Personalized Image Semantic Segmentation

Yu Zhang[1] Chang-Bin Zhang[1] Peng-Tao Jiang [1] Ming-Ming Cheng[1*] Feng Mao[2]
[1]TKLNDST, CS, Nankai University [2]Alibaba Group
zhangyuygss@gmail.com pt.jiang@mail.nankai.edu.cn cmm@nankai.edu.cn

Abstract

Semantic segmentation models trained on public datasets have achieved great success in recent years. However, these models didn't consider the personalization issue of segmentation though it is important in practice. In this paper, we address the problem of personalized image segmentation. The objective is to generate more accurate segmentation results on unlabeled personalized images by investigating the data's personalized traits. To open up future research in this area, we collect a large dataset containing various users' personalized images called PSS (Personalized Semantic Segmentation). We also survey some recent researches related to this problem and report their performance on our dataset. Furthermore, by observing the correlation among a user's personalized images, we propose a baseline method that incorporates the inter-image context when segmenting certain images. Extensive experiments show that our method outperforms the existing methods on the proposed dataset. The code and the PSS dataset are available at https://mmcheng.net/pss/.

1. Introduction

Semantic segmentation is a well-studied task in the computer vision society. The goal of this task is to assign a semantic label to each pixel of a given image. As with other computer vision tasks, deep learning has greatly empowered semantic segmentation [4,5,30,33,37,57,58] with its great representation learning ability. These state-of-the-art methods mainly focus on the publicly available datasets like Pascal VOC [10], ADE20K [61], CityScapes [7], in which images are assumed to be independent and identically distributed. However, this assumption does not stand in real-world scenarios. For example, in mobile photography, a user may take pictures to record his/her own life and form a personalized image set. On the one hand, the personalized data does not have an identical distribution with public datasets, leading to a generalization issue when

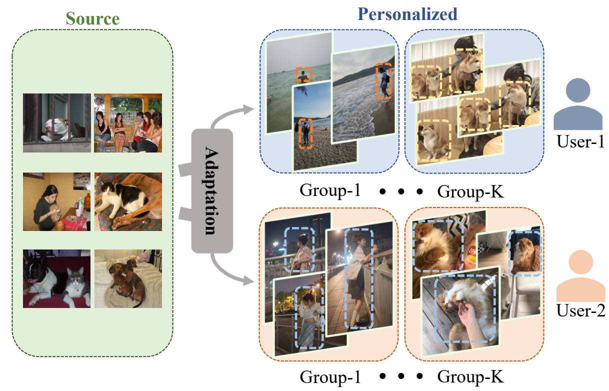

Figure 1. Samples of the proposed personalized image dataset. We investigate how to utilize the personalized property of users' images when adapting from source data. One notable property we can observe is that images from the same user are correlated (similar objects, scenes, etc.).

adopting well-trained segmentation models trained on public datasets. On the other hand, as shown in Fig. 1, images from the same user are correlated. It yields potential studies on utilizing this interrelated property to facilitate segmentation.

This paper addresses *personalized image segmentation*, a problem that has not been discussed in previous works. The difficulties mainly lie in the following two aspects: (i) Firstly, there exists a large distribution gap between the public datasets and a user's personal data. A simple method is to use extra annotations of the user's data to train the model, which is very costly. So it is urged to learn directly from unlabeled personalized data. However, there is no available personalized dataset to learn. (ii) Additionally, the personalized images from the same user usually have some personal traits. How to properly utilize these personalized traits in semantic segmentation remains an unaddressed problem. Despite the above difficulties, there are great demands for the personalized image segmentation in practice. For example, camera apps may need to generate high-quality segmentation masks for user's images to assist photography. To tackle these challenges and open up further studies on personalized image segmentation, we propose a personalized

*MM Cheng is the corresponding author.

dataset called PSS and a baseline personalized segmentation method based on the proposed dataset.

The PSS dataset contains 15 individual users' personalized images, resulting in 10080 images in total. For each user's personalized data, we randomly select around 30% of images and annotate their pixel-level semantic masks for validation. For easier adaptation from existing datasets to our personalized data, we consider the 20 common object classes as in the PASCAL VOC [10] dataset. To our knowledge, the PSS dataset is the first dataset that focuses on the personalization issue of image segmentation. It enables researchers to study segmentation problems by utilizing personalized traits.

On the challenge of learning with personal traits, previous studies on vision tasks' personalization issues [16, 22, 40] usually extract a global memory from personal data to represent the user's preference or personality. The extracted memory is then adopted as a prior for specific downstream tasks. However, we consider personalization from a broader perspective, *i.e*., the personal traits lie in each person's images, which is unnecessary to be extracted as a global representation. In fact, experiments indicate that a global representation extracted for a user fails in our scenario. The failure is reasonable since in the semantic segmentation problem, we need to predict the class of each pixel for an image. Though certain users' images have their characteristics, there are still various classes of objects and scenes among these images. A global representation for all pixels would be too ambiguous. To avoid the ambiguity of global representation while learning to segment with personalized traits. We propose to investigate the contextual connection between personalized images and utilize them among correlated images locally. Specifically, we first cluster a person's images into several groups so that images from the same group share similar objects or backgrounds. Then within each group, we extract multiple local region representations. For the prediction of each pixel in that group, we consult with correlated regional representations using an attention mechanism.

Note that no labeled training images are provided in the proposed personalized dataset. We tackle this problem by domain adaptation from the existing labeled dataset (as source) to the personalized images (as target). There are many works on the problem of unsupervised domain adaptation semantic segmentation (UDASS) [3, 9, 23, 25, 28, 34, 35, 50]. While our personalized images are correlated with each other, current UDASS methods all treat target images as independently distributed. They can't capture the personality traits in the personal data. In our baseline approach, we incorporate a group context module into the domain adaptation framework. It allows the network to adapt from existing dataset to the personalized images while taking advantage of the personal traits in the personalized images.

The contributions of this work are two folds:

- We first propose the personalization issue of image semantic segmentation and collect a personalized image dataset called PSS, containing 15 different users' data.
- We select some recent works related to this problem and report their performances on our dataset. Furthermore, we propose a baseline method that studies personal traits by learning local region representations. The approach achieves state-of-the-art performance on the proposed personalized dataset.

2. Related Work

2.1. Personalization Research

Personalization issue has been discussed in many computer vision and NLP tasks. [36] use personality traits to enhance the machine translation system. [16] proposed a personalized classifier for food image classification. [40] predict caption and hashtag for social media images by exploring personality from the user's previous posts. [22] studied image enhancement based on the user's preference. These methods usually focus on learning a global representation from existing data, which serves as a prior when facing new data. In this paper, we explore personalized image semantic segmentation, a problem that has not been discussed previously. In our proposed problem, personalization can be investigated from the user's global trait, and the personalized images' correlated property.

2.2. Learning from Correlated Data

A key challenge of the proposed personalized image segmentation is to learn from the correlated images of the same person, *i.e*., to extract complementary semantics while discarding misleading semantics. Co-segmentation [12, 18, 26, 62] and Co-saliency detection [11, 56] aim to mine the common semantic objects in grouped images, where each group contains the same class of objects. Li *et al*. [26] proposed a recurrent network architecture to explore the common semantic representation. Zhang *et al*. [56] utilize the common classification features to discover the general area of the target. These methods usually learn group representation for each group, which serves as a prior when segmenting images from that group. Our situation is more complex since the personalized data may contain different objects in different images, and we need to segment all these classes instead of just one.

2.3. Domain Adaptation for Semantic Segmentation

The objective of personalized image segmentation is to predict segmentation masks for unlabeled personalized images with existing datasets and models. A similar task

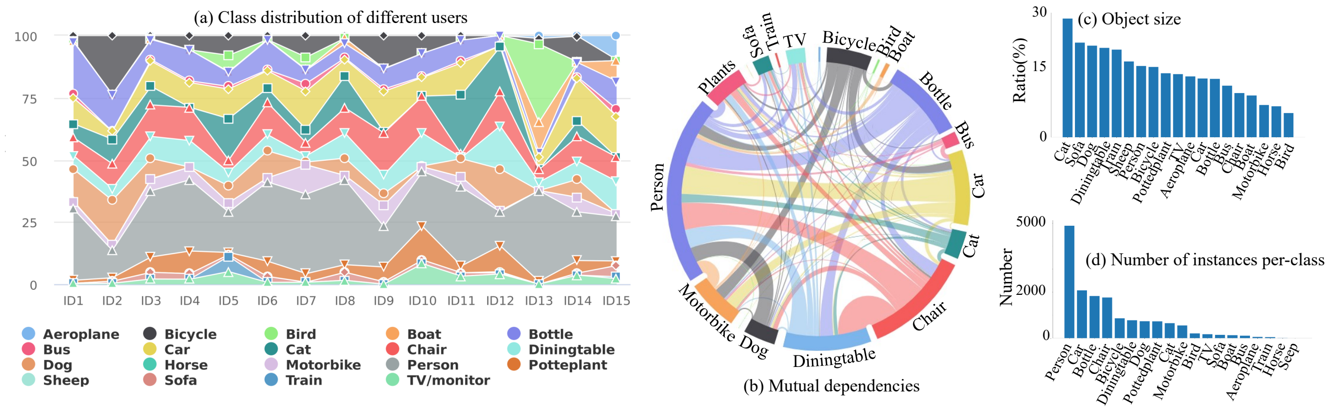

Figure 2. Statistics of the PSS dataset. (a) The proportion of different object classes of each user's personalized data. (b) The mutual dependencies of different classes over the dataset. (c) The average size of different classes. (d) The number of instances per-class.

that has been well studied recently is unsupervised domain adaptation for semantic segmentation. We will call it UDASS in the rest of the paper. Given a labeled source dataset and unlabeled target dataset, UDASS aims at tackling the distribution mismatch between the source dataset and the target dataset and make the model generalize well from source to target. One line of work [6, 15, 24, 38, 45–47, 51] for UDASS use adversarial-based methods to align the distribution shift of source and target domains. Another line of work [19, 21, 27, 29, 48, 54, 55, 63] focus on learning strategy: use curriculum learning or self-training strategies to babysit the network towards learning good semantics for the target domain. The main difference between our problem and UDASS is that we do not consider images of the target domain independently. Instead, we consider images from the same person as correlated parts. Similar images may provide useful information for others. Besides, current UDASS methods mainly focus on urban scene datasets like Cityscapes [7], GTA5 [42], and SYNTHIA [43], where they do domain adaptation between synthetic images and real images. In the proposed dataset, we focus on personalized images of common objects. Our dataset provides a more diverse and realistic personalized scenario that can also evaluate the effectiveness of domain adaptation approaches.

2.4. Datasets for Semantic Segmentation

As with other computer vision tasks, datasets play a key role in the research of image segmentation. Recent datasets have greatly empowered deep learning-based segmentation frameworks [4, 13, 57]. PASCAL VOC [10] and COCO [31] are datasets focusing on images of common objects. ADE20K [61] also focuses on common objects but with more fine-grained class labels like object parts. CityScapes [7], GTA5 [42], SYNTHIA [43], and Synscapes [49] are datasets of urban scenes. Though many datasets have been proposed recently, none of these datasets consider the personalization issue in segmentation. In this paper, we collect the PSS dataset from various users. Our dataset concentrates on images of common objects with different users' personalized traits. It can be a good start for researches in personalized image segmentation. It can also provide a good benchmark for other segmentation tasks like domain adaptation.

3. Proposed Dataset

3.1. Dataset Collection

To mimic real-world personalized data distribution, we directly collect our dataset from different volunteers. Each volunteer is asked to export images in his/her mobile phone or camera to form his/her personal data. For privacy, the volunteers are asked to look through the images and filter out images that he/she is not willing to make public. Our dataset focuses on the 20 classes as in PASCAL VOC [10]. Eventually, we get a large-scale dataset with 10080 images composed of 15 users' personalized data. Each personalized data can have different data distribution than other users and may have some of its high-/low-level statistics useful for semantic segmentation.

3.2. Data Annotation

We asked several well-trained experts to annotated the collected personalized data. Both image-level and pixel-level annotations are provided in our dataset.

Image-Level Annotation. Consistency with [10], all images in our dataset are labeled with the class label of objects that appeared. On the one hand, the image-level labels can be used for data analysis. On the other hand, the image-level labels may benefit the personalized semantic segmentation due to their success in many weakly supervised segmentation approaches [1, 17, 20, 32, 39]. We show the object class distribution for each user and the mutual dependencies of different classes in Fig. 2(a) and Fig. 2(b).

Pixel-Level Annotation. The challenge of personalized image segmentation is to generate segmentation masks on unlabeled personalized images. For model evaluation on our dataset, we provide pixel-wise annotation for around 30% of each user's data. For each pixel-wise annotated image, object regions belong to the 20 classes. As in PASCAL VOC [10], they are labeled with certain values, resulting in a pixel-wise mask indicating the class of every pixel in the image. We show the average size and the number of instances for different classes in Fig. 2(c) and Fig. 2(d).

3.3. Dataset Characteristics

Personalized Data. The most important characteristic of our dataset is personalization. This naturally leads to intra-user coherency, *i.e.*, certain user's data has its trait that might be coherent among different images, which can be utilized to facilitate learning. On the other hand, different user's images differ in both low-level (*e.g.*, light condition, picture quality) and high-level (*e.g.*, image contents, background) properties. The diversity of data distribution between different users requires segmentation models to adjust to certain user's data. More details of the intra-user coherency and the inter-user distribution gap can be found in the supplementary material.

Realistic Data. Our personalized dataset is very close to realistic scenarios. The realism lies in two folds. Firstly, our dataset is directly collected from different users. These images faithfully reflect what they care about and take pictures in daily life, which means our dataset's results can reflect the effectiveness of different practice methods. As the examples shown in the supplementary material: some users have more images of foods or pets about their daily life, while others have more images of beautiful scenery. This indicates the importance of personalized segmentation. Secondly, the object classes in our dataset are long-tailed distributed, as illustrated in Fig. 2(d). Some objects are more likely to be filmed while others are not, *e.g.*, there might be *"person"* in most of the images, while only a few instances of *"boat."* How to settle the unbalanced class distribution problem can be an interesting direction to explore.

4. Proposed Approach

In this section, we introduce our proposed baseline method for personalized image segmentation.

Overview. Consider source data with images $\{I_s \subset \mathbb{R}^{3\times H\times W}\}$ and its C-class segmentation labels $\{L_s \subset \mathbb{R}^{C\times H\times W}\}$, the unlabeled personalized data $\{I_p \subset \mathbb{R}^{3\times H\times W}\}$. Our approach's key idea is to utilize the correlation between personalized images $\{I_p\}$ by using context from other images of the same user. We show the architecture of our approach in Fig. 3. Our personalized image segmentation framework has two major steps: a domain adaptation step and a pseudo label refinement step. In the first step, we adapt from source data to personalized data with an adversarial based domain adaptation framework. During training, we incorporate our proposed group region context module to utilize the inter-image context in personalized data. In the second step, we select easy images in the personalized data as pseudo labels with entropy maps. The pseudo labels are used as ground-truth of the easy images to guide the segmentation network.

4.1. Adversarial Based Domain Adaptation

We start by introducing the adversarial based domain adaptation technique we used in step one. Denoting a segmentation network as S, it takes image I_s as input and outputs a soft prediction map $P_s = S(I_s) \in \mathbb{R}^{C\times H\times W}$, where each value $P_s^{(c,h,w)}$ indicates the probability that pixel $I_s^{(h,w)}$ belongs to class c. Given I_s's ground-truth Y_s, a cross-entropy loss:

$$\mathcal{L}_{seg} = -\sum_{h,w}\sum_{c} Y_s^{c,h,w} log(P_s^{(c,h,w)}) \tag{1}$$

is optimized to train the segmentation network. Besides the segmentation loss, an adversarial training paradigm is adopted to align the distribution discrepancy between source data $\{I_s\}$ and personalized data $\{I_p\}$. Given the source and personalized image's segmentation prediction P_s and P_p, we compute their entropy map by

$$E_s^{h,w} = \sum_{c} -P_s^{c,h,w} log(P_s^{c,h,w}). \tag{2}$$

A discriminator D is trained to predict the domain labels of E_s and E_p. By training the segmentation network S to fool D, we can close the distribution gap between the predictions of source and personalized data. The adversarial loss is formulated as:

$$\mathcal{L}_{adv}(I_s, I_p) = -\sum_{h,w} log(1 - D(E_s^{h,w})) + log(D(E_p^{h,w})). \tag{3}$$

This adversarial paradigm can align the distribution mismatch between source data and our personalized data. However, it takes each image in personalized data individually, thus fails to consider the correlation within $\{I_p\}$. For this purpose, we propose a group region context module to utilize the inter-image context of the personalized data.

4.2. Group Context Module

We design a simple group context module to utilize the correlated property of the proposed personalized dataset. We first cluster each user's personalized data into multiple groups. Each group contains images with similar semantics. Within each group, we extract soft region-wise context representations for all the images. During segmentation, all

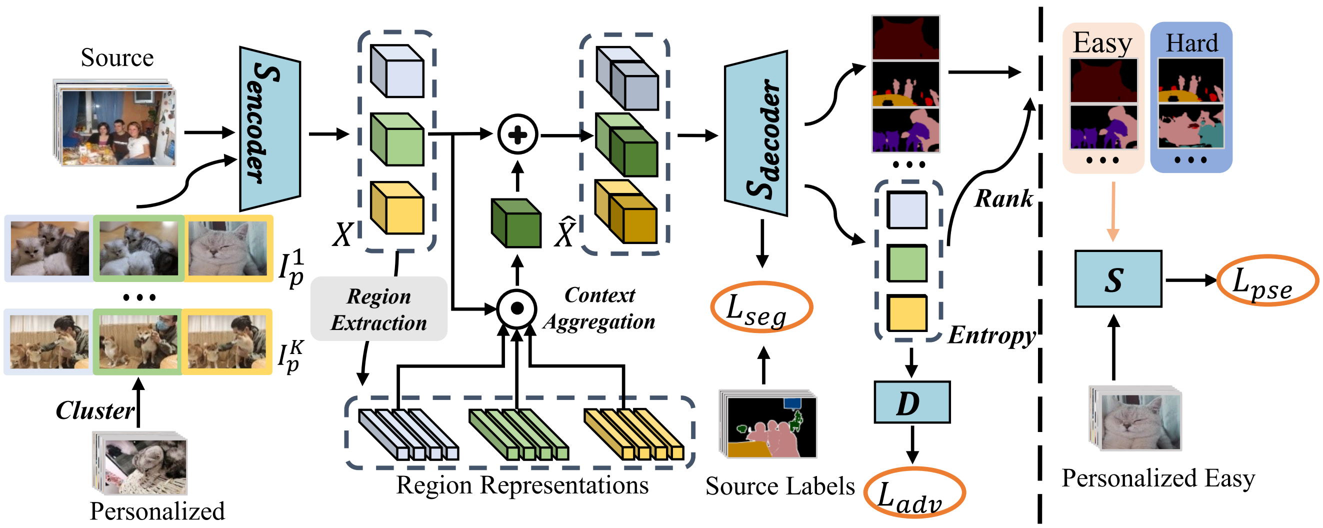

Figure 3. The pipeline of our approach for personalized image segmentation. Our model consists of two steps. The first step is the domain adaptation step, as in (a). In the second step, we further add a pseudo label loss L_{pse} as in (b). In (a), we first cluster the personalized data into K groups. Then in each group, we enhance the image representation X with regional group context to obtain $\hat{X}$. For simplicity, we only show three images for each group, and we only show the group region context aggregation process for the image tagged green.

soft region-wise context representations are inferred to help training.

For a user's personalized data $\{I_p\}$, we feed them into ResNet-50 [14] pretrained on ImageNet [8] and get the representation $\{F_p \in \mathbb{R}^{2048}\}$ before the last fully connected layer. Then we adopt the K-means clustering algorithm on $\{F_p\}$, obtaining K groups of images as $\{\{I_p^1\}, \{I_p^2\}, \cdots, \{I_p^K\}\}$. Consider the segmentation network as the composition of encoder $S_{encoder}$ and decoder $S_{decoder}$. The encoder takes image I_p from group k as input and outputs intermediate representation $X = S_{encoder}(I_p) \in \mathbb{R}^{CH\times W\times H}$, CH and W, H indicate the channel and spatial size of X, respectively. Our group context module F_{group} learns an enhanced representation $\hat{X} = F_{group}(X) \in \mathbb{R}^{CH\times W\times H}$ by utilizing the group context. There are two steps in the group context module: *region context extraction* and *group region context aggregation*.

Region context extraction. Inspired by [53], we partition image I_p into C soft object regions. C is the number of object classes. Using the aux output $P_p \in \mathbb{R}^{C\times W\times H}$ of the segmentation network. We compute each soft region's representation as

$$f_c = \sum_i r_{ci} X_i, \tag{4}$$

where i denotes the spatial location, X_i represents pixel i. r_{ci} is the weight of pixel i computed by softmax normalization of $P_{pi} \in \mathbb{R}^C$ over the C dimensions as $r_{ci} = softmax(P_{pi})_c$. For a group with N images, we can extract $N \times C$ region representations for this group.

Group region context aggregation. Given a group's region representations $\{f_{i,j} | i \in [1, C], j \in [1, N]\}$, we compute group context representation for each pixel in X by weighted aggregation of group regions:

$$c_{h,w} = \rho(\sum_{i,j} w_{(i,j),(h,w)} \sigma(f_{i,j})). \tag{5}$$

Here, ρ and σ are two linear transformation functions. The weight $w_{(\tilde{i},\tilde{j}),(h,w)}$ is computed by measuring the relation between the pixel $X_{h,w}$ and region representation $f_{\tilde{i},\tilde{j}}$ as

$$w_{(\tilde{i},\tilde{j}),(h,w)} = \frac{e^{s(X_{h,w}, f_{\tilde{i},\tilde{j}})}}{\sum_{i\in[1,K],j\in[1,N]} e^{s(X_{h,w}, f_{i,j})}}, \tag{6}$$

where $s(X_{h,w}, f_{i,j})$ is a relation function formulated as $s(X_{h,w}, f_{i,j}) = \phi(X_{h,w})^T \varphi(f_{i,j})$, ϕ and φ are two transform functions implemented with one fully connected layer.

After obtaining the group context, we can enhance the pixel representation as:

$$\hat{X}_{h,w} = \psi([X_{h,w}, c_{h,w}]). \tag{7}$$

$[*, *]$ denotes concatenation, and ψ is a linear transformation. The result representation $\hat{X}$ will be fed into the decoder and output the prediction map: $\hat{P}_p = S_{decoder}(\hat{X})$. For each pixel in X, the group region context enhancing module aggregates representations of similar regions in the same group as the group context, which provides extra information for the segmentation network.

Method	Backbone	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Mean
No-DA	ResNet-50	45.39	51.99	48.95	47.60	58.03	48.15	56.86	62.45	48.23	45.14	62.37	51.68	48.56	48.13	41.57	51.01
AdaptSeg [45]		46.87	52.16	50.06	48.51	59.78	51.39	57.12	63.41	50.99	46.15	60.68	52.84	50.32	50.69	43.08	52.27
MaxSquare [6]		48.28	52.50	50.61	50.54	61.39	54.60	59.36	63.43	50.67	46.49	62.94	52.68	49.65	48.99	46.00	53.20
FDA [52]		50.12	53.70	53.22	50.76	60.29	55.01	58.18	65.89	53.28	46.49	62.09	56.10	48.93	51.38	47.03	54.16
ADVENT [47]		53.39	57.33	52.42	52.51	64.63	55.04	60.61	61.69	55.34	49.18	66.05	57.83	56.04	54.38	52.34	56.59
MRNet [60]		**54.05**	58.62	54.29	53.17	61.72	57.24	62.20	66.46	56.75	50.27	66.76	54.20	53.87	54.38	51.38	57.02
OURS-S1		52.90	59.12	54.74	55.82	64.97	**60.38**	61.78	68.12	56.99	51.21	69.42	60.44	57.05	54.41	54.51	58.79
OURS-S2		53.28	**60.39**	**54.81**	**56.02**	**66.87**	60.11	**63.77**	**69.09**	**57.44**	**52.66**	**70.42**	**60.77**	**58.50**	**56.84**	**54.85**	**59.72**
No-DA	VGG-16	33.68	33.56	35.50	35.49	39.52	37.55	36.23	47.95	34.35	32.86	50.95	41.48	39.24	30.90	34.51	37.58
AdaptSeg [45]		32.70	37.65	37.16	33.54	40.55	41.11	43.17	52.12	36.95	31.83	49.04	40.97	33.54	31.49	34.06	38.39
MaxSquare [6]		36.17	32.99	38.81	37.36	42.64	42.03	49.88	50.06	37.99	35.93	51.33	41.98	36.27	36.35	37.13	40.46
FDA [52]		34.61	36.75	35.53	36.60	38.36	40.07	45.21	52.57	37.79	35.01	49.59	41.93	33.72	35.01	36.27	39.27
ADVENT [47]		39.89	44.39	39.88	40.01	49.89	44.24	47.99	54.59	43.84	38.29	53.00	43.07	42.83	40.02	41.36	44.22
MRNet [60]		34.40	41.18	36.67	32.18	44.63	38.12	41.99	46.78	39.51	36.54	39.39	44.17	35.93	37.17	38.35	39.13
OURS-S1		41.87	45.73	43.14	**44.04**	52.44	47.45	**52.32**	56.92	45.61	**42.67**	54.94	**48.38**	44.24	41.67	45.98	47.16
OURS-S2		**43.24**	**47.89**	**44.67**	44.00	**53.27**	**50.68**	52.18	**57.86**	**46.84**	42.34	**56.56**	46.28	**47.02**	**42.98**	**47.01**	**48.19**

Table 1. FIoU results for different methods with ResNet-50 [14] and VGG-16 [44]. The column number indicates the 15 user IDs. The column "Mean" denotes the mean performance overall IDs. Best results are highlighted in **bold**.

Method	Backbone	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Mean
No-DA	ResNet-50	28.05	29.18	30.78	33.05	42.52	31.31	35.85	28.63	39.60	36.99	33.15	38.51	29.78	32.75	31.85	33.47
AdaptSeg [45]		31.69	28.87	30.50	35.09	45.83	32.55	36.70	33.83	36.43	36.49	34.09	41.23	31.02	35.52	34.40	34.95
MaxSquare [6]		28.72	28.91	31.81	36.45	40.09	33.94	38.85	31.21	35.85	32.23	28.58	34.16	33.58	30.35	34.78	33.30
FDA [52]		31.94	31.16	32.39	36.11	45.35	35.76	37.46	30.93	42.91	**43.28**	**37.51**	38.09	29.31	**37.25**	35.76	36.35
ADVENT [47]		36.04	34.04	36.98	39.98	43.76	40.52	41.59	29.69	36.26	39.19	33.46	39.05	38.17	33.43	37.44	37.31
MRNet [60]		**38.27**	**35.02**	36.98	36.54	43.99	**40.90**	40.22	36.26	32.35	33.10	36.26	31.78	37.77	35.89	32.24	36.51
OURS-S1		36.61	34.43	31.88	40.36	44.25	33.64	38.14	32.25	39.87	38.69	37.20	42.44	**39.60**	30.37	**42.18**	37.46
OURS-S2		33.85	33.38	**38.40**	**41.36**	**46.73**	37.58	**44.19**	**36.87**	**44.66**	42.03	37.42	**43.71**	35.12	34.18	37.89	**39.16**
No-DA	VGG-16	15.78	17.19	17.80	21.41	21.54	18.35	19.07	15.40	21.67	22.80	18.55	21.06	21.66	18.96	22.95	19.61
AdaptSeg [45]		16.59	19.04	18.96	23.81	21.43	23.12	25.47	16.26	23.33	22.60	19.08	20.20	22.12	20.10	24.30	21.09
MaxSquare [6]		18.46	18.19	18.46	22.29	26.01	23.88	25.36	17.07	25.01	25.37	19.99	20.56	24.52	20.82	25.90	22.13
FDA [52]		17.17	17.82	20.07	24.44	23.82	25.69	25.22	15.56	23.31	25.53	21.14	20.68	20.82	19.26	24.65	21.68
ADVENT [47]		27.32	21.95	**25.21**	25.46	35.50	23.75	28.18	**25.03**	**32.47**	29.73	24.76	25.00	26.74	24.76	26.31	26.81
MRNet [60]		20.30	16.46	20.92	27.62	29.70	25.46	27.74	22.30	30.64	26.46	17.64	27.12	23.43	23.86	26.08	24.38
OURS-S1		24.40	**24.93**	20.31	31.01	35.54	30.67	28.81	20.68	27.45	**32.06**	27.63	**31.44**	27.68	23.87	**33.53**	28.00
OURS-S2		**25.53**	24.40	22.62	**33.55**	**35.73**	**31.86**	**32.14**	21.84	28.62	30.57	**30.26**	24.97	**28.82**	**25.25**	31.91	**28.54**

Table 2. MIoU results for different methods with ResNet-50 [14] and VGG-16 [44]. The column number indicates the 15 user IDs. The column "Mean" denotes the mean performance overall IDs. Best results are highlighted in **bold**.

4.3. Refine with Pseudo Label

Besides a first step domain adaptation, recent domain adaptation methods [38,59] for semantic segmentation usually adopt pseudo labels to refine the network further. We also adopt this training paradigm with the predictions of personalized images in our approach. The entropy map E_p introduced previously is an indicator of the uncertainty of the segmentation network for image I_p. Prediction with low uncertainty usually means the input image is simple, and the result has high reliability. So we choose predictions with low entropy values as pseudo labels. Note that each person's personalized data is relatively small compared with datasets like VOC and CityScapes, which alone doesn't have enough data for training the segmentation network. So unlike in [38], we add the pseudo labels to the network with an extra segmentation loss $\mathcal{L}_{pse}$ instead of replacing the source dataset with pseudo labels.

5. Experiments

5.1. Datasets and Evaluation Metrics

We collect our personalized dataset to have the same classes as PASCAL VOC [10]. So during training, we use the augmented VOC training set as the source dataset, which has 10582 labeled images with 20 classes of objects. Mean Intersection over Union (MIoU) is adopted for quantitative evaluation. Note that the personalized data are usually long-tailed distributed, which means the classes are very unbalanced. MIoU may be distorted due to this unbalance. So we further use another metric called Foreground Intersection over Union (FIoU). FIoU reflects the mean IoU over images instead of classes. Specifically, we first compute the foreground IoU IoU_i for image i, then compute the mean IoU of all images $\sum_i^N IoU_i$.

Methods	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Mean
None	39.89	44.39	39.88	40.01	49.89	44.24	47.99	54.59	43.84	38.29	53.00	43.07	42.83	40.02	41.36	44.22
Global	38.89	42.13	42.13	38.96	50.88	45.09	51.62	55.67	44.23	38.51	49.80	45.70	43.20	39.84	41.71	44.56
OURS	41.87	45.73	43.14	44.04	52.44	47.45	52.32	56.92	45.61	42.67	54.94	48.38	44.24	41.67	45.98	47.16

Table 3. Ablation of the group context module. "None" and "Global" denote no context and global context, respectively.

Groups	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Mean
1	42.39	46.11	42.95	43.79	52.45	46.15	51.23	55.88	45.57	42.51	55.30	47.89	43.66	39.54	44.12	46.64
10	42.49	45.52	42.75	44.07	52.64	46.54	50.85	56.52	45.81	42.25	54.75	46.86	44.62	41.18	45.36	46.81
80	41.87	45.73	43.14	44.04	52.44	47.45	52.32	56.92	45.61	42.67	54.94	48.38	44.24	41.67	45.98	47.16
200	42.11	45.66	43.14	43.33	52.08	45.85	52.05	56.51	45.78	42.13	53.75	46.73	43.80	41.59	45.35	46.66

Table 4. Ablation study of different numbers of groups. Columns indicate different user IDs. FIoU is reported.

Methods	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Mean
MixSample	40.92	43.18	40.73	40.55	49.84	46.18	50.42	57.22	42.92	39.39	54.26	45.83	43.67	38.74	44.90	45.25
MixAll	42.54	45.11	41.73	39.87	49.58	43.74	52.64	56.88	43.82	38.55	55.26	47.18	42.33	38.97	45.16	45.56
Personal	41.87	45.73	43.14	44.04	52.44	47.45	52.32	56.92	45.61	42.67	54.94	48.38	44.24	41.67	45.98	47.16

Table 5. Experiments on the mixed image set. "MixAll" denotes mix all the user's image, "MixSample" samples 1/15 from "MixAll" to have a similar size with each personalized data.

5.2. Implementation Details

We use ResNet50 [14] pretrained from ImageNet [8] as the backbone of the segmentation network. A PSP module [57] is equipped to the segmentation network as in [47]. The inputs for adaptation training are source images and labels, grouped target images. To simplify training and save computation, we do not use all image regions in a group to build the group context. Instead, we constrain each batch to be in the same group, then use the images in each batch to compute it. The random crop is adopted for image augmentation. All the inputs are resized to 320×320 during training. In the pseudo label refinement step, a select rate of $r = 0.5$ is used for choosing reliable predictions. The masked pseudo label pixels are set to 255. To simplify the training process and save GPU memory, we do not process a whole group at one iteration. Instead, we just make sure all images of a bath are from the same group. The batch size of all experiments is set to 8 in this paper. We use the SGD optimizer [2] with a learning rate of 2.5×10^{-4}, The momentum and weight decay are set to 0.9 and 10^{-4}, respectively. The codes are implemented with PyTorch [41] library.

5.3. Performace Comparison

We report the performances of some selected domain adaptation methods on our dataset, including AdaptSeg [45], MaxSquare [6], FDA [52], ADVENT [47], and MRNet [60]. All these methods treat target images individually without considering the correlated property of personalized images. All models are trained with VOC [10] as the source and personalized data as the target. Methods like MRNet [60] only use the target pseudo labels to supervise the segmentation network in step 2, which leads to poor performance since the number of our personalized data is relatively small. So we add extra supervision of VOC [10] label for such methods. The results are tested on the annotated validation split of the personalized dataset. We report the results of FIoU in Tab. 1 and MIoU in Tab. 2. We denote our method without pseudo label refinement and our full model as *OURS-S1* and *OURS-S2*, respectively.

Overall, with ResNet50 [14] as the backbone, *OURS-S1* obtains 37.46 MIoU and 58.79 FIoU. Compared to the baseline method ADVENT, it improves the performance by 0.15 and 2.20, respectively, which indicates the effectiveness of our group context module. Note that the MIoU improvement of 0.15 is relatively slight compared with FIoU. We conjecture it is caused by the long-tailed property of the personalized data. Since the group context module incorporates other images' context to help to learn, it tends to perform better on classes with many images while may damage the results for rare classes. When evaluating MIoU, the results can be affected by these rare classes. We provide the class IoU results for different users in the supplementary material. By utilizing pseudo labels, *OURS-S2* obtains 39.16 MIoU and 59.72 FIoU, further improves the performance by 1.7 and 0.93, respectively. We show some of the predicted masks in Fig. 4.

6. Discussion

6.1. Effectiveness of Group Context

In this section, we study the effectiveness of our proposed group context module by comparing it with two baselines: *None* and *Global*. *None* refers to directly use the feature X from the encoder without context. *Global* denotes

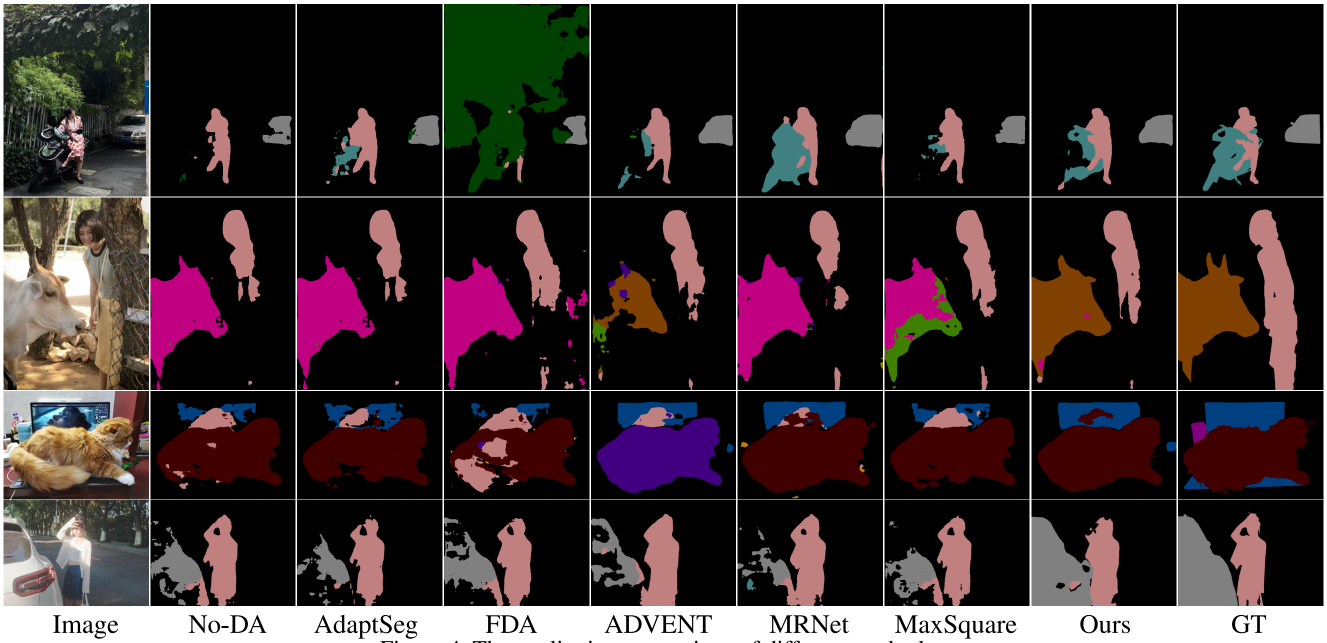

Figure 4. The qualitative comparison of different methods.

using a global group context to enhance representation as in object co-segmentation methods [26]. Experiments here are conducted with backbone VGG-16 [44]. As reported in Tab. 3, the *None* baseline achieves 44.22 FIoU on average. *Global* slightly improves the performance by 0.34, which indicates that a global group representation is not effective enough in our situation. *OURS* improves the performance by 2.94, which shows the effectiveness of the proposed group context module.

6.2. Significance of Personalized Training

In this section, we merge all the images from all the users to form a large image set *MixAll*, a subset of 672 images *MixSample* is then randomly sampled from *MixAll*. We train our model on these image sets and evaluate the model on different users' data. Results are reported in Tab. 5. With around 15 times of target images, the *MixAll* achieves 45.56 FIoU, which is lower than 47.16 of *Personal*, *i.e.*, training from corresponding personalized data. The result shows the value of learning from personalized data.

6.3. Number of Groups

In this section, we cluster each user's personalized data into different numbers of groups and investigate how the number of groups influences the segmentation performance at inference time. As in Tab. 4. Different rows indicate different numbers of groups. With *Groups=1*, all images of certain users are treated as in one group. When computing a group region context, irrelevant images might be considered and confusing the network. With *Groups=200*, the number of images in each group is too small, thus can't provide enough context for the group context module. On average, we get a better FIoU of 47.16 with a group number of 80 compared with other numbers. However, we may still notice that different users have the best result with different group numbers. We conjecture that it is caused by the distribution gap between different users. The results indicate that we need different numbers of groups for different users. In the future, we'll study more flexible methods to cluster the personalized images rather than using a fixed group number.

7. Conclusion

In this paper, we address the personalization issue in image semantic segmentation. We first collect a large personalized image dataset PSS with 15 users' data. Our dataset can be a good start for investigating the personalization issue in segmentation. The challenges of the personalized image segmentation problem are two folds. One is how to learn from different users' unlabeled data; another is how to utilize the personalized traits in certain user's data. By utilizing the personalized images' correlated property, we propose a baseline method that adopts the inter-image context to facilitate segmentation. For future work, we will explore more sophisticated ways to learn from the unlabeled data. We will also investigate how to improve the performance of the group context module in rare classes.

Acknowledgement. This research was supported by NSFC (61922046), S&T innovation project from Chinese Ministry of Education, BNRist(No.BNR2020KF01001) and the Fundamental Research Funds for the Central Universities (Nankai University, NO. 63213090).

References

[1] Jiwoon Ahn and Suha Kwak. Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 4981–4990, 2018. 3

[2] Léon Bottou. Large-scale machine learning with stochastic gradient descent. In *International Conference on Computational Statistics*, pages 177–186. Springer, 2010. 7

[3] Wei-Lun Chang, Hui-Po Wang, Wen-Hsiao Peng, and Wei-Chen Chiu. All about structure: Adapting structural information across domains for boosting semantic segmentation. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 1900–1909, 2019. 2

[4] 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. 1, 3

[5] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Eur. Conf. Comput. Vis.*, pages 801–818, 2018. 1

[6] Minghao Chen, Hongyang Xue, and Deng Cai. Domain adaptation for semantic segmentation with maximum squares loss. In *Int. Conf. Comput. Vis.*, pages 2090–2099, 2019. 3, 6, 7

[7] 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 *IEEE Conf. Comput. Vis. Pattern Recog.*, 2016. 1, 3

[8] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 248–255, 2009. 5, 7

[9] 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 *Int. Conf. Comput. Vis.*, pages 982–991, 2019. 2

[10] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *Int. J. Comput. Vis.*, 88(2):303–338, 2010. 1, 2, 3, 4, 6, 7

[11] Deng-Ping Fan, Zheng Lin, Ge-Peng Ji, Dingwen Zhang, Huazhu Fu, and Ming-Ming Cheng. Taking a deeper look at co-salient object detection. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2020. 2

[12] Junwei Han, Rong Quan, Dingwen Zhang, and Feiping Nie. Robust object co-segmentation using background prior. *IEEE Trans. Image Process.*, 27(4):1639–1651, 2017. 2

[13] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Int. Conf. Comput. Vis.*, pages 2961–2969, 2017. 3

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 770–778, 2016. 5, 6, 7

[15] Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei Efros, and Trevor Darrell. Cycada: Cycle-consistent adversarial domain adaptation. In *Int. Conf. Mach. Learn.*, pages 1989–1998, 2018. 3

[16] Shota Horiguchi, Sosuke Amano, Makoto Ogawa, and Kiyoharu Aizawa. Personalized classifier for food image recognition. *IEEE Trans. Multimedia*, 20(10):2836–2848, 2018. 2

[17] Qibin Hou, Peng-Tao Jiang, Yunchao Wei, and Ming-Ming Cheng. Self-erasing network for integral object attention. In *Adv. Neural Inform. Process. Syst.*, 2018. 3

[18] Kuang-Jui Hsu, Yen-Yu Lin, and Yung-Yu Chuang. Deepco3: Deep instance co-segmentation by co-peak search and co-saliency detection. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2019. 2

[19] Jiaxing Huang, Shijian Lu, Dayan Guan, and Xiaobing Zhang. Contextual-relation consistent domain adaptation for semantic segmentation. In *Eur. Conf. Comput. Vis.*, pages 705–722, 2020. 3

[20] Peng-Tao Jiang, Qibin Hou, Yang Cao, Ming-Ming Cheng, Yunchao Wei, and Hong-Kai Xiong. Integral object mining via online attention accumulation. In *Int. Conf. Comput. Vis.*, pages 2070–2079, 2019. 3

[21] Guoliang Kang, Yunchao Wei, Yi Yang, Yueting Zhuang, and Alexander G Hauptmann. Pixel-level cycle association: A new perspective for domain adaptive semantic segmentation. In *Adv. Neural Inform. Process. Syst.*, 2020. 3

[22] Han-Ul Kim, Young Jun Koh, and Chang-Su Kim. Pienet: Personalized image enhancement. In *Eur. Conf. Comput. Vis.*, 2020. 2

[23] Myeongjin Kim and Hyeran Byun. Learning texture invariant representation for domain adaptation of semantic segmentation. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 12975–12984, 2020. 2

[24] Minsu Kim, Sunghun Joung, Seungryong Kim, JungIn Park, Ig-Jae Kim, and Kwanghoon Sohn. Cross-domain grouping and alignment for domain adaptive semantic segmentation. In *AAAI Conf. Artif. Intell.*, 2021. 3

[25] Suhyeon Lee, Junhyuk Hyun, Hongje Seong, and Euntai Kim. Unsupervised domain adaptation for semantic segmentation by content transfer. In *AAAI Conf. Artif. Intell.*, 2021. 2

[26] Bo Li, Zhengxing Sun, Qian Li, Yunjie Wu, and Anqi Hu. Group-wise deep object co-segmentation with co-attention recurrent neural network. In *Int. Conf. Comput. Vis.*, pages 8519–8528, 2019. 2, 8

[27] Guangrui Li, Guoliang Kang, Wu Liu, Yunchao Wei, and Yi Yang. Content-consistent matching for domain adaptive semantic segmentation. In *Eur. Conf. Comput. Vis.*, pages 440–456, 2020. 3

[28] Yunsheng Li, Lu Yuan, and Nuno Vasconcelos. Bidirectional learning for domain adaptation of semantic segmentation. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2019. 2

[29] Qing Lian, Fengmao Lv, Lixin Duan, and Boqing Gong. Constructing self-motivated pyramid curriculums for cross-domain semantic segmentation: A non-adversarial approach. In *Int. Conf. Comput. Vis.*, pages 6758–6767, 2019. 3

[30] Guosheng Lin, Anton Milan, Chunhua Shen, and Ian Reid. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 1925–1934, 2017. 1

[31] Tsungyi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollar, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Eur. Conf. Comput. Vis.*, pages 740–755, 2014. 3

[32] Yun Liu, Yu-Huan Wu, Pei-Song Wen, Yu-Jun Shi, Yu Qiu, and Ming-Ming Cheng. Leveraging instance-, image- and dataset-level information for weakly supervised instance segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2020. 3

[33] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3431–3440, 2015. 1

[34] Yawei Luo, Ping Liu, Tao Guan, Junqing Yu, and Yi Yang. Significance-aware information bottleneck for domain adaptive semantic segmentation. In *Int. Conf. Comput. Vis.*, pages 6778–6787, 2019. 2

[35] 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 *IEEE Conf. Comput. Vis. Pattern Recog.*, 2019. 2

[36] Shachar Mirkin, Scott Nowson, Caroline Brun, and Julien Perez. Motivating personality-aware machine translation. In *Conf. Empir. Meth. Natur. Lang. Process.*, 2015. 2

[37] Hyeonwoo Noh, Seunghoon Hong, and Bohyung Han. Learning deconvolution network for semantic segmentation. In *Int. Conf. Comput. Vis.*, pages 1520–1528, 2015. 1

[38] Fei Pan, Inkyu Shin, Francois Rameau, Seokju Lee, and In So Kweon. Unsupervised intra-domain adaptation for semantic segmentation through self-supervision. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3764–3773, 2020. 3, 6

[39] George Papandreou, Liang-Chieh Chen, Kevin P Murphy, and Alan L Yuille. Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation. In *Int. Conf. Comput. Vis.*, pages 1742–1750, 2015. 3

[40] Cesc Chunseong Park, Byeongchang Kim, and Gunhee Kim. Attend to you: Personalized image captioning with context sequence memory networks. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 6432–6440, 2017. 2

[41] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Adv. Neural Inform. Process. Syst.*, pages 8024–8035, 2019. 7

[42] Stephan R Richter, Vibhav Vineet, Stefan Roth, and Vladlen Koltun. Playing for data: Ground truth from computer games. In *Eur. Conf. Comput. Vis.*, pages 102–118, 2016. 3

[43] 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 *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3234–3243, 2016. 3

[44] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *Int. Conf. Learn. Represent.*, 2015. 6, 8

[45] 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 *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 7472–7481, 2018. 3, 6, 7

[46] Yi-Hsuan Tsai, Kihyuk Sohn, Samuel Schulter, and Manmohan Chandraker. Domain adaptation for structured output via discriminative patch representations. In *Int. Conf. Comput. Vis.*, pages 1456–1465, 2019. 3

[47] Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, and Patrick Pérez. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2517–2526, 2019. 3, 6, 7

[48] 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 *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 12635–12644, 2020. 3

[49] Magnus Wrenninge and Jonas Unger. Synscapes: A photorealistic synthetic dataset for street scene parsing. *arXiv preprint arXiv:1810.08705*, 2018. 3

[50] Zuxuan Wu, Xintong Han, Yen-Liang Lin, Mustafa Gokhan Uzunbas, Tom Goldstein, Ser Nam Lim, and Larry S Davis. Dcan: Dual channel-wise alignment networks for unsupervised scene adaptation. In *Eur. Conf. Comput. Vis.*, 2018. 2

[51] Yanchao Yang, Dong Lao, Ganesh Sundaramoorthi, and Stefano Soatto. Phase consistent ecological domain adaptation. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2020. 3

[52] Yanchao Yang and Stefano Soatto. Fda: Fourier domain adaptation for semantic segmentation. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 4085–4095, 2020. 6, 7

[53] Yuhui Yuan, Xilin Chen, and Jingdong Wang. Object-contextual representations for semantic segmentation. *arXiv preprint arXiv:1909.11065*, 2019. 5

[54] Yang Zhang, Philip David, Hassan Foroosh, and Boqing Gong. A curriculum domain adaptation approach to the semantic segmentation of urban scenes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2019. 3

[55] Yang Zhang, Philip David, and Boqing Gong. Curriculum domain adaptation for semantic segmentation of urban scenes. In *Int. Conf. Comput. Vis.*, pages 2020–2030, 2017. 3

[56] Zhao Zhang, Wenda Jin, Jun Xu, and Ming-Ming Cheng. Gradient-induced co-saliency detection. In *Eur. Conf. Comput. Vis.*, 2020. 2

[57] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2017. 1, 3, 7

[58] Wei Zhen, Sun Yao, and J Lin. Learning adaptive receptive fields for deep image parsing networks. *Computational Visual Media*, 4(3):1–14, 2018. 1

[59] Zhedong Zheng and Yi Yang. Unsupervised scene adaptation with memory regularization in vivo. In *Int. Joint Conf. Artif. Intell.*, 2019. 6

[60] Zhedong Zheng and Yi Yang. Rectifying pseudo label learning via uncertainty estimation for domain adaptive semantic segmentation. *Int. J. Comput. Vis.*, 2020. 6, 7

[61] Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Scene parsing through ade20k dataset. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2017. 1, 3

[62] Chenyang Zhu, Kai Xu, Siddhartha Chaudhuri, Li Yi, Leonidas J. Guibas, and Hao Zhang. Adacoseg: Adaptive shape co-segmentation with group consistency loss. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2020. 2

[63] Yang Zou, Zhiding Yu, BVK Vijaya Kumar, and Jinsong Wang. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In *Eur. Conf. Comput. Vis.*, pages 289–305, 2018. 3