Unbiased Subclass Regularization for Semi-Supervised Semantic Segmentation

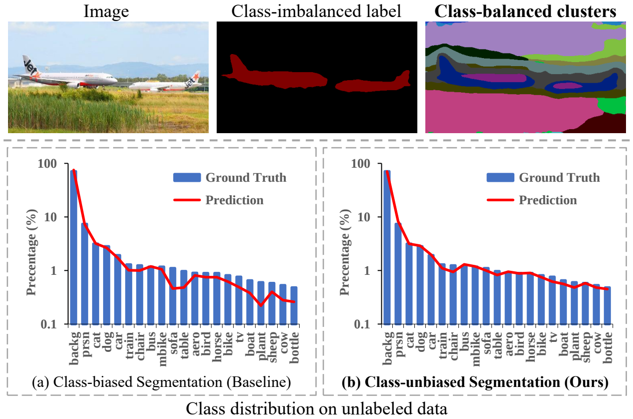
Dayan Guan, Jiaxing Huang, Aoran Xiao, Shijian Lu*
Singtel Cognitive and Artificial Intelligence Lab for Enterprises, Nanyang Technological University
{Dayan.Guan, Jiaxing.Huang, Aoran.Xiao, Shijian.Lu}@ntu.edu.sg

Abstract

Semi-supervised semantic segmentation learns from small amounts of labelled images and large amounts of unlabelled images, which has witnessed impressive progress with the recent advance of deep neural networks. However, it often suffers from severe class-bias problem while exploring the unlabelled images, largely due to the clear pixel-wise class imbalance in the labelled images. This paper presents an unbiased subclass regularization network (USRN) that alleviates the class imbalance issue by learning class-unbiased segmentation from balanced subclass distributions. We build the balanced subclass distributions by clustering pixels of each original class into multiple subclasses of similar sizes, which provide class-balanced pseudo supervision to regularize the class-biased segmentation. In addition, we design an entropy-based gate mechanism to coordinate learning between the original classes and the clustered subclasses which facilitates subclass regularization effectively by suppressing unconfident subclass predictions. Extensive experiments over multiple public benchmarks show that USRN achieves superior performance as compared with the state-of-the-art.

1. Introduction

Semantic segmentation aims to assign a human-defined class label to each pixel of an image, which is a fundamental task in the computer vision research. With the recent advance of deep neural networks [9, 19, 69], we can learn a very accurate segmentation model while a large amount of labelled training images are available. However, collecting a large amount of pixel-wise semantic labels is laborious and time-consuming, which has become a bottleneck in semantic segmentation research [11, 15, 39]. Semi-supervised semantic segmentation, which aims to learn from a small amount of labelled images and a large amount of unlabelled images, has been attracting increasing attention for addressing the image annotation challenge.



(a) Class-biased Segmentation (Baseline)

(b) Class-unbiased Segmentation (Ours)

Figure 1. Motivation of our work: In semi-supervised semantic segmentation, the segmentation model trained using *class-imbalanced labels* (of the labelled data) tends to produce *class-biased segmentation* on the unlabelled data. We create *class-balanced clusters* with balanced subclass distribution, learning from which alleviates the class imbalance issue and produces *class-unbiased segmentation* on the unlabelled data. We obtain the class-balanced clusters by clustering pixels of each original class into multiple subclasses of similar size. Best viewed in color.

Most existing studies tackle the challenge of semi-supervised semantic segmentation by applying either consistency-training [35, 46, 47] or self-training [10, 20, 29, 43, 45, 68] to the unlabelled data. However, they often suffer from constrained segmentation accuracy largely due to the segmentation model that is trained by using the labelled data. As illustrated in Fig. 1, the model trained using the labelled data is class biased due to the class-imbalance of the labelled data. This leads to class-biased segmentation of the unlabelled data which accumulates and finally degrades the whole semi-supervised learning. Though a few studies [20, 64] attempt to handle the class imbalance issue by selecting more pseudo labels for minority classes during self-training, these pseudo labels are often noisy as they are generated from class-biased segmentation. Note the class imbalance issue has been widely studied in supervised learning via re-sampling [5, 6, 34, 55, 62],

[*] Corresponding author.

re-weighting [12, 22, 38, 49] and meta-learning [49, 53, 63], but these works require labels to rectify biased predictions and are thus inapplicable to the unlabelled data in semi-supervised semantic segmentation.

In this work, we propose an unbiased subclass regularization network (USRN) that tackles the class-imbalance issue and regularizes class-biased segmentation by *generating* class-unbiased segmentation. Leveraging the segmentation backbone as learnt from the original class distribution, USRN introduces an auxiliary segmentation task as supervised by a set of class-balanced clusters for producing class-unbiased segmentation on the unlabelled data. We obtain the class-balanced clusters from the labelled data by clustering pixels of each original class into multiple subclasses of similar size. As illustrated in Fig. 1, the USRN trained using class-balanced clusters can produce clearly more class-unbiased segmentation for the unlabelled data. In addition, the segmentation with the original classes could be interfered by the segmentation with the generated subclasses due to their different convergence speeds. We design an entropy-based gate mechanism to address this issue, where the learning with the auxiliary subclasses will be stopped (i.e., no back-propagation) when the subclass predictions are less confident than the original class predictions. Extensive experiments over multiple public benchmarks demonstrate the effectiveness of our designed network.

The contribution of this work is threefold. *First*, we propose an unbiased subclass regularization network that explores class-unbiased segmentation to alleviate the class imbalance issue in semi-supervised semantic segmentation. *Second*, we design an entropy-based gate mechanism that coordinates the concurrent learning from the original classes and the generated subclasses effectively. *Third*, extensive experiments show the superior effectiveness of our designed network as compared with the state-of-the-art.

2. Related Works

2.1. Supervised Semantic Segmentation

With recent progress of deep learning, supervised semantic segmentation has made remarkable progress by designing various architectures. FCN in [42] is the first end-to-end trainable network with fully convolutional layers for semantic segmentation. The subsequent studies improve [42] by employing encoder-decoder structures [3, 9, 51], mutli-scale inputs [8, 13, 37], feature pyramid spatial pooling [41, 69], attention mechanism [16, 70] or dilated convolutions [7, 9, 61, 67]. For example, Deeplabv3+ in [9] combines low-level and high-level features to refine the object boundaries of segmentation results. However, training these supervised segmentation networks requires large amounts of annotated data which is often laborious and time-consuming to collect. Our work aims to alleviate the data annotation constraint by exploring large amounts of unlabeled data together with limited amount of labeled data.

2.2. Semi-Supervised Semantic Segmentation

Semi-supervised segmentation aims to explore tremendous unlabeled data with supervision from limited labeled data, which is most relevant to domain adaptive segmentation where labeled data is obtained from another domain [2, 18, 26–28, 60, 66]. Most existing studies address this challenge by either consistency-training [17, 25, 32, 36, 48, 58, 65, 71, 72] or self-training [1, 4, 21, 23, 24, 29, 30, 33, 50, 54, 57, 73]. Specifically, consistency-training maintains the consistency of segmentation of each unlabeled sample under different perturbations. For example, CCT [47] applies two same-structured segmentation networks with different initialization to produce differently perturbed samples. CAC [35] enforces context-aware consistency between representations from the same unlabeled image augmented with different context information. Self-training instead generates pseudo labels on unlabeled data to re-train networks. For example, GCT [31] introduces a flaw detector to correct the defects in pseudo labels. DBSN [68] designs distribution-specific batch normalization for robust pseudo labels generation. CPS [10] produces pseudo labels from one segmentation network to supervise the other segmentation network with the same structure yet different initialization. However, both consistency-training and self-training suffer from clear pixel-wise class imbalance in labeled data. Our method can mitigate the class imbalance issue in semi-supervised segmentation effectively.

2.3. Class-Imbalance Learning

Class imbalance issue has been widely studied in supervised learning. For example, re-sampling based methods [5, 6, 34, 62] re-balance the biased networks according to sample sizes for each class. Re-weighting based methods [12, 22, 38, 49] adaptively adjust the loss weight for different training samples with different class labels. Meta-learning based methods [49, 53, 63] use the validation loss calculated from selected class-balanced labeled samples as the meta objective to optimize networks. However, all these methods rely on labels to address the class imbalance issue and cannot be directly applied to unlabeled data in semi-supervised learning. Recently, several studies attempt to handle the class imbalance issue in semi-supervised learning. For example, CReST [64] selects pseudo labels more frequently for minority classes according to the estimated class distribution. DARS [20] employs adaptive threshold to select more pseudo labels for minority class during self-training. However, these methods tends to generate noisy pseudo labels from class-biased segmentation of unlabeled data. We address the class imbalance issue by constructing and learning from class-balanced subclasses.

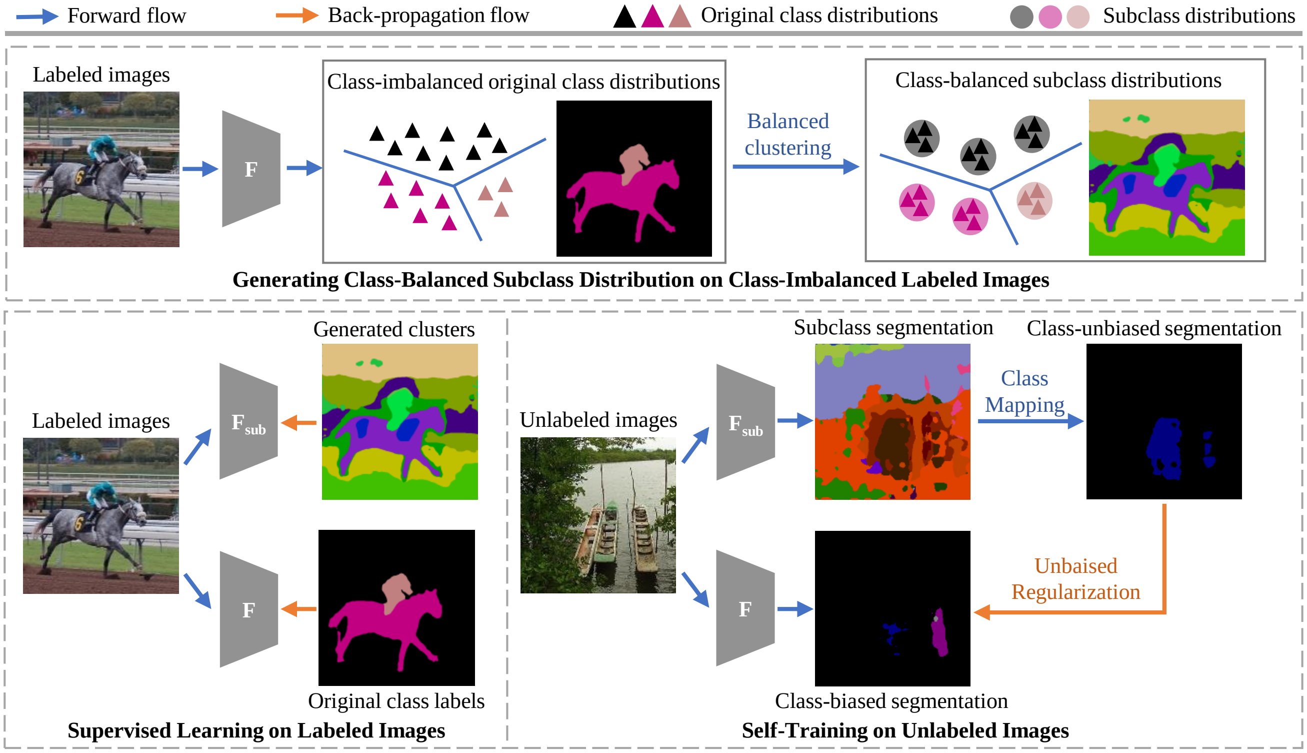

Forward flow
Back-propagation flow
Original class distributions
Subclass distributions
Labeled images
F
Class-imbalanced original class distributions
Balanced clustering
Class-balanced subclass distributions
Generating Class-Balanced Subclass Distribution on Class-Imbalanced Labeled Images
Generated clusters
Labeled images
F_sub
F
Original class labels
Supervised Learning on Labeled Images
Unlabeled images
Subclass segmentation
Class Mapping
Class-unbiased segmentation
Unbaised Regularization
Class-biased segmentation
Self-Training on Unlabeled Images

Figure 2. Overview of unbiased subclass regularization network (USRN): USRN regularizes *class-biased segmentation* from original class distribution with *class-unbiased segmentation* from subclass distributions. We generate *class-balanced subclass distribution* by clustering features from *class-imbalanced original class distribution* into multiple groups of similar size. Specifically, USRN performs supervised learning by training a class-biased model F and a class-unbiased model F_{sub} with *original class labels* and *generated clusters*, respectively, for labelled images under semi-supervised setup. For unlabelled images, USRN performs self-training by applying *class-unbiased segmentation* from F_{sub} to regularize *class-biased segmentation* from F. We obtain class-unbiased segmentation by mapping the *subclass segmentation* (as produced by F_{sub}) from subclass space to original class space. Best viewed in color.

3. Method

3.1. Problem Definition

This work focuses on semi-supervised semantic segmentation. Given images $X_l \subset \mathbb{R}^{H\times W\times 3}$ with pixel-level semantic labels $\hat{y} \subset (1, C)^{H\times W}$ and unlabelled images $X_u \subset \mathbb{R}^{H\times W\times 3}$ (H, W and C denote image height, image width and class number, respectively), the goal is to learn a segmentation model F that can fit both labelled and unlabelled data and work well on unseen images. Existing methods [10, 20, 29, 31, 35, 43, 45, 47, 73] combine supervised learning on labelled images and unsupervised learning on unlabelled image to tackle the semi-supervised challenge. For labelled images, they adopt cross entropy loss as supervised loss $\mathcal{L}_s$ to train F. For unlabelled images, they adopt consistency regularization loss [31, 47] or self-training loss [10, 20, 29, 35, 43, 45, 73] as unsupervised loss $\mathcal{L}_u$ to train F. The overall objective is a weighted combination of supervised and unsupervised losses:

$$\mathcal{L} = \mathcal{L}_s(X_l, Y) + \lambda_u \mathcal{L}_u(X_u), \qquad (1)$$

where λ_u is a balancing weight. With this objective function, supervised and unsupervised learning would benefit each other thanks to their complementary nature [59].

Though consistency-training and self-training can learn from the unlabelled images effectively, their performance is often constrained by the quality of the supervised model that is trained by using the labelled images. Specifically, the labelled images often suffer from a clear class-imbalance issue which directly leads to class-biased model and further class-biased segmentation on the unlabelled images. Such class-biased segmentation accumulates during the process of consistency-training or self-training which finally degrades the overall performance of semi-supervised semantic segmentation. We define this problem as a class-imbalance issue in semi-supervised semantic segmentation, and design a class-balanced subclass regularization network to address the class-imbalance problem.

3.2. Unbiased Subclass Regularization

We design an unbiased subclass regularization network (USRN) for addressing the class-imbalance issue in semi-supervised segmentation, as shown in Fig. 2. With labelled images in semi-supervised segmentation, USRN first trains a class-biased model F (by learning from the class-imbalanced labelled images) and then produces a class-balanced subclass distribution by clustering the F-produced features of the labelled images. With the class-balanced subclass distribution, a class-unbiased model F_{sub} can be trained which tends to produce class-unbiased segmentation while applied to unlabelled images in semi-supervised segmentation.

Generating class-balanced subclass distribution. USRN learns class-unbiased model by generating class-balanced clusters. With the labelled images (with class imbalanced annotations), USRN first trains a supervised segmentation model F and then applies F to each labelled image to extract semantic features. It then adopts balanced k-means clustering [40] to group the extracted semantic features into multiple clusters of similar size. The generated class-balanced clusters $\hat{y}^{\star} \subset (1, C_{sub})^{H \times W}$ (C_{sub} is the number of clustered subclasses) directly give a balanced subclass distribution with the labelled images. In our implementation, we empirically set the cluster size as the size of the smallest class in the original annotations.

Supervised learning on labeled data. USRN performs supervised learning with both original and subclass annotations. For each labelled image x_l, we feed a weakly augmented image $\mathcal{A}^w(x_l)$ to F to obtain original class prediction $p_l^w = F(\mathcal{A}^w(x_l))$ and the same input to F_{sub} to obtain subclass prediction $p_l^{w\star} = F_{sub}(\mathcal{A}^w(x_l))$. Here, $\mathcal{A}^w$ is a weak augmentation function, *i.e.*, random scaling, cropping and horizontal flipping. Given p_l^w with its original class label $\hat{y} \subset Y$ and $p_l^{w\star}$ with its class-balanced cluster $\hat{y}^{\star} \subset Y^{\star}$, a multi-distribution supervised loss $\mathcal{L}_s^{md}$ can be defined by:

$$\mathcal{L}_s^{md} = \mathcal{L}_{ce}(p_l^w, \hat{y}) + \lambda_{sub}\mathcal{L}_{ce}(p_l^{w\star}, \hat{y}^{\star}), \tag{2}$$

where $\mathcal{L}_{ce}$ is cross-entropy loss and λ_{sub} is a balancing weight.

Self-training on unlabeled data. USRN preforms self-training to update F with class-unbiased pseudo label as generated from the subclass distribution. For each unlabelled sample x_u, we feed a weakly augmented image $\mathcal{A}^w(x_u)$ to F to obtain original class prediction $p_u^w = F(\mathcal{A}^w(x_u))$ and the same input to F_{sub} to obtain subclass prediction $p_u^{w\star} = F_{sub}(\mathcal{A}^w(x_u))$. To generate unbiased pseudo labels for the original class supervision, we first map the prediction $p_u^{w\star}$ from the subclass space $(1, C_{sub})^{H \times W}$ to the original class space $(1, C)^{H \times W}$ (this process denoted by $\mathcal{M}$), and then define a function $\mathcal{S}$ to select pseudo labels from the mapped predictions in an online manner. We define the pseudo-label selection function $\mathcal{S}$ by:

$$\mathcal{S}(p) = \mathbb{1}_{[p^{(c)} > \gamma]}(p^{(c)}), \tag{3}$$

where p refers to the predictions, $\mathbb{1}$ is a function that returns the class index c if the condition is true or the 'ignore' class index otherwise, and γ is a confidence threshold. Note there is no back-propagation for the 'ignore' class in training.

To alleviate over-fitting in self-training, the pseudo label generated from weakly augmented version of an image $\mathcal{A}^w(x_u)$ is used to supervise the segmentation from the strongly augmented version of the same image $\mathcal{A}^s(x_u)$. Here, $\mathcal{A}^s$ is a strong augmentation function, *i.e.*, random color jitters and Gaussian blur. With p_u^w and $\hat{p}_u^{w\star}$ (one-hot vector computed from $p_u^{w\star}$ using softmax) from $\mathcal{A}^w(x_u)$, we simultaneously feed $\mathcal{A}^s(x_u)$ to F to obtain the original class prediction $p_u^s = F(\mathcal{A}^s(x_u))$, and perform subclass regularized self-training with the loss $\mathcal{L}_{st}$:

$$\mathcal{L}_{st} = \mathcal{L}_{ce}(p_u^s, \mathcal{S}(\mathcal{M}(\hat{p}_u^{w\star}) \cdot p_u^w)) \tag{4}$$

In addition, USRN performs self-training on subclass distributions to update F_{sub}. With $p_u^{w\star}$ from $p_u^{w\star}$ as in Eq. 4, we simultaneously feed $\mathcal{A}^s(x_u)$ to F_{sub} to obtain the subclass prediction $p_u^{s\star}$, and perform subclass self-training with the loss $\mathcal{L}_{st}$:

$$\mathcal{L}_{st}^{sub} = \mathcal{L}_{ce}(p_u^{s\star}, \mathcal{S}(p_u^{w\star})). \tag{5}$$

3.3. Entropy-based Gate Mechanism

The proposed USRN employs subclass predictions to regularize original-class predictions. As the subclass distributions are derived from the original-class distribution, learning from the subclass distributions is more complex and tends to be slower as compared with that from the original-class distributions under the same learning policy (*e.g.*, optimizer, learning rate, weight decay rate, etc). This could introduce undesired regularization. Specifically, the original-class learning could produce more confident and correct predictions than the subclass learning as the original-class learning converges faster than the subclass learning in training. The semi-supervised learning will degrade if the original-class predictions are regularized by the subclass predictions under such circumstance.

To address this problem, we design an entropy-based selection function to avoid regularizing the confident original-class predictions p with unconfident subclass predictions $p^{\star}$. The entropy-based selection function is defined by:

$$\mathcal{S}_e(p^{\star}, p) = \mathbb{1}_{[\mathcal{E}(p^{\star}) < \mathcal{E}(p)]}(\mathcal{S}(\hat{p}^{\star}) \cdot p), \tag{6}$$

where $\mathcal{E}$ is the entropy function as defined in [52].

Given the original predictions (*i.e.*, p_u^s and p_u^w from strongly and weakly augmented versions of the same image) and the subclass prediction (*i.e.*, $p_u^{w\star}$ from the weakly augmented version) as in Eq. 4, we reformulate the self-training loss in Eq. 4 and define an entropy-based self-training loss $\mathcal{L}_{st}^e$ as follows:

$$\mathcal{L}_{st}^e = \mathcal{L}_{ce}(p_u^s, \mathcal{S}_e(\mathcal{M}(p_u^{w\star}), p_u^w)) \tag{7}$$

Combining the losses in Eqs. 2, 5 and 7, the overall training objective of the unbiased subclass regularization network (USRN) can be formulated as follows:

$$\mathcal{L}_{\text{USRN}} = \mathcal{L}_s^{md} + \lambda_u(\mathcal{L}_{st}^{e} + \lambda_{sub}\mathcal{L}_{st}^{sub}). \tag{8}$$

4. Experiments

4.1. Experimental Setting

Datasets. We conducted main experiments on the dataset PASCAL VOC [15] by following previous work [10,31,35,47]. The dataset consists of 10,582 images for training and 1,456 images for evaluation, and the image resolution varies from 192 × 282 to 500 × 500. It provides pixel-wise annotations with 21 semantic classes. To perform comprehensive validation, we also conducted experiments on the dataset Cityscapes [11] which contains 2,975 images for training and 500 images for evaluation and all images have the same resolution of 1024 × 2048. Cityscapes provides pixel-wise labels with 19 semantic classes.

Implementation details. Both segmentation backbone model F and auxiliary segmentation model F_{sub} adopt Deeplabv3+ [9] with ResNet-50 [19] pre-trained on ImageNet [14], where F and F_{sub} share layers that extract low-level features in ResNet-50. All network models are optimized by mini-batch stochastic gradient descent (SGD) with a base learning rate of 10^{-3}, a momentum of 0.9 and a weight decay of 10^{-4}. The weak augmentation function $\mathcal{A}^w$ (*i.e.*, random scaling, cropping and horizontal flipping) and the strong augmentation function $\mathcal{A}^s$ (*i.e.*, random color jitters and Gaussian blur) are the same as in [35]. The confidence threshold γ is set to 0.75 and all the balancing weights (*i.e.*, λ_{sub} and λ_u) are directly set to 1. During evaluation, each image is tested only on the segmentation backbone and the mean intersection-over-union (mIoU) is adopted as the evaluation metric.

4.2. Comparison with the state-of-the-art

We compare USRN with state-of-the-art methods [10, 20, 31, 35, 47, 68] over PASCAL VOC and Cityscapes datasets [11, 15]. Tables 1 and 2 shows experimental results. For PASCAL VOC dataset, we randomly split 1/64, 1/32 and 1/16 of the trainset (including 165, 331 and 662 training images, respectively) as labelled data, and the remaining training images as unllabeled data. As the number of training images in Cityscapes dataset is less than that in PASCAL VOC, we randomly split 1/32, 1/16 and 1/8 of the trainset (including 93, 186 and 372 training images, respectively) in Cityscapes dataset as labelled data, and the remaining training images as unlabelled data. State-of-the-art methods are implemented with various segmentation backbones and choose different splits of the trainset in experiments. For fair comparisons, we reproduce some experimental result by using the official code so that all the methods can be compared with the same split of labelled data as well as the same segmentation backbone.

Method	Publication	1/64	1/32	1/16
Baseline	-	52.4	59.2	63.9
GCT [31]	ECCV 20	-	-	64.1
CCT [47]	CVPR 20	-	-	65.2
DARS [20]	ICCV 21	56.9	64.5	68.4
DBSN [68]	ICCV 21	57.5	64.6	69.8
CAC [35]	CVPR 21	56.5	65.1	70.1
CPS [10]	CVPR 21	57.9	64.8	68.2
USRN (Ours)	-	**61.7**	**68.6**	**72.3**
Oracle	-	76.8	76.8	76.8

Table 1. Quantitative comparison with the state-of-the-art over the dataset PASCAL VOC [15]. We randomly split 1/64, 1/32 and 1/16 of the trainset (including 165, 331 and 662 training images, respectively) as labeled data, and the remaining training images as unlabeled data for semi-supervised learning. The *Baseline* and *Oracle* are trained with supervised loss by using the split labelled training data and the whole trainset, respectively.

Method	Publication	1/32	1/16	1/8
Baseline	-	59.8	64.3	68.9
GCT [31]	ECCV 20	-	65.8	71.3
CCT [47]	CVPR 20	-	66.4	72.5
DARS [20]	ICCV 21	61.9	66.9	73.7
DBSN [68]	ICCV 21	62.2	67.3	73.5
CAC [35]	CVPR 21	62.2	69.4	74.0
CPS [10]	CVPR 21	62.5	69.8	74.4
USRN (Ours)	-	**64.6**	**71.2**	**75.0**
Oracle	-	78.3	78.3	78.3

Table 2. Quantitative comparison of USRN with the state-of-the-art over the dataset Cityscapes [11]. We randomly split 1/32, 1/16 and 1/8 of the trainset (including 93, 186 and 372 training images, respectively) as labeled data, and the remaining training images as unlabeled data for semi-supervised learning. The *Baseline* and *Oracle* are trained with supervised loss by using the split labelled training data and the whole trainset, respectively.

As Tables 1 and 2 show, the proposed USRN outperforms the state-of-the-art consistently over the two datasets with different splits of labelled training data. The superior performance is largely attributed to the proposed unbiased subclass regularization that effectively addresses the class imbalance issue in semi-supervised segmentation. For smaller splits of the labelled training data, USRN outperforms the state-of-the-art with larger margins by 3.8% and 2.1% in mIoU for 1/64 split of PASCAL VOC and 1/32 split of Cityscapes, respectively. In particular, the performance of state-of-the-art methods is largely constrained by the quality of the segmentation model that is trained by using the class-imbalanced labelled data. Since deep convolutional neural networks tends to overfit with small datasets

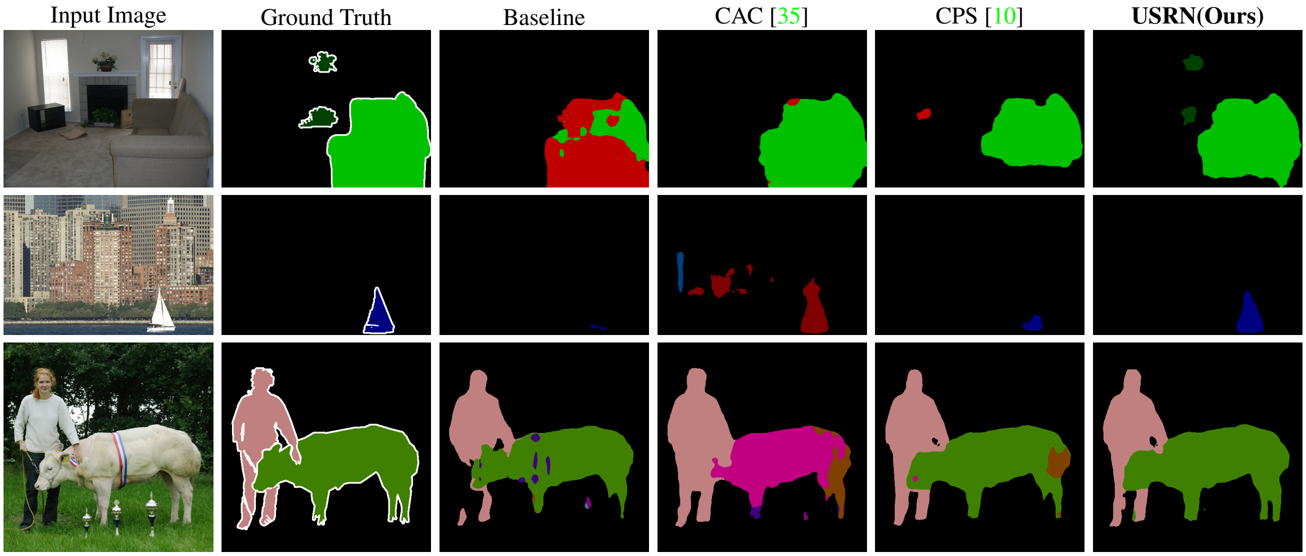


Figure 3. Qualitative comparison of USRN with the state-of-the-art over 1/32 split of PASCAL VOC dataset. USRN can obtain more accurate semantic segmentation especially for pixels that are inaccurately segmented as the most dominant class (e.g., back-ground class as visualized in black color) by state-of-the-art methods [10, 35].

as proved in [56], the class imbalance issue is more severe when training with fewer labelled data which degrades the performance of state-of-the-art methods greatly. While using larger splits of labelled data, the gaps between our method and the *Oracle* as trained by using the whole train-set are 4.5% in mIoU for 1/16 split of PASCAL VOC and 3.3% in mIoU for 1/8 split of Cityscapes. Such experimental results show that our method can learn accurate segmentation models with a small amount of labelled training data, demonstrating its potential in reducing labelling efforts in deep network training.

We also provide qualitative comparisons over 1/32 split of PASCAL VOC dataset. We compare USRN with state-of-the-art methods [10, 35] and the *Baseline* that is trained with supervised loss only. The qualitative results are well aligned with the quantitative results as illustrated in Fig. 3. It can be observed that USRN produces more accurate segmentations than state-of-the-art methods especially for those inaccurately segmented pixels that belong to the most dominant class. The qualitative experimental results further validate that USRN can better handle the class imbalance issue in semi-supervised semantic segmentation.

4.3. Ablation studies

We conducted extensive ablation studies to examine how the proposed USRN achieves the superior semi-supervised semantic segmentation. We performed all the ablation studies over 1/32 split of PASCAL VOC dataset, where USRN can achieve a mIoU of 68.6% under default settings. Specifically, we examine different designs in USRN including different USRN components, different clustering strategies for class-balanced clusters generation, sharing features in different level (between the segmentation backbone F and the auxiliary segmentation model F_{sub}), and parameter analysis of the confidence threshold γ for pseudo label selection.

Model	MSL	OST	USR	SST	EGM	mIoU
Model I	✓					59.1
Model II	✓	✓				64.1
Model III	✓	✓	✓			65.0
Model IV	✓	✓	✓	✓		67.1
USRN	✓	✓	✓	✓	✓	68.6

Table 3. Ablation study on different components (*i.e.*, MSL, OST, USR, SST and EGM) of USRN over 1/32 split of PASCAL VOC dataset. Here, MSL, OST, USR, SST and EGM are abbreviations of multi-distribution supervised learning, original self-training, unbiased subclass regularization, subclass self-training and entropy-base gate mechanism, respectively.

Different Components. We conducted ablation studies on different components of USRN to examine their effectiveness as shown in Table. 3. Specifically, we trained five models over 1/32 split of PASCAL VOC dataset including: 1) **Model I** that is trained with labeled data only using the multi-distribution supervised learning (MDL) loss $\mathcal{L}_s^{md}$ in Eq. 2; 2) **Model II** that performs self-training on original class distributions only using the MDL loss and the original self-training (OST) loss as in [10, 35, 54]; 3) **Model III** that performs unbiased subclass regularization (USR) directly on the OST in **Model II** by using the MDL loss and the proposed self-training loss $\mathcal{L}_{st}$ as defined in Eq. 4; 4)

Model IV that includes subclass self-training (SST) loss in Eq. 5 into **Model III** for training the auxiliary segmentation model on unlabelled data; and 5) **USRN** that introduces the entropy-based gate mechanism (EGM) into **Model IV** to coordinate the concurrent learning from the original classes and the generated subclasses.

As Table 3 shows, both **Model II** and **Model III** outperform **Model I** by large margins, demonstrating the effectiveness of self-training in semi-supervised segmentation. Without SST, the performance of **Model III** still outperforms **Model II**, which shows that the subclass segmentation model trained with the labelled data only can produce high-quality pseudo labels on unlabelled data. With SST, **Model IV** outperforms **Model III** by 2.1% in mIoU thanks to updating the subclass segmentation model by self-training on unlabelled data. With the updated auxiliary segmentation model, more accurate subclass segmentation can be produced to generate unbiased pseudo labels for updating the segmentation backbone. Finally, **USRN** further improves **Model IV** by 1.5% in mIoU, which validates the effectiveness of the proposed entropy-based gate mechanism.

Clustering algorithm	Original CBR	Subclass CBR	mIoU
Normal k-means	33.8%	96.4%	68.0
Balanced k-means	33.8%	**99.5%**	**68.6**

Table 4. Comparisons of normal k-means [44] with balanced k-means [40] while applying USRN to 1/32 split of PASCAL VOC dataset. Here, the class balance rate (CBR) is defined in Eq. 9, where CBR=100% means the number of pixels within each class is equal (*i.e.*, extreme class balance) and CBR=0% means all pixels are labeled with only one class (*i.e.*, extreme class imbalance).

Clustering Strategy. In Section. 3.2, we adopted balanced k-means clustering [40] to generate class-balanced subclass annotations. To measure the class balance of annotations, we define a new metric named class balance rate (CBR) which can be formulated as follows:

$$\mathrm{CBR} = 1 - \frac{\sigma_c}{\sigma_c^{\star}} = 1 - \frac{\sqrt{\frac{1}{C}\sum_{n=1}^{C} n_c^2 - (\frac{1}{C}\sum_{n=1}^{C} n_c)^2}}{\sqrt{\frac{1}{C}(\sum_{n=1}^{C} n_c)^2 - (\frac{1}{C}\sum_{n=1}^{C} n_c)^2}}, \quad (9)$$

where n_c is the number of pixels within each class $c \subset (1, C)$ for given annotations, σ_c is standard deviation of $\{n_1, n_2, \cdots, n_C\}$ and $\sigma_c^{\star}$ is standard deviation of $\{\sum_{n=1}^{C} n_c, 0, \cdots, 0\}$, *i.e.*, all pixels are labeled with only one class (extreme class imbalance).

As shown in Table 4, the CBR of subclass annotations is almost 100% which is much higher than the CBR of the original annotations. This demonstrates that we obtain class-balanced subclass annotations from class-imbalanced original annotations successfully. We can also observe that the subclass annotations generated with normal k-means [44] is also quite class-balanced (CBR=96.4%), and USRN model trained with such annotations can achieve comparable accuracy as the USRN trained with default clustering strategy (*i.e.*, balanced k-means). This shows that our method is robust to different clustering strategies.

Sharing Features	GPU Occupation	mIoU
No sharing	9.76 Gb×2	67.3
Low-level features	8.75 Gb×2	**68.6**
Both low-level and high-level features	6.99 Gb×2	67.8

Table 5. The impact of feature sharing between the segmentation backbone F and the auxiliary network F_{sub} over 1/32 split of PASCAL VOC dataset: USRN trained with default setting (*i.e.*, sharing low-level features) achieved the best mIoU with a little computation overhead during training. Note that the computational cost is equal for all the settings during inference.

Sharing Features. Recent supervised segmentation models [3,9,69,70] achieved high accuracy by integrating multi-level features. In default setting of USRN, the segmentation backbone F and the auxiliary segmentation model F_{sub} share layers that extract low-level features. We further evaluate the impact of sharing features between F and F_{sub}. As shown in Table 5, USRN trained with default setting (*i.e.*, sharing low-level features) achieves the highest accuracy in mIoU as compared with USRN trained with other settings (*i.e.*, 'no sharing' and sharing multi-level features). The setting of 'no sharing' has the lowest accuracy, which demonstrates that original class segmentation and auxiliary subclass segmentation are complementary to each other. The reason why sharing high-level features (*i.e.*, semantic features) degrades the accuracy of USRN is that original class segmentation and auxiliary subclass segmentation require to learn different semantic features as semantic information of these two tasks is different.

γ	0.55	0.65	0.75	0.85	0.95	0.99
mIoU	67.4	67.7	68.6	68.6	68.5	68.1

Table 6. Sensitivity of the confidence threshold γ in Eq. 3: USRN is stable when γ changes in a range from 0.75 to 0.95. The experiments are conducted over 1/32 split of PASCAL VOC dataset.

Parameter Analysis. The confidence threshold γ in Eq. 3 is an important hyper-parameter for generating high-quality class-unbiased pseudo labels. We evaluate USRN with different γ and Table 6 show experimental results. It can be observed that USRN is very stable when γ changes in a range from 0.75 to 0.95. While γ is smaller than 0.75, the per-

Method	back.	aero.	bicy.	bird	boat	bott.	bus	car	cat	chair	cow	table	dog	horse	motor	pers.	plant	sheep	sofa	train	tv	mIoU
Baseline	89.9	73.6	33.8	75.1	42.0	54.4	80.0	75.8	78.9	24.7	50.2	43.1	72.6	50.2	68.2	77.2	34.9	64.8	30.6	67.6	55.1	59.2
CReST [64]	90.5	77.0	**38.6**	74.8	48.2	52.1	83.3	76.0	82.9	24.9	61.2	49.8	79.6	63.7	71.2	77.3	41.5	65.9	34.8	74.7	59.1	63.2
DARS [20]	91.3	82.6	37.4	81.9	50.5	58.6	**88.5**	**82.9**	82.8	25.5	56.3	49.1	75.3	64.6	73.6	79.7	**42.2**	64.0	37.1	73.4	57.9	64.5
USRN (Ours)	**91.9**	**84.1**	36.1	**84.9**	**52.8**	**66.4**	87.9	81.8	**86.4**	**26.5**	**75.2**	**58.6**	**83.0**	**73.3**	**74.7**	**80.2**	40.7	**76.2**	**42.0**	**78.5**	**59.8**	**68.6**

Table 7. Quantitative comparisons of USRN with multiple class-imbalance learning methods for semi-supervised semantic segmentation. The experiments are conducted over 1/32 split of PASCAL VOC dataset.

formance of USRN degrades because the predicted pseudo labels tend to become noisy. While γ is larger than 0.95, USRN suffers from over-fitting because the very high confidence threshold returns very limited pseudo labels. We set γ at 0.75 by default in our implemented USRN.

4.4. Discussion

Comparison with Class-Imbalance Methods: The proposed USRN explores class-unbiased segmentation to address the class imbalance issue in semi-supervised segmentation. Recently, several studies [20, 64] attempt to handle the class imbalance issue in semi-supervised learning. We compare USRN with these methods and Table 7 shows experimental results. It can been seen that USRN achieves the best overall performance (*i.e.*, 68.6 in mIoU) and the best per-class accuracy on 17 out of all 21 classes. The superior performance shows that exploring class-unbiased segmentation from balanced subclass distributions is more effective than selecting more pseudo labels for minority classes in self-training as in [20, 64].

Method	Base	+ USRN	Gain
DARS [20]	64.5	69.0	+4.5
CPS [10]	64.8	69.2	+4.4
CAC [35]	65.1	70.0	+4.9

Table 8. The proposed USRN complements with state-of-the-art methods [10, 20, 35] over 1/32 split of PASCAL VOC dataset: the performance of all tested state-of-the-art methods can be improved greatly with the integration of USRN.

Complementary Studies: We also investigate whether the proposed USRN can complement with state-of-the-art methods [10, 20, 35] as compared in Section 4.2. We integrate our proposed unbiased subclass regularization networks into the state-of-the-art methods to perform this study. Table 8 shows experimental results. It can be observed that the integration of USRN improves performance greatly across all tested state-of-the-art methods which employ either consistency-training [35] or self-training [10, 20].

Different Segmentation Architectures: We further study whether USRN can work well with different semantic segmentation architectures. We studied three widely used segmentation architectures including PSPNet [69], PSANet [70] and Deeplabv3+ [9] and Table 9 shows experimental results. It can be observed that the proposed USRN outperforms the *Baseline* model with large margins consistently with the three architectures. This shows that USRN can work well with different semantic segmentation architectures that apply pyramid spatial pooling [69], attention mechanism [70] and dilated convolutions [9].

Architecture	Baseline	USRN	Gain
PSPNet [69]	49.7	65.4	+15.7
PSANet [70]	56.5	66.5	+10.0
Deeplabv3+ [9]	59.2	68.6	+9.4

Table 9. The proposed USRN can work well with different semantic segmentation architectures [9, 69, 70] with significant performance improvement as compared with the *Baseline* over 1/32 split of PASCAL VOC dataset.

5. Conclusion

This paper presents an unbiased subclass regularization network that explores class-unbiased segmentation to address class imbalance issue in semi-supervised segmentation. Specifically, the class-biased segmentation learnt in imbalanced original class distributions is regularized by the class-unbiased segmentation learnt in balanced subclass distributions. To coordinate the concurrent learning from the original class and the generated subclass, an entropy-based gate mechanism is designed to suppress unconfident subclass predictions for facilitating subclass regularization. Comprehensive experiments demonstrate the effectiveness of our method in semi-supervised segmentation. In the future, we will investigate how the idea of unbiased subclass regularization perform in other semi-supervised learning tasks such as semi-supervised image classification and semi-supervised object detection.

Acknowledgement. This study is supported under the RIE2020 Industry Alignment Fund – Industry Collaboration Projects (IAF-ICP) Funding Initiative, as well as cash and in-kind contribution from Singapore Telecommunications Limited (Singtel), through Singtel Cognitive and Artificial Intelligence Lab for Enterprises (SCALE@NTU).

References

[1] Iñigo Alonso, Alberto Sabater, David Ferstl, Luis Montesano, and Ana C. Murillo. Semi-supervised semantic segmentation with pixel-level contrastive learning from a class-wise memory bank. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8219–8228, October 2021. 2

[2] Nikita Araslanov and Stefan Roth. Self-supervised augmentation consistency for adapting semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15384–15394, 2021. 2

[3] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017. 2, 7

[4] David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin A Raffel. Mixmatch: A holistic approach to semi-supervised learning. In *Advances in Neural Information Processing Systems*, pages 5049–5059, 2019. 2

[5] Mateusz Buda, Atsuto Maki, and Maciej A Mazurowski. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 106:249–259, 2018. 1, 2

[6] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority oversampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002. 1, 2

[7] 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 transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017. 2

[8] Liang-Chieh Chen, Yi Yang, Jiang Wang, Wei Xu, and Alan L Yuille. Attention to scale: Scale-aware semantic image segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3640–3649, 2016. 2

[9] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018. 1, 2, 5, 7, 8

[10] Xiaokang Chen, Yuhui Yuan, Gang Zeng, and Jingdong Wang. Semi-supervised semantic segmentation with cross pseudo supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2613–2622, 2021. 1, 2, 3, 5, 6, 8

[11] 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 *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016. 1, 5

[12] Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie. Class-balanced loss based on effective number of samples. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9268–9277, 2019. 2

[13] Jifeng Dai, Kaiming He, and Jian Sun. Convolutional feature masking for joint object and stuff segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3992–4000, 2015. 2

[14] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 5

[15] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010. 1, 5

[16] Jun Fu, Jing Liu, Haijie Tian, Yong Li, Yongjun Bao, Zhiwei Fang, and Hanqing Lu. Dual attention network for scene segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3146–3154, 2019. 2

[17] Dayan Guan, Jiaxing Huang, Shijian Lu, and Aoran Xiao. Scale variance minimization for unsupervised domain adaptation in image segmentation. *Pattern Recognition*, 112:107764, 2021. 2

[18] Dayan Guan, Jiaxing Huang, Aoran Xiao, and Shijian Lu. Domain adaptive video segmentation via temporal consistency regularization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8053–8064, 2021. 2

[19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1, 5

[20] Ruifei He, Jihan Yang, and Xiaojuan Qi. Re-distributing biased pseudo labels for semi-supervised semantic segmentation: A baseline investigation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6930–6940, 2021. 1, 2, 3, 5, 8

[21] Hanzhe Hu, Fangyun Wei, Han Hu, Qiwei Ye, Jinshi Cui, and Liwei Wang. Semi-supervised semantic segmentation via adaptive equalization learning. *Advances in Neural Information Processing Systems*, 34, 2021. 2

[22] Chen Huang, Yining Li, Chen Change Loy, and Xiaoou Tang. Learning deep representation for imbalanced classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5375–5384, 2016. 2

[23] Jiaxing Huang, Dayan Guan, Aoran Xiao, and Shijian Lu. Cross-view regularization for domain adaptive panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10133–10144, 2021. 2

[24] Jiaxing Huang, Dayan Guan, Aoran Xiao, and Shijian Lu. Fsdr: Frequency space domain randomization for domain generalization. *arXiv preprint arXiv:2103.02370*, 2021. 2

[25] Jiaxing Huang, Dayan Guan, Aoran Xiao, and Shijian Lu. Model adaptation: Historical contrastive learning for unsupervised domain adaptation without source data. *Advances in Neural Information Processing Systems*, 34, 2021. 2

[26] Jiaxing Huang, Dayan Guan, Aoran Xiao, and Shijian Lu. Rda: Robust domain adaptation via fourier adversarial attacking. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8988–8999, 2021. 2

[27] Jiaxing Huang, Dayan Guan, Aoran Xiao, and Shijian Lu. Multi-level adversarial network for domain adaptive semantic segmentation. *Pattern Recognition*, 123:108384, 2022. 2

[28] Jiaxing Huang, Shijian Lu, Dayan Guan, and Xiaobing Zhang. Contextual-relation consistent domain adaptation for semantic segmentation. In *European Conference on Computer Vision*, pages 705–722. Springer, 2020. 2

[29] Xinyue Huo, Lingxi Xie, Jianzhong He, Zijie Yang, Wengang Zhou, Houqiang Li, and Qi Tian. Atso: Asynchronous teacher-student optimization for semi-supervised image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1235–1244, 2021. 1, 2, 3

[30] Mostafa S Ibrahim, Arash Vahdat, Mani Ranjbar, and William G Macready. Semi-supervised semantic image segmentation with self-correcting networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12715–12725, 2020. 2

[31] Zhanghan Ke, Di Qiu, Kaican Li, Qiong Yan, and Rynson WH Lau. Guided collaborative training for pixel-wise semi-supervised learning. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16*, pages 429–445. Springer, 2020. 2, 3, 5

[32] Zhanghan Ke, Daoye Wang, Qiong Yan, Jimmy Ren, and Rynson WH Lau. Dual student: Breaking the limits of the teacher in semi-supervised learning. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 6728–6736, 2019. 2

[33] Jaehyung Kim, Youngbum Hur, Sejun Park, Eunho Yang, Sung Ju Hwang, and Jinwoo Shin. Distribution aligning refinery of pseudo-label for imbalanced semi-supervised learning. *Advances in Neural Information Processing Systems*, 33, 2020. 2

[34] Jaehyung Kim, Jongheon Jeong, and Jinwoo Shin. M2m: Imbalanced classification via major-to-minor translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13896–13905, 2020. 1, 2

[35] Xin Lai, Zhuotao Tian, Li Jiang, Shu Liu, Hengshuang Zhao, Liwei Wang, and Jiaya Jia. Semi-supervised semantic segmentation with directional context-aware consistency. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1205–1214, 2021. 1, 2, 3, 5, 6, 8

[36] Junnan Li, Caiming Xiong, and Steven CH Hoi. Comatch: Semi-supervised learning with contrastive graph regularization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9475–9484, 2021. 2

[37] Guosheng Lin, Chunhua Shen, Anton Van Den Hengel, and Ian Reid. Efficient piecewise training of deep structured models for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3194–3203, 2016. 2

[38] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017. 2

[39] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014. 1

[40] Weibo Lin, Zhu He, and Mingyu Xiao. Balanced clustering: A uniform model and fast algorithm. In *IJCAI*, pages 2987–2993, 2019. 4, 7

[41] Wei Liu, Andrew Rabinovich, and Alexander C Berg. Parsenet: Looking wider to see better. *arXiv preprint arXiv:1506.04579*, 2015. 2

[42] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015. 2

[43] Wenfeng Luo and Meng Yang. Semi-supervised semantic segmentation via strong-weak dual-branch network. In *European Conference on Computer Vision*, pages 784–800. Springer, 2020. 1, 3

[44] J MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, volume 5, pages 281–298. University of California Press, 1967. 7

[45] Robert Mendel, Luis Antonio De Souza, David Rauber, João Paulo Papa, and Christoph Palm. Semi-supervised segmentation based on error-correcting supervision. In *European Conference on Computer Vision*, pages 141–157. Springer, 2020. 1, 3

[46] Sudhanshu Mittal, Maxim Tatarchenko, and Thomas Brox. Semi-supervised semantic segmentation with high-and low-level consistency. *IEEE transactions on pattern analysis and machine intelligence*, 2019. 1

[47] Yassine Ouali, Céline Hudelot, and Myriam Tami. Semi-supervised semantic segmentation with cross-consistency training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12674–12684, 2020. 1, 2, 3, 5

[48] Mengshi Qi, Yunhong Wang, Jie Qin, and Annan Li. Kegan: Knowledge embedded generative adversarial networks for semi-supervised scene parsing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5237–5246, 2019. 2

[49] Mengye Ren, Wenyuan Zeng, Bin Yang, and Raquel Urtasun. Learning to reweight examples for robust deep learning. In *International Conference on Machine Learning*, pages 4334–4343. PMLR, 2018. 2

[50] Zhongzheng Ren, Raymond Yeh, and Alexander Schwing. Not all unlabeled data are equal: Learning to weight data in semi-supervised learning. *Advances in Neural Information Processing Systems*, 33, 2020. 2

[51] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 2

[52] Claude Elwood Shannon. A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423, 1948. 4

[53] 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. *Advances in Neural Information Processing Systems*, 32:1919–1930, 2019. 2

[54] Kihyuk Sohn, David Berthelot, Chun-Liang Li, Zizhao Zhang, Nicholas Carlini, Ekin D Cubuk, Alex Kurakin, Han Zhang, and Colin Raffel. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *arXiv preprint arXiv:2001.07685*, 2020. 2, 6

[55] Nimit Sohoni, Jared Dunnmon, Geoffrey Angus, Albert Gu, and Christopher Ré. No subclass left behind: Fine-grained robustness in coarse-grained classification problems. *Advances in Neural Information Processing Systems*, 33:19339–19352, 2020. 1

[56] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014. 6

[57] Fariborz Taherkhani, Ali Dabouei, Sobhan Soleymani, Jeremy Dawson, and Nasser M Nasrabadi. Self-supervised wasserstein pseudo-labeling for semi-supervised image classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12267–12277, 2021. 2

[58] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *Advances in neural information processing systems*, pages 1195–1204, 2017. 2

[59] Jesper E Van Engelen and Holger H Hoos. A survey on semi-supervised learning. *Machine Learning*, 109(2):373–440, 2020. 3

[60] Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, and Patrick Pérez. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2517–2526, 2019. 2

[61] Panqu Wang, Pengfei Chen, Ye Yuan, Ding Liu, Zehua Huang, Xiaodi Hou, and Garrison Cottrell. Understanding convolution for semantic segmentation. In *2018 IEEE winter conference on applications of computer vision (WACV)*, pages 1451–1460. IEEE, 2018. 2

[62] Yiru Wang, Weihao Gan, Jie Yang, Wei Wu, and Junjie Yan. Dynamic curriculum learning for imbalanced data classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5017–5026, 2019. 1, 2

[63] Yu-Xiong Wang, Deva Ramanan, and Martial Hebert. Learning to model the tail. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 7032–7042, 2017. 2

[64] Chen Wei, Kihyuk Sohn, Clayton Mellina, Alan Yuille, and Fan Yang. Crest: A class-rebalancing self-training framework for imbalanced semi-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10857–10866, 2021. 1, 2, 8

[65] Qizhe Xie, Zihang Dai, Eduard Hovy, Thang Luong, and Quoc Le. Unsupervised data augmentation for consistency training. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 6256–6268. Curran Associates, Inc., 2020. 2

[66] Yanchao Yang and Stefano Soatto. Fda: Fourier domain adaptation for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4085–4095, 2020. 2

[67] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015. 2

[68] Jianlong Yuan, Yifan Liu, Chunhua Shen, Zhibin Wang, and Hao Li. A simple baseline for semi-supervised semantic segmentation with strong data augmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8229–8238, October 2021. 1, 2, 5

[69] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017. 1, 2, 7, 8

[70] Hengshuang Zhao, Yi Zhang, Shu Liu, Jianping Shi, Chen Change Loy, Dahua Lin, and Jiaya Jia. Psanet: Point-wise spatial attention network for scene parsing. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 267–283, 2018. 2, 7, 8

[71] Yuanyi Zhong, Bodi Yuan, Hong Wu, Zhiqiang Yuan, Jian Peng, and Yu-Xiong Wang. Pixel contrastive-consistent semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7273–7282, 2021. 2

[72] Yanning Zhou, Hang Xu, Wei Zhang, Bin Gao, and Pheng-Ann Heng. C3-semiseg: Contrastive semi-supervised segmentation via cross-set learning and dynamic class-balancing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7036–7045, 2021. 2

[73] Yuliang Zou, Zizhao Zhang, Han Zhang, Chun-Liang Li, Xiao Bian, Jia-Bin Huang, and Tomas Pfister. Pseudoseg: Designing pseudo labels for semantic segmentation. In *International Conference on Learning Representations*, 2021. 2, 3