtual optimization, meta optimization and actual optimization.

Note that this is a 'virtual' step, indicating parameters in segmentation net are not actually updated to $\hat{\mathbf{w}}^{i+1}$.

Similar to well-known MAML [12] with second-order back-propagation, in the meta optimization step (Figure 3 (b)), we update T by minimizing the cross entropy loss on meta data with the feedback from the updated parameters $\hat{\mathbf{w}}^{i+1}$ as follows:

$$\widetilde{T}^{i+1} = T^i + \gamma_m \nabla_T \sum_{m \in \mathcal{M}} y_m \log f(x_m, \hat{\mathbf{w}}^{i+1}(T^i)), \quad (7)$$

where γ_m is the learning rate of meta optimization. The intuition behind the meta optimization step is to obtain an optima of $\widetilde{T}^{i+1}$ with a low empirical risk and a high generalization ability. After the back propagation updating parameters, $\widetilde{T}^{i+1}$ may contains negative values. Therefore, we first utilize $\widetilde{T}^{i+1} = \max\left(\widetilde{T}^{i+1}, 0\right)$ to enable the non-negative matrix and then perform normalization along the row direction $T_{jk}^{i+1} = \widetilde{T}_{jk}^{i+1}/\sum \widetilde{T}_{j\cdot}^{i+1}$ to ensure transition probabilities of class j are summed to 1.

In the actual optimization step (Figure 3 (c)), the noisy pseudo-labeled target data and the labeled source data are simultaneously used to optimize the segmentation net via:

$$\begin{aligned} \mathbf{w}^{i+1} = \mathbf{w}^i &+ \gamma_a \nabla_{\mathbf{w}} \sum_{s \in \mathcal{S}} y_s \log f(x_s, \mathbf{w}) \\ &+ \gamma_a \nabla_{\mathbf{w}} \sum_{t \in \mathcal{T}} \widehat{y}_t \log[f(x_t, \mathbf{w}) T^{i+1}], \end{aligned} \quad (8)$$

where γ_a is the learning rate. Through the alternating optimization strategy, both the NTM (T) and the segmentation net $f(\cdot)_{\mathbf{w}}$ can be gradually ameliorated based on the optimal solution computed in the last step.

3.5. MetaCorrection

Since low-level layers usually contain detailed features while high-level features often encode task-specific infor-

mation, it may be beneficial to transfer the knowledge from the deep layer to guide the adaptation of low-level features. Previous deep supervision approaches [4, 40] directly forces the low-level output to mimic the one-hot pseudo label computed from high-level output layer, which may bring about supervision bias and eliminate useful detailed information in low-level layer due to the capacity gap [8, 25].

Aiming at solving the above problem, we incorporate the proposed DMLC to generate the matched and compatible supervision signal for the low-level features to enhance the adaptation. In particular, we introduce additional NTM for the loss correction and utilize the domain-aware meta-learning method to estimate the specific noise distribution for shallow supervisions, thereby bridging the gap between the low-level outputs and the pseudo label obtained from deep features. The overall training objective of our Meta-Correction framework can be extended from Eq. (5) to be:

$$\mathcal{L}_{MC} = \mathcal{L}_{seg}^{\mathcal{S}}(X_{\mathcal{S}}, Y_{\mathcal{S}}) + \sum_l \alpha_l \mathcal{L}_{DMLC}^{\mathcal{T}(l)}(X_{\mathcal{T}}, \widehat{Y}_{\mathcal{T}}), \quad (9)$$

where l indicates the level used for DMLC and α_l is the weighting factor for supervision signal of lth layer. Note that $l = 0$ denotes the output layer.

With the proposed MetaCorrection framework, NTMs embedded in the segmentation net are capable of assessing the individual noise distributions in pseudo labels for different layers. Moreover, our MetaCorrection framework can obtain corrected loss functions in a data-driven manner, and these generated compatible supervision signals for different levels of features can further boost the learning of model.

4. Experiments

4.1. Datasets

We evaluate the performance of our methods on two challenging synthetic-to-real UDA semantic segmentation tasks and a medical image segmentation task. Two synthetic datasets, GTA5 [33] and SYNTHIA [34], and a real dataset, Cityscapes [10], are utilized to perform UDA synthetic-to-real semantic segmentation tasks, including two scenarios: GTA5→Cityscapes and SYNTHIA→Cityscapes. Moreover, two public prostate MRI datasets are adopted to perform UDA from Decathlon [38] to NCI-ISBI13 [29].

GTA5 contains 24,966 images captured from a video game. Pixel-wise annotations with 33 classes are provided, but only 19 classes are utilized for compatibility with Cityscapes. **SYNTHIA** consists of 9,400 synthetic images, and annotations with 16 classes are used for adaptation. **Cityscapes** is a real-world semantic segmentation dataset collected in driving scenarios. Training set, including 2,975 unlabeled images, is regarded as the target domain data for training. Evaluations are performed on 500 validation images with manual annotations. **Decathlon** is a comprehensive medical image segmentation dataset, including

Table 1. Results of adapting GTA5 to Cityscapes. The mechanism 'AL' and 'ST' stand for adversarial learning and self-training.

Methods	mech.	road	sidewalk	building	wall	fence	pole	traffic lgt	traffic sgn	veg.	terrain	sky	person	rider	car	truck	bus	train	motor	bike	mIoU
										GTA5 → Cityscapes											
AdaptSegNet [40]	AL	86.5	36.0	79.9	23.4	23.3	23.9	35.2	14.8	83.4	33.3	75.6	58.5	27.6	73.7	32.5	35.4	3.9	30.1	28.1	42.4
PatchAlign [41]	AL	92.3	51.9	82.1	29.2	25.1	24.5	33.8	33.0	82.4	32.8	82.2	58.6	27.2	84.3	33.4	46.3	2.2	29.5	32.3	46.5
LTIR [20]	AL	**92.9**	55.0	85.3	34.2	31.1	34.9	40.7	34.0	85.2	_40.1_	87.1	61.0	31.1	82.5	32.3	42.9	0.3	36.4	46.1	50.2
CBST [51]	ST	91.8	53.5	80.5	32.7	21.0	34.0	28.9	20.4	83.9	34.2	80.9	53.1	24.0	82.7	30.3	35.9	16.0	25.9	42.8	45.9
CRST [50]	ST	91.0	55.4	80.0	33.7	21.4	37.3	32.9	24.5	85.0	34.1	80.8	57.7	24.6	84.1	27.8	30.1	26.9	26.0	42.3	47.1
MaxSquare [4]	ST	89.4	43.0	82.1	30.5	21.3	30.3	34.7	24.0	85.3	39.4	78.2	**63.0**	22.9	84.6	_36.4_	43.0	5.5	34.7	33.5	46.4
MLSL [17]	ST	89.0	45.2	78.2	22.9	27.3	_37.4_	46.1	_43.8_	82.9	18.6	61.2	60.4	26.7	85.4	35.9	44.9	_36.4_	**37.2**	49.3	49.0
PyCDA [23]	ST	90.5	36.3	84.4	32.4	28.7	34.6	36.4	31.5	**86.8**	37.9	78.5	62.3	21.5	_85.6_	27.9	34.8	18.0	22.9	49.3	47.4
IntraDA [30]	ST	90.6	37.1	82.6	30.1	19.1	29.5	32.4	20.6	_85.7_	**40.5**	79.7	58.7	31.1	**86.3**	31.5	_48.3_	0.0	30.2	35.8	46.3
CAG-UDA [46]	ST	90.4	51.6	83.8	34.2	27.8	**38.4**	25.3	**48.4**	85.4	38.2	78.1	58.6	**34.6**	84.7	21.9	42.7	**41.1**	29.3	37.2	50.2
Source only	–	75.8	16.8	77.2	12.5	21.0	25.5	30.1	20.1	81.3	24.6	70.3	53.8	26.4	49.9	17.2	25.9	6.5	25.3	36.0	36.6
Ours (single DMLC)	ST	92.5	_55.1_	_85.9_	_36.9_	_32.4_	34.7	41.4	37.0	85.3	37.8	_87.4_	62.7	_31.8_	84.5	**36.8**	48.2	2.2	34.3	47.3	_51.2_
Ours (MetaCorrection)	ST	_92.8_	**58.1**	**86.2**	**39.7**	**33.1**	36.3	_42.0_	38.6	85.5	37.8	**87.6**	_62.8_	31.7	84.8	35.7	**50.3**	2.0	_36.8_	**48.0**	**52.1**

32 prostate MRI scans obtained from 3T (Siemens TIM). Annotations outline the peripheral zone (PZ) and transition zone (TZ). **NCI-ISBI13** consists of 40 labeled prostate MRI scans obtained from 1.5 T (Philips Achieva). We utilized 30 training scans as unlabeled target data to perform UDA training and other 10 scans for evaluation.

4.2. Network and Training Details

Segmentation Net We adopt DeepLab-v2 [3] backbone with pre-trained ResNet-101 [15] encoder as our segmentation net. Subsequently, Atrous Spatial Pyramid Pooling (ASPP) is employed after the last layer of encoder with dilated rates {6, 12, 18, 24}. Finally, an up-sampling layer along with a softmax operator is applied to obtain the final segmentation result with the matched size of input image.

We construct the above-mentioned segmentation net and apply the NTM to the output layer as our single DMLC model. For the MetaCorrection framework, we additionally extract low-level feature maps from the $conv4$ layer of ResNet-101 and introduce an ASPP module as the auxiliary classifier with output $f^1(x_t, \mathbf{w})$. An additional NTM $T^{(1)}$ is incorporated to generate a compatible supervision signal for low-level output. α_0 and α_1 in Eq. (9) are set as 1, 0.1.

Domain Predictor The feature maps extracted from the encoder in the segmentation net are utilized for the pixel-level domain prediction. We adopt a similar structure with DCGAN [32], which is composed of five cascaded 4×4 convolution layers with output channel numbers {64, 128, 256, 512, 1}. Then the domain prediction is obtained with the same resolution of the input image, and the threshold coefficient τ is set as 0.5 to construct the meta data set.

Implementation Details Our methods are implemented with the PyTorch library on Nvidia Tesla V100. The Stochastic Gradient Descent is utilized as our optimizer, where the momentum is 0.9 and the weight decay is 1e-3. We adopt polynomial learning rate scheduling to optimize the segmentation net with the initial learning rate of $\gamma_a = 2.5e - 4$. The batch size is set as 3, and the maximum iteration number is 150000. For the updating of meta net, $\gamma_v = 1e - 4$ and $\gamma_m = 0.11$ are set in our implementation.

The performances of our methods in synthetic-to-real scenarios are evaluated by the widely utilized performance metrics, intersection-over-union (IoU) of each class and the mean IoU (mIoU). For the prostate zonal segmentation, $Dice$ scores for PZ, TZ and the whole prostate (WP) are employed to measure the accuracy of segmentation results.

4.3. Results on GTA5→Cityscapes

We first verify the effectiveness of our approachs in the GTA5→Cityscapes scenario, and the corresponding comparison results are listed in Table 1 with the first and second best results highlighted in bold and underline. For a fair comparison, all the competed models adopt DeepLab-v2 backbone network with pre-trained ResNet-101 as encoder. Overall, our MetaCorrection framework surpasses all other models with a promising mIoU of 52.1%, outperforming the model trained only on the source data by a significant increment of 15.5% in mIoU. Compared with domain alignment methods [40, 41, 20], the proposed method shows superior performance. For example, PatchAlign [41] leverages the patch-level information to encourage the domain alignment, yielding 46.5% mIoU, which is inferior to our approach. Our MetaCorrection model, as a self-training based method, also outperforms other related pseudo label learning works [51, 4, 17, 23, 30, 46], demonstrating the effectiveness of the proposed method in alleviating the noise problem. Moreover, the proposed methods also show superior performance in terms of the per class IoU score, especially in the minor categories (e.g., 'motor') and ambiguous (e.g., 'road' and 'sidewalk') categories.

4.4. Results on SYNTHIA → Cityscapes

We then utilize SYNTHIA as the source domain data and display comparison results of our methods and other state-of-the-art methods [40, 41, 20, 51, 4, 17, 23, 30, 46] on the validation set of Cityscapes, as listed in the Table 2. We consider the IoU and mIoU of both the 16 classes and a subset of 13 classes following the standard experimental setting [30]. Since the domain shift is more evident in this scenario, the performance is slightly worse. Our MetaCor-

Table 2. Results of adapting SYNTHIA to Cityscapes. mIoU* denotes the mean IoU of 13 classes, excluding the classes with ∗.

									SYNTHIA → Cityscapes										
Methods	mech.	road	sidewalk	building	wall*	fence*	pole*	light	sign	veg.	sky	person	rider	car	bus	mbike	bike	mIoU	mIoU*
AdaptSegNet [40]	AL	84.3	42.7	77.5	–	–	–	4.7	7.0	77.9	82.5	54.3	21.0	72.3	32.2	18.9	32.3	–	46.7
PatchAlign [41]	AL	82.4	38.0	78.6	8.7	0.6	26.0	3.9	11.1	75.5	84.6	53.5	21.6	71.4	32.6	19.3	31.7	40.0	46.5
LTIR [20]	AL	**92.6**	**53.2**	79.2	–	–	–	1.6	7.5	78.6	84.4	52.6	20.0	82.1	34.8	14.6	39.4	–	49.3
CBST [51]	ST	68.0	29.9	76.3	10.8	1.4	33.9	22.8	29.5	77.6	78.3	60.6	28.3	81.6	23.5	18.8	39.8	42.6	48.9
CRST [50]	ST	67.7	32.2	73.9	10.7	1.6	**37.4**	22.2	31.2	80.8	80.5	60.8	**29.1**	82.8	25.0	19.4	45.3	43.8	50.1
MaxSquare [4]	ST	82.9	40.7	80.3	10.2	0.8	25.8	12.8	18.2	82.5	82.2	53.1	18.0	79.0	31.4	10.4	35.6	41.4	48.2
MLSL [17]	ST	59.2	30.2	68.5	**22.9**	1.0	36.2	**32.7**	28.3	**86.2**	75.4	**68.6**	27.7	82.7	26.3	**24.3**	**52.7**	45.2	51.0
PyCDA [23]	ST	75.5	30.9	83.3	20.8	0.7	32.7	27.3	33.5	84.7	85.0	64.1	25.4	85.0	45.2	21.2	32.0	**46.7**	53.3
IntraDA [30]	ST	84.3	37.7	79.5	5.3	0.4	24.9	9.2	8.4	80.0	84.1	57.2	23.0	78.0	**38.1**	20.3	36.5	41.7	48.9
CAG-UDA [46]	ST	84.7	40.8	81.7	7.8	0.0	35.1	13.3	22.7	84.5	77.6	64.2	27.8	80.9	19.7	22.7	48.3	44.5	51.5
Source only	–	55.6	23.8	74.6	9.2	0.2	24.4	6.1	12.1	74.8	79.0	55.3	19.1	39.6	23.3	13.7	25.0	33.5	38.6
Ours (single DMLC)	ST	92.3	53.0	80.2	7.7	**2.8**	26.9	11.4	8.1	83.1	**85.2**	58.9	20.5	**85.5**	35.9	21.0	41.8	44.6	52.1
Ours (MetaCorrection)	ST	**92.6**	52.7	**81.3**	8.9	2.4	28.1	13.0	7.3	83.5	85.0	60.1	19.7	84.8	37.2	21.5	43.9	45.1	52.5

Table 3. Results of adapting Decathlon to NCI-ISBI13.

Method	mech.	PZ ($Dice$)	TZ ($Dice$)	WP ($Dice$)
CBST [51]	ST	38.22	70.14	64.31
MRENT [50]	ST	40.82	72.39	67.68
MaxSquare [4]	ST	37.45	69.61	63.34
Source only	–	28.48	52.57	47.56
Ours (single DMLC)	ST	42.03	74.09	69.38
Ours (MetaCorrection)	ST	**43.25**	**74.31**	**70.87**

Table 4. Impact of different pseudo labels. 'Pseudo Label' denotes we employ pseudo labels generated by the corresponding model.

Method	Pseudo Label	GTA5 → Cityscapes	Δ
AdaptSegNet [40]	—	42.4	–
Self-training (MRENT [50])	AdaptSegNet	45.1	2.7
Self-training (Threshold [51])	AdaptSegNet	44.4	2.0
Self-training (Ucertainty [48])	AdaptSegNet	46.1	3.7
Ours (single DMLC)	AdaptSegNet	45.9	3.5
Ours (MetaCorrection)	AdaptSegNet	**47.3**	**4.9**
LTIR [20]	—	50.2	–
Self-training (MRENT [50])	LTIR	50.6	0.4
Ours (single DMLC)	LTIR	51.2	1.0
Ours (MetaCorrection)	LTIR	**52.1**	**1.9**
Source only	—	36.6	–
Self-training (MRENT [50])	Source	39.6	3.0
Ours (single DMLC)	Source	43.8	7.2
Ours (MetaCorrection)	Source	**44.5**	**7.9**

rection framework still achieves promising results in comparison to other competed methods. Specifically, the proposed method achieves 45.1% mIoU of 16 categories and 52.5% mIoU*of 13 categories.

4.5. Results on Decathlon → NCI-ISBI13

Domain discrepancy is common in clinical practice, e.g., MRIs obtained from different scanners and sites. Hence, we further assess the performance of our methods on UDA prostate segmentation, and the quantitative comparison results are listed in Table 3. It is observed that our methods exhibit superior segmentation performance in comparison to the self-training based methods, CBST [51], MRENT [50], MaxSquare [4]. For example, both single DMLC and MetaCorrection methods outperform CBST with increments of 5.07%, 6.56% in WP $Dice$ score. This observation proves the impact of our approaches in medical image analysis.

4.6. Ablation Study

Ablation Experiments. To investigate the effects of individual components of our proposed model, we design ablation studies under three adaptation scenarios (Table 1, 2, 3) and with three baseline models (Table 4). Compared with the 'Source only' lower bound, our baseline network with single DMLC boosts the mIoU to 51.2% with an increment of 14.6% in GTA5 → Cityscapes case. Then we introduce auxiliary supervision signals for low-level layers, which also contributes to the performance gain and increases the mIoU to 52.1%, as in Table 1. Moreover, incorporating DMLC to different baseline models, the MetaCorrection framework consistently improves the performance over single DMLC in terms of mIoU score, as in Table 4.

Comparison with Self-training based UDA Models. To validate the effectiveness of DMLC, we compare it with three typical self-training based UDA models [48, 50, 51]. As listed in Table 4, the proposed MetaCorrection framework (row 7) is superior to other self-training methods, including entropy minimization [50] (row 3), handcrafted threshold [51] (row 4), uncertainty based rectification [48] (row 5), yielding increments of 2.2%, 2.9%, 1.2% mIoU. Through entropy minimization, considerable noisy labels inevitably result in unsatisfactory performance. Other self-training based methods [48, 51] filtered out noisy samples with respect to the confidence score and uncertainty, but lost useful information in those omitted samples. Therefore, the increments are owing to that our method can preserve all data distribution and distill effective information from all samples with learned NTM for unbiased self-training.

Robustness to Various Types of Noise. We further explore the robustness of our MetaCorrection framework under different types of noise. Specifically, we adopt AdaptSegNet [40], LTIR [20], source only model to generate noisy pseudo labels and apply our MetaCorrection to mitigate the noise problem. As in Table 4, these three models have significantly improved performance with the proposed MetaCorrection framework. For example, MetaCorrection (row 7) improves the performance of AdaptSegNet (row 2) from 42.4% to 47.3% mIoU.
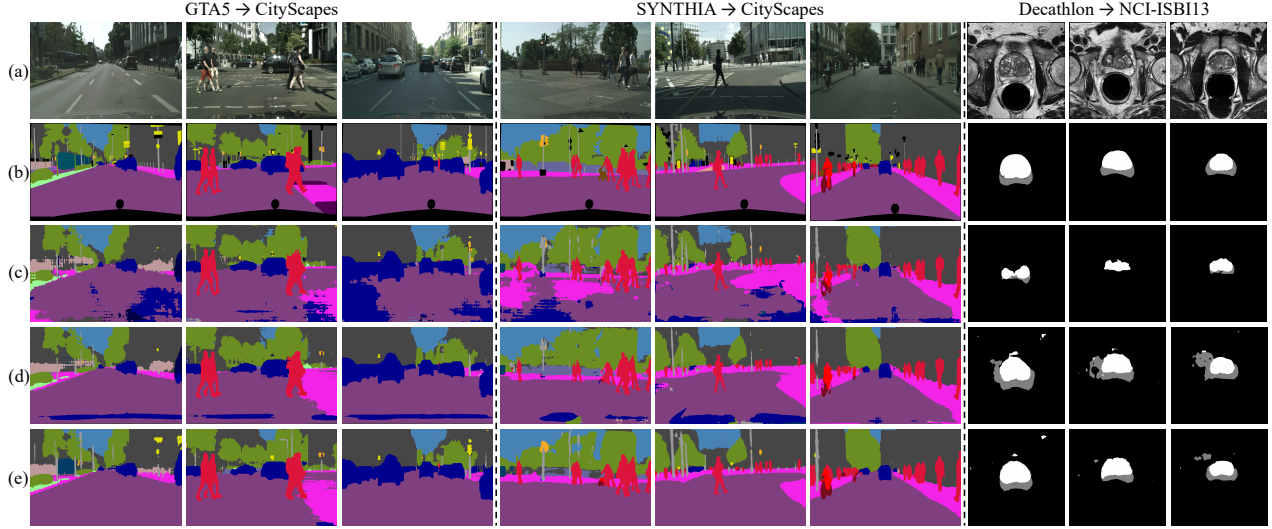
Figure 4. Qualitative results of UDA semantic segmentation. (a) Target image, (b) Ground truth, Predictions from (c) source only model, (d) self-training based MRENT model [50], (e) ours (MetaCorrection).
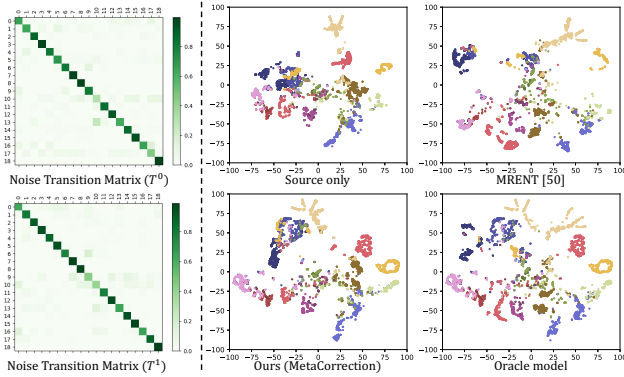

Figure 5. Left: Visualization of NMTs $T^{(0)}$ and $T^{(1)}$. Right: The t-SNE visualization of embedded features.

4.7. Visualization Results

Segmentation Visualization. As illustrated in Figure 4, we provide some typical qualitative segmentation results of target data on all three benchmarks. Obviously, the self-training method [50] could significantly promote the performance in comparison to the source model. Besides, in contrast to the baseline self-training method with conventional entropy minimization, the proposed MetaCorrection framework has better scalability to confused categories (e.g., 'rider' and 'bike') and small-scale objectives (e.g., 'traffic sgn'). We speculate the reason is that pseudo labels usually contain considerable noises in the minor categories and ambiguous categories. The proposed method rectifies the supervision signals and prevent such mistakes, leading to more reasonable segmentation predictions.

NTM Visualization. We visualize the learned NTMs of output layer (T^0) and shallow layer (T^1), as in Figure 5. It is obvious that different layers exhibit variant noise transition probability, indicating the varying noise distributions

of deep and shallow layers. The proposed MetaCorrection framework can generate matched supervision signals for individual layers to enhance the deep adaptation.

Feature Visualization. We use t-SNE [27] to visualize the feature representations of source only model, self-training [50], our MetaCorrection and oracle model (i.e., fine-tuning the segmentation net with the labeled target data), as illustrated in Figure 5. It is observed that our MetaCorrection model obtains the most matched feature distribution with that of the oracle model in comparison to source only and self-training [50] models. This observation demonstrates that our method can provide correct supervision signal for target data through the learnable NTM. Moreover, our feature representations exhibit the clearest clusters compared with other baseline methods, revealing the discriminative capability of our method.

5. Conclusion

In this paper, we have proposed a MetaCorrection framework, where the Domain-aware Meta Loss Correction (DMLC) strategy is advanced for UDA in the context of semantic segmentation, aiming for addressing the noise problem in self-training based UDA methods. The DMLC incorporates a learnable noise transition matrix (NTM) to bridge the noisy pseudo labels and ground truth labels for loss correction of the target domain, and NTM is derived through the proposed domain-aware meta-learning strategy in a data-driven manner. The model-agnostic DMLC can be flexibly applied to other models and datasets. Moreover, we consider the capacity gap between deep and shallow layers, and provide compatible supervisions for different levels to ensure the deep adaptation of the proposed MetaCorrection. The experimental results demonstrate that our methods achieve superior results to state-of-the-art methods.

References

[1] Eric Arazo, Diego Ortego, Paul Albert, Noel O'Connor, and Kevin McGuinness. Unsupervised label noise modeling and loss correction. In *ICML*, pages 312–321, 2019. 3

[2] Cheng Chen, Qi Dou, Hao Chen, Jing Qin, and Pheng-Ann Heng. Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation. In *AAAI*, volume 33, pages 865–872, 2019. 1, 2

[3] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.*, 40(4):834–848, 2017. 6

[4] Minghao Chen, Hongyang Xue, and Deng Cai. Domain adaptation for semantic segmentation with maximum squares loss. In *ICCV*, pages 2090–2099, 2019. 1, 2, 5, 6, 7

[5] Minghao Chen, Shuai Zhao, Haifeng Liu, and Deng Cai. Adversarial-learned loss for domain adaptation. In *AAAI*, pages 3521–3528, 2020. 1

[6] Yuhua Chen, Wen Li, Xiaoran Chen, and Luc Van Gool. Learning semantic segmentation from synthetic data: A geometrically guided input-output adaptation approach. In *CVPR*, pages 1841–1850, 2019. 1, 2

[7] Yun-Chun Chen, Yen-Yu Lin, Ming-Hsuan Yang, and Jia-Bin Huang. Crdoco: Pixel-level domain transfer with cross-domain consistency. In *CVPR*, pages 1791–1800, 2019. 1, 2

[8] Jang Hyun Cho and Bharath Hariharan. On the efficacy of knowledge distillation. In *ICCV*, pages 4794–4802, 2019. 5

[9] Jaehoon Choi, Taekyung Kim, and Changick Kim. Self-ensembling with gan-based data augmentation for domain adaptation in semantic segmentation. In *ICCV*, pages 6830–6840, 2019. 1, 2

[10] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *CVPR*, pages 3213–3223, 2016. 1, 5

[11] Liang Du, Jingang Tan, Hongye Yang, Jianfeng Feng, Xiangyang Xue, Qibao Zheng, Xiaoqing Ye, and Xiaolin Zhang. Ssf-dan: Separated semantic feature based domain adaptation network for semantic segmentation. In *ICCV*, pages 982–991, 2019. 1, 2

[12] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *ICML*, pages 1126–1135, 2017. 5

[13] Jacob Goldberger and Ehud Ben-Reuven. Training deep neural-networks using a noise adaptation layer. In *ICLR*, 2017. 3

[14] Xiaoqing Guo and Yixuan Yuan. Semi-supervised wce image classification with adaptive aggregated attention. *Med. Image Anal.*, page 101733, 2020. 2

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016. 6

[16] Dan Hendrycks, Mantas Mazeika, Duncan Wilson, and Kevin Gimpel. Using trusted data to train deep networks on labels corrupted by severe noise. In *NeurIPS*, pages 10456–10465, 2018. 3, 4

[17] Javed Iqbal and Mohsen Ali. Mlsl: Multi-level self-supervised learning for domain adaptation with spatially independent and semantically consistent labeling. In *WACV*, pages 1864–1873, 2020. 1, 2, 3, 6, 7

[18] Stephen James, Paul Wohlhart, Mrinal Kalakrishnan, Dmitry Kalashnikov, Alex Irpan, Julian Ibarz, Sergey Levine, Raia Hadsell, and Konstantinos Bousmalis. Sim-to-real via sim-to-sim: Data-efficient robotic grasping via randomized-to-canonical adaptation networks. In *CVPR*, pages 12627–12637, 2019. 1

[19] Lu Jiang, Zhengyuan Zhou, Thomas Leung, Li-Jia Li, and Li Fei-Fei. Mentornet: Learning data-driven curriculum for very deep neural networks on corrupted labels. In *ICML*, pages 2304–2313, 2018. 3

[20] Myeongjin Kim and Hyeran Byun. Learning texture invariant representation for domain adaptation of semantic segmentation. In *CVPR*, pages 12975–12984, 2020. 1, 2, 6, 7

[21] Hao Li and Maoguo Gong. Self-paced convolutional neural networks. In *IJCAI*, pages 2110–2116, 2017. 3

[22] Yunsheng Li, Lu Yuan, and Nuno Vasconcelos. Bidirectional learning for domain adaptation of semantic segmentation. In *CVPR*, pages 6936–6945, 2019. 1, 2

[23] Qing Lian, Fengmao Lv, Lixin Duan, and Boqing Gong. Constructing self-motivated pyramid curriculums for cross-domain semantic segmentation: A non-adversarial approach. In *ICCV*, pages 6758–6767, 2019. 1, 2, 3, 6, 7

[24] Jian Liang, Ran He, Zhenan Sun, and Tieniu Tan. Exploring uncertainty in pseudo-label guided unsupervised domain adaptation. *Pattern Recognit.*, 96:106996, 2019. 1

[25] Benlin Liu, Yongming Rao, Jiwen Lu, Jie Zhou, and Cho-Jui Hsieh. Metadistiller: Network self-boosting via meta-learned top-down distillation. In *ECCV*, pages 694–709. Springer, 2020. 5

[26] Yawei Luo, Liang Zheng, Tao Guan, Junqing Yu, and Yi Yang. Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation. In *CVPR*, pages 2507–2516, 2019. 1, 2

[27] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *J. Mach. Learn Res.*, 9(Nov):2579–2605, 2008. 8

[28] Ke Mei, Chuang Zhu, Jiaqi Zou, and Shanghang Zhang. Instance adaptive self-training for unsupervised domain adaptation. In *ECCV*, 2020. 1, 2

[29] Bloch Nicolas, Madabhushi Anant, Huisman Henkjan, Freymann John, Kirby Justin, et al. Nci-isbi 2013 challenge: Automated segmentation of prostate structures. *Cancer Imaging Arch.*, 2015. 5

[30] Fei Pan, Inkyu Shin, Francois Rameau, Seokju Lee, and In So Kweon. Unsupervised intra-domain adaptation for semantic segmentation through self-supervision. In *CVPR*, pages 3764–3773, 2020. 1, 2, 3, 6, 7

[31] Giorgio Patrini, Alessandro Rozza, Aditya Krishna Menon, Richard Nock, and Lizhen Qu. Making deep neural networks robust to label noise: A loss correction approach. In *CVPR*, pages 1944–1952, 2017. 3, 4

[32] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015. 6

[33] Stephan R Richter, Vibhav Vineet, Stefan Roth, and Vladlen Koltun. Playing for data: Ground truth from computer games. In *ECCV*, pages 102–118. Springer, 2016. 5

[34] German Ros, Laura Sellart, Joanna Materzynska, David Vazquez, and Antonio M Lopez. The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In *CVPR*, pages 3234–3243, 2016. 5

[35] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Guided curriculum model adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation. In *ICCV*, pages 7374–7383, 2019. 1, 2, 3

[36] Jun Shu, Qi Xie, Lixuan Yi, Qian Zhao, Sanping Zhou, Zongben Xu, and Deyu Meng. Meta-weight-net: Learning an explicit mapping for sample weighting. In *NeurIPS*, pages 1919–1930, 2019. 3, 4

[37] Jun Shu, Qian Zhao, Zengben Xu, and Deyu Meng. Meta transition adaptation for robust deep learning with noisy labels. *arXiv preprint arXiv:2006.05697*, 2020. 2, 3, 4

[38] Amber L Simpson, Michela Antonelli, Spyridon Bakas, Michel Bilello, Keyvan Farahani, Bram Van Ginneken, Annette Kopp-Schneider, Bennett A Landman, Geert Litjens, Bjoern Menze, et al. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *arXiv preprint arXiv:1902.09063*, 2019. 5

[39] Luan Tran, Kihyuk Sohn, Xiang Yu, Xiaoming Liu, and Manmohan Chandraker. Gotta adapt'em all: Joint pixel and feature-level domain adaptation for recognition in the wild. In *CVPR*, pages 2672–2681, 2019. 1

[40] Yi-Hsuan Tsai, Wei-Chih Hung, Samuel Schulter, Kihyuk Sohn, Ming-Hsuan Yang, and Manmohan Chandraker. Learning to adapt structured output space for semantic segmentation. In *CVPR*, pages 7472–7481, 2018. 1, 2, 5, 6, 7

[41] Yi-Hsuan Tsai, Kihyuk Sohn, Samuel Schulter, and Manmohan Chandraker. Domain adaptation for structured output via discriminative patch representations. In *ICCV*, pages 1456–1465, 2019. 1, 2, 6, 7

[42] Qian Wang and Toby Breckon. Unsupervised domain adaptation via structured prediction based selective pseudo-labeling. In *AAAI*, volume 34, pages 6243–6250, 2020. 1, 2

[43] Zhen Wang, Guosheng Hu, and Qinghua Hu. Training noise-robust deep neural networks via meta-learning. In *CVPR*, pages 4524–4533, 2020. 3, 4

[44] Zhonghao Wang, Mo Yu, Yunchao Wei, Rogerio Feris, Jinjun Xiong, Wen-mei Hwu, Thomas S Huang, and Honghui Shi. Differential treatment for stuff and things: A simple unsupervised domain adaptation method for semantic segmentation. In *CVPR*, pages 12635–12644, 2020. 1

[45] Minghao Xu, Jian Zhang, Bingbing Ni, Teng Li, Chengjie Wang, Qi Tian, and Wenjun Zhang. Adversarial domain adaptation with domain mixup. In *AAAI*, pages 6502–6509, 2020. 1

[46] Qiming Zhang, Jing Zhang, Wei Liu, and Dacheng Tao. Category anchor-guided unsupervised domain adaptation for semantic segmentation. In *NeurIPS*, pages 435–445, 2019. 1, 2, 3, 6, 7

[47] Zizhao Zhang, Han Zhang, Sercan O Arik, Honglak Lee, and Tomas Pfister. Distilling effective supervision from severe label noise. In *CVPR*, pages 9294–9303, 2020. 2, 3

[48] Zhedong Zheng and Yi Yang. Rectifying pseudo label learning via uncertainty estimation for domain adaptive semantic segmentation. *Int. J. Comput. Vis.*, pages 1–15, 2021. 1, 2, 7

[49] Yi Zhu, Zhongyue Zhang, Chongruo Wu, Zhi Zhang, Tong He, Hang Zhang, R Manmatha, Mu Li, and Alexander Smola. Improving semantic segmentation via self-training. *arXiv preprint arXiv:2004.14960*, 2020. 1, 2, 3

[50] Yang Zou, Zhiding Yu, Xiaofeng Liu, BVK Kumar, and Jinsong Wang. Confidence regularized self-training. In *ICCV*, pages 5982–5991, 2019. 1, 2, 3, 6, 7, 8

[51] Yang Zou, Zhiding Yu, BVK Vijaya Kumar, and Jinsong Wang. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In *ECCV*, pages 289–305, 2018. 1, 2, 6, 7