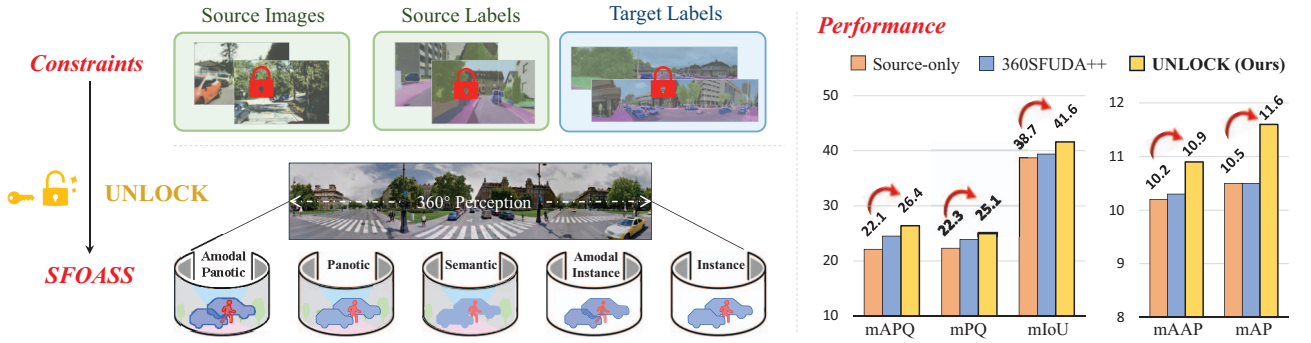
Unlocking Constraints: Source-Free Occlusion-Aware Seamless Segmentation

Yihong Cao[1,2,*] Jiaming Zhang[3,4,*] Xu Zheng[5,6] Hao Shi[7] Kunyu Peng[3] Hang Liu[1]
Kailun Yang[1,†] Hui Zhang[1,†]
[1]Hunan University [2]Hunan Normal University [3]Karlsruhe Institute of Technology [4]ETH Zürich
[5]HKUST(GZ) [6]INSAIT, Sofia University "St. Kliment Ohridski" [7]Zhejiang University



(a) Unlocking the constraints of SFOASS with the proposed UNLOCK framework.

(b) Performance comparison of SFOASS.

Figure 1. **UNLOCK** framework ***(a)*** solves the Source-Free Occlusion-Aware Seamless Segmentation (SFOASS), enabling segmentation with 360° viewpoint coverage and occlusion-aware reasoning while adapting without requiring source data and target labels, and ***(b)*** outperforms existing SFDA methods on the Real-to-Real scenario [7] across all five segmentation metrics (*i.e.*, mAPQ for amodal panoptic, mPQ for panoptic, mIoU for semantic, mAAP for amodal instance, and mAP for instance).

Abstract

Panoramic image processing is essential for omni-context perception, yet faces constraints like distortions, perspective occlusions, and limited annotations. Previous unsupervised domain adaptation methods transfer knowledge from labeled pinhole data to unlabeled panoramic images, but they require access to source pinhole data. To address these, we introduce a more practical task, i.e.*, Source-Free Occlusion-Aware Seamless Segmentation (SFOASS), and propose its first solution, called* ***UNconstrained Learning Omni-Context Knowledge (UNLOCK)****. Specifically, UNLOCK includes two key modules: Omni Pseudo-Labeling Learning and Amodal-Driven Context Learning. While adapting without relying on source data or target labels, this framework enhances models to achieve segmentation with 360° viewpoint coverage and occlusion-aware reasoning. Furthermore, we benchmark the proposed SFOASS task through both real-to-real and synthetic-to-real adaptation settings. Experimental results show that our source-free method achieves performance comparable to source-dependent methods, yielding state-of-the-art scores of 10.9 in mAAP and 11.6 in mAP, along with an absolute improvement of +4.3 in mAPQ over the source-only method. All data and code will be made publicly available at* https://github.com/yihong-97/UNLOCK.

1. Introduction

Scene understanding is a foundational task in computer vision, essential for various downstream applications like autonomous driving [44, 77], virtual reality [11, 58], and robotics [3, 60]. Despite significant advances in the field, current methods still face challenges in achieving both comprehensive and human-like perceptions of surrounding environments, limiting their ability to fully understand scenes. Multiple critical research questions remain unsolved.

1) How to efficiently expand to 360° field of view? Traditional scene understanding methods [9, 21, 70] tailored for pinhole imagery, are limited by a narrow Field of View (FoV). In contrast, panoramic vision offers 360° view in a single shot, which has the potential to facilitate more comprehensive scene understanding [1, 19]. However, directly applying traditional methods to panoramic data often results in unreliable performance due to challenges such as

*Equal contribution. †Corresponding authors (e-mail: kailun.yang@hnu.edu.cn, zhanghuihby@126.com).

distortions [92]. For effective panoramic understanding, it is essential to learn omni-context knowledge when transferring from the pinhole domain to the panoramic domain. This knowledge transfer is crucial for accurate feature interpretation and wide-FoV representation, ultimately enabling seamless 360° scene understanding [84, 85].

2) How can amodal prediction be improved to extend the depth of view? Amodal perception [34, 53], the ability to predict an occluded object's complete shape, is key to robust scene understanding [83] and mirrors human-like perception. Extending the depth of view (vertical to field of view) means that models can achieve seamless perception, which refers to pixel-level segmentation of the background and occlusion-aware instance-level segmentation of the foreground. Beyond the previous amodal segmentation methods [16, 98], how to improve occlusion-aware reasoning like OASS [7] in both directions (depth and field of view) simultaneously remains challenging.

3) How can models adapt to panoramic domains without source data? Apart from the field and depth of view, the scarcity of labeled data presents a significant limitation in panoramic vision. To address the lack of labeled data, Unsupervised Domain Adaptation (UDA) leverages labeled data from the source domain while training on unlabeled data from the target domain [7, 24, 25, 28, 32, 65]. However, UDA methods still rely on access to source domain, which can often be challenging or even impossible due to privacy concerns or commercial restrictions on source datasets, such as in autonomous driving [40, 95, 96]. This restriction leads us to source-free learning, *i.e*. how to adapt models to panoramic domains without source data. However, it remains under-explored in amodal panoramic vision.

In this work, to address the aforementioned challenges in scene understanding, we extend the OASS [7] task and introduce a more practical task, *i.e.*, Source-Free Occlusion-Aware Seamless Segmentation (SFOASS). To tackle this task, we propose its first solution, called **UNconstrained Learning Omni-Context Knowledge (UNLOCK)**. As shown in Fig. 1a, UNLOCK improves panoramic seamless segmentation performance under adaptation constraints without requiring source data and target labels. Specifically, to effectively capture the domain-invariant knowledge from the source domain, a method called Omni Pseudo-Labeling Learning (OPLL) is proposed. Besides, to learn intra-domain knowledge and integrate the domain-invariant knowledge, we put forward the Amodal-Driven Context Learning (ADCL) strategy. Together, OPLL and ADCL adapt the source model trained on pinhole data to target panoramic images while enhancing panoramic perception and occlusion-aware reasoning. With OPLL as the key and ADCL performing the unlocking action, they work in synergy to unlock the constraints in SFOASS, thereby enabling seamless perception of the model towards panoramic data.

Aside from Real-to-Real source-free adaptation on the *KITTI360-APS→BlendPASS* benchmark [7], we pioneer Synthetic-to-Real adaptation in OASS and SFOASS and introduce *AmodalSynthDrive→BlendPASS* to investigate the potential of adapting synthetic panoramic images to the real panoramic image domain. Extensive experiments are conducted on both SFOASS benchmarks. Without access to the original source data (neither images nor labels), our UNLOCK framework can obtain on-par performance as compared to UDA-based methods that require all source data. Surprisingly, as shown in Fig. 1b, results on the Real-to-Real scenario [7] show an absolute improvement of $+4.3$ in mAPQ, reaching 26.4, compared to the source-only method, and outperforms UDA methods using $12K$ images in instance-level segmentation, with state-of-the-art scores of 10.9 in mAAP and 11.6 in mAP, respectively.

In this work, we propose contributions as follows:

- We introduce the Source-Free Occlusion-Aware Seamless Segmentation (SFOASS) task and its first solution, UNconstrained Learning Omni-Context Knowledge (UNLOCK). The UNLOCK framework enables models to achieve segmentation with 360° viewpoint coverage and occlusion-aware reasoning while adapting without the need for source data and target labels.
- To fully exploit the knowledge from source and target domains, Omni Pseudo-Labeling Learning (OPLL) and Amodal-Driven Context Learning (ADCL) strategies are introduced in UNLOCK to achieve occlusion-aware seamless segmentation for panoramic images.
- Extensive experiments conducted on two SFOASS scenarios, Real-to-Real and Synthetic-to-Real, demonstrate the effectiveness of the UNLOCK framework, highlighting its ability to achieve competitive performance.

2. Related Work

Seamless segmentation. Seamless segmentation, as introduced by [7], aims to achieve amodal-level segmentation for large-FoV images, enabling unified, occlusion-aware scene understanding. Wide-FoV segmentation on fisheye [12, 56, 59, 78, 80] and panoramic images [22, 35, 75, 76, 97] facilitates comprehensive 360° scene comprehension [74]. Panoramic panoptic segmentation enhances scene understanding by providing instance-level insights [17, 30, 47, 52]. Amodal segmentation extends this concept by predicting both visible and occluded object regions [4, 10, 15, 23, 38, 41, 61, 63]. Li *et al*. [34] pioneered amodal instance segmentation, introducing occlusion-aware segmentation through iterative regression. Subsequent datasets adapted for amodal segmentation [98] and occlusion classification [53] have driven further advances. Shape and contour priors [8, 18, 36, 37, 69] have further refined segmentation accuracy. In addition to instance segmentation, amodal panoptic segmentation has been explored [5, 26, 98]. Mo-

han *et al*. [48, 49] combined semantic and amodal segmentation for amodal panoptic segmentation. Recent methods, including pix2gestalt [51], AISDiff [64], and Xu *et al*. [72], leverage diffusion model priors for amodal segmentation. Although seamless segmentation [7] unifies aforementioned segmentation paradigms to form occlusion-aware scene understanding for panoramic vision, our work emphasizes source-free seamless segmentation, aiming for high accuracy even with restricted access to labeled source data.

Source-free domain adaptation. Existing works dealing with domain gaps on semantic segmentation tasks mostly focus on UDA where labeled source domain data is required [20, 62, 68, 71, 84, 84, 85, 91, 93, 94, 99]. Source-Free Domain Adaptation (SFDA) [6, 14, 33, 40, 43, 54, 67, 79, 81] is a more practical approach compared with UDA, as it removes the need for access to source data during the adaptation process. As sub-directions of scene segmentation, domain adaptation of panoptic segmentation is challenging as it must distinguish and integrate semantic and instance information across domains, as highlighted in existing works [27, 45, 46, 55, 87]. To overcome the limitations of the FoV, domain adaptation in panoramic segmentation facilitates comprehensive scene understanding across a full 360° perception, while effectively harnessing the rich knowledge from label-dense pinhole domains. Researchers have explored both UDA [31, 84, 85, 88–90, 94] and SFDA [95, 96] settings for panoramic segmentation. Jaus *et al*. [29, 30] illustrate the need for panoramic panoptic segmentation in autonomous driving. Zheng *et al*. [95, 96] for the first time, explored SFDA in panoramic semantic segmentation. However, the SFDA challenge remains unexplored for OASS [7]. In this work, we introduce the novel SFOASS task and benchmark it across two scenarios. We also proposed the first solution UNLOCK, a novel method designed to unlock constraints posed by SFOASS.

3. UNLOCK Framework

3.1. Overview

OASS. In the OASS [7], the objective is to use a labeled source pinhole domain $\mathcal{D}^{pin}=\{x_i^{pin}, y_i^{pin}\}_{i=1}^{N^{pin}}$ and an unlabeled target panoramic domain $\mathcal{D}^{pan}=\{x_i^{pan}\}_{i=1}^{N^{pan}}$ to obtain an OASS model F that performs well in $\mathcal{D}^{pan}$. Both domains share the same C categories, which can be further divided into C^{stu} *Stuff* classes and C^{thi} *Thing* classes. Ultimately, the F should be able to output five segmentation results with the input panoramic image at once: semantic, instance, amodal instance, panoptic, and amodal panoptic maps. Typically, F is designed to contain a shared encoder and three detection branches: semantic, instance, and amodal instance branches, each outputting corresponding predictions $p_i^{sem}, p_i^{ins}, p_i^{ains}$, denoted as

$$p_i^{sem}, p_i^{ins}, p_i^{ains} = F(x_i), \tag{1}$$

where $p_i^{sem} \in \mathbb{R}^{H\times W\times C}$ represents the semantic prediction of the input image $x_i \in \mathbb{R}^{H\times W\times 3}$. p_i^{ins} and p_i^{ains} represent the instance and amodal instance predictions, respectively, with each including $class \in \mathbb{R}^{C^{thi}}$, $score \in [0, 1]$, and $mask \in \mathbb{R}^{H\times W}$ predictions for jth object. During adaptation, the label y_i of x_i consists of three components: semantic y_i^{sem}, instance y_i^{ins}, and amodal instance y_i^{ains} labels, used for supervised learning in their respective branches. In the inference phase, the predictions from all three branches are fused to produce five segmentation maps of OASS.

SFOASS. Considering privacy and storage limitations, this work introduces a more practical task, *i.e.*, SFOASS. In the adaptation phase, the source pinhole domain $\mathcal{D}^{pin}$ is locked and inaccessible. Given a source model F^{pin} well-trained from $\mathcal{D}^{pin}$, and $\mathcal{D}^{pan}$, the goal of SFOASS is to adapt the source model F^{pin} with only unlabeled panoramic data $\{x_i^{pan}\}_{i=1}^{N^{pan}}$, obtaining a target model F^{pan} that performs well in the target panoramic domain. (For ease of expression, the superscript "pan" for target panoramic domain will be omitted henceforth.)

How can models adapt to panoramic domains without source data, while expanding the field and depth of view? To answer this, we propose a novel method called **UNconstrained Learning Omni-Context Knowledge (UNLOCK)**. It can further enhance the performance of panoramic seamless prediction while improving occlusion-aware reasoning. Specifically, as shown in Fig. 2, in response to the challenge that existing self-training methods are ineffective for the SFOASS task, we designed an **Omni Pseudo-Labeling Learning (OPLL)** approach. This approach leverages all predictions to generate omni soft labels. These generated labels serve as the "key" to the proposed UNLOCK. Furthermore, we propose a brand-new **Amodal-Driven Contextual Learning (ADCL)** strategy, which effectively resolves the conflict between the real object shapes and contextual knowledge for the occluded object. Ultimately, the ADCL holds the "key" and performs the "unlocking" action, completing the source-free adaptation to the panoramic domain.

3.2. Omni Pseudo-Labeling Learning

In the SFOASS, only unlabeled target panoramic images $\{x_i\}_{i=1}^{N}$ are accessible. Common solutions [2, 6, 33] typically involve generating pseudo-labels of target images for training. Pseudo-label generation usually entails filtering model predictions based on pre-defined thresholds. However, we found that directly using the filtered pseudo-labels for self-training in the SFOASS led to performance degradation for instance and amodal instance branches compared to the source model. The OASS task consists of a pixel-wise semantic branch and two object-wise instance-level branches (instance and amodal instance) based on region proposals. For the instance-level branches, each object is

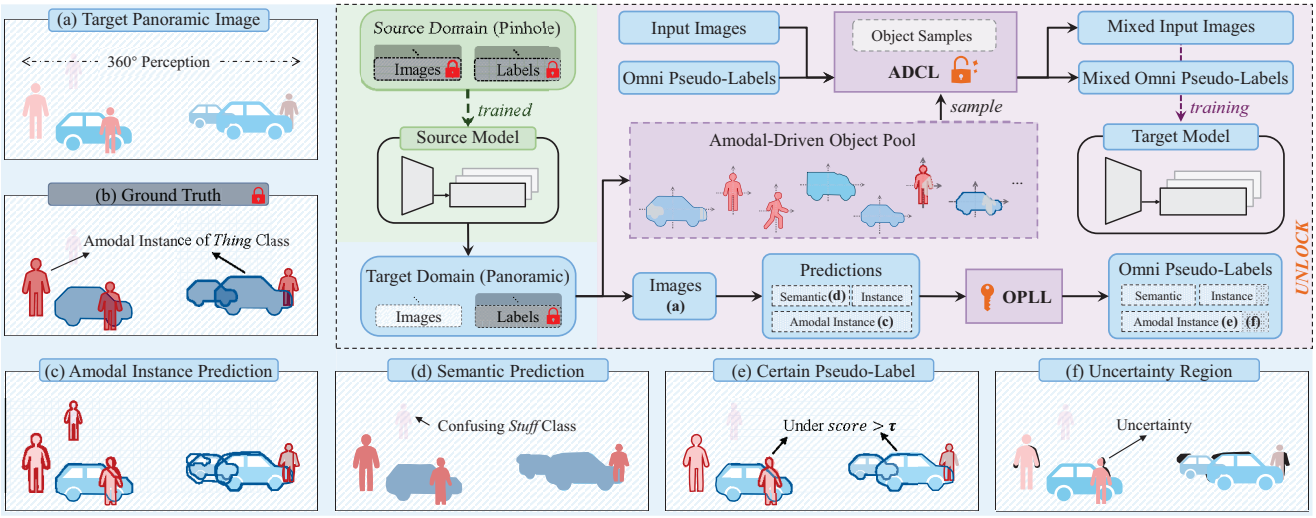


Figure 2. Illustration of the proposed UNLOCK. To address the challenge of the inaccessible source domain and target labels in SFOASS, we propose Omni Pseudo-labeling Learning (OPLL) and Amodal-Driven Contextual Learning (ADCL) modules. OPLL serves as the key, and ADCL, holding the key, unlocks the constraints of SFOASS, enabling effective adaptation of the target model to the panoramic domain.

assigned a binary mask for binary cross-entropy loss. However, inaccurate pseudo-labels can hinder learning. When object parts are mistakenly labeled as background, even if the model correctly detects the objects, the pseudo-labels may force it to incorrectly classify them as absent.

(a) Incorrect predictions (cyclists in red) of the *Stuff* class.

(b) Omni pseudo-label generation with CS thresholds.

Figure 3. Solution to the challenge of predictions from the source model in OPLL, illustrated with the amodal instance branch.

To address this, we developed the OPLL method, which introduces generating omni pseudo-labels using Class-wise Self-tuning (CS) thresholds and applying uncertainty-guided instance loss. Since the two instance-level branches perform the same operations, we will introduce one branch as an example below. Specifically, we first utilized the semantic branch's predictions to impose additional constraints on the predictions of both instance-level branches. As shown in Fig. 3a, we observed that the instance-level branch, constrained by local information, may misclassify the statue in the shop window as a cyclist of *Thing* class. In contrast, the semantic branch, aided by global information, can usually understand the scenes correctly. For each panoramic image x_i, we obtain the *Thing* mask m^{thi} by the semantic prediction, then use it to revise the object mask of instance-level predictions $p_i^{il}, il \in \{ins, ains\}$:

$$m^{thi} = \begin{cases} 1, & \text{if argmax} p^{sem} \in C^{thi} \\ 0, & \text{else} \end{cases}, \tag{2}$$

$$\hat{p}_i^{il} = p_i^{il} \cap m^{thi}. \tag{3}$$

Then, considering the inconsistent number of objects and the significant imbalance in the source model's capabilities across different *Thing* classes (as shown in *SourceOnly* of Table 2), we propose a simple and effective CS threshold method for the predictions of the SFOASS task. For each branch, the CS thresholds $\boldsymbol{\tau} = \{\tau_c\}_{c=1}^{C^{thi}}$ is determined by:

$$\tau_c = \text{argmax}_{\tau \in \{\tau^{fix}, \tau^{per}\}} N_c^{\tau}, \; c \in C^{thi}, \tag{4}$$

where $N_c^{\tau^{fix}}$ represents the number of objects of class c whose the $score$ exceeds τ^{fix} across all predictions $\{\hat{p}_i\}_{i=1}^N$, and $N_c^{\tau^{per}}$ denotes the number of objects of class c in the top τ^{per} percentage when the $score$ of all predictions $\{\hat{p}_i\}_{i=1}^N$ are sorted in descending order. For the semantic branch, the $score$ is replaced with the maximum predicted probability of p_i^{sem}. This approach generates high-quality pseudo-labels while effectively addressing the issue of low confidence scores for certain classes, caused by the source model F^{pin} inconsistent performance across classes.

Then, we filter out high-quality predictions based on CS thresholds $\boldsymbol{\tau}$ as certain pseudo-labels $\hat{y}_i^{cer}$:

$$\hat{y}_i^{cer} = \{\hat{p}_i^{(j)} | \hat{p}_i^{(j)} > \boldsymbol{\tau}\}, \tag{5}$$

where $\hat{p}_i^{(j)}$ represent the prediction of jth object in predictions p_i for input image x_i. To fully utilize the predictions,

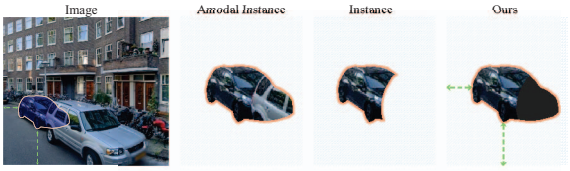


Figure 4. Comparison of object extraction methods: Amodal instance extraction mixes information from other objects, while instance-based extraction may yield incomplete shapes. Our method avoids these issues and preserves spatial awareness.

we treat the remaining masks of the other predictions as uncertainty region $\hat{y}_i^{uncer}$ for the current image:

$$\hat{y}_i^{uncer} = (1 - \mathbf{1}_{\{\sum_j \hat{y}_i^{cer} > 0\}}) \cap (\mathbf{1}_{\{\sum_j (\hat{p}_i^{(j)} | \hat{p}_i^{(j)} < \boldsymbol{\tau}) > 0\}}), \tag{6}$$

where $\mathbf{1}_{\{\cdot\}}$ is indicator function. The left-hand side of $\cap$ in Eq. 6 ensures that the object region of $\hat{y}_i^{cer}$ is not misassigned as the uncertain region. They forms the omni pseudo-labels $\hat{y}_i = \{\hat{p}_i^{(1)}, \hat{p}_i^{(2)}, ..., \hat{y}_i^{uncer}\}$ of the training panoramic image x_i. We believe that all predictions from both instance-level branches of the source model can provide effective knowledge. For the semantic branch, the j is replaced with the pixel spatial location. Finally, for each training panoramic image x_i, we obtain three types of omni pseudo-labels $\hat{y}_i^{sem}, \hat{y}_i^{ins}, \hat{y}_i^{ains}$ for training. During the adaptation process of the target model F^{pan}, following [7], the cross-entropy loss is used for the semantic branch. For the instance-level branches, we introduce a novel uncertain-guided Binary Cross Entropy (BCE) loss $\mathcal{L}_{ur}$, specifically tailored for the SFOASS task.

$$\mathcal{L}_{ur} = (1 - \hat{y}_i^{uncer}) \odot \text{BCE}(F^{pan}(x_i), \hat{y}_i^{cer}). \tag{7}$$

This approach effectively mitigates prediction errors of the target panoramic data arising from the source pinhole model, allowing the target model to focus on high-quality object samples during the adaptation process.

3.3. Amodal-Driven Contextual Learning

Through OPLL, we leverage the source model to generate omni pseudo-labels of the unlabeled panoramic data, obtaining the "key" for adapting the model to the target domain. It primarily extracts domain-invariant knowledge from the pinhole-tolerant source model. We further introduce the ADCL strategy to learn intra-domain knowledge while integrating domain-invariant knowledge. It holds the "key" to complete the "unlocking" action, and achieves the adaptation to the panoramic images in a source-free manner.

Mixing is an effective strategy often used in UDA methods [2, 50, 82]. The typical approach involves randomly extracting objects from other data and pasting them into the current training image to form new mixed training samples. However, this approach is not suitable for the OASS model with an amodal instance branch, and it can confuse semantic context and hinder model learning. Specifically, since the two instance branches in the OASS model are independent, their predictions are not correlated, leaving us with the option of relying on one of the predictions. As shown by the car in Fig. 4, the amodal instance prediction captures the complete shape of the car. However, this prediction only provides the full mask for the object, meaning that the occluded areas in the mask may inadvertently include parts of other objects. If these masks are directly used for mixing, incorrect contextual information can be introduced. On the other hand, relying on instance predictions focuses only on non-occluded regions, making it difficult for the model to learn the full shape of objects. This dual challenge complicates the learning process in SFOASS tasks.

First, we use OPLL with stricter thresholds τ'^{fix} and τ'^{per} to build a buffer pool $\mathcal{B}$ of K high-quality object samples, relying solely on predictions from the amodal instance branch to accurately capture object shapes.

$$\begin{aligned} \mathcal{B} =& \{o^{(1)}, o^{(2)}, ..., o^{(K)}\}, \\ & \text{where } o^{(k)} \in \{\hat{y}_i^{ains \to cer} | (\tau'^{fix}, \tau'^{per})\}_{i=1}^{N^{pan}}. \end{aligned} \tag{8}$$

The $o^{(k)}$ represents the filtered certain object obtained through OPLL from the panoramic data. Additionally, for each $o^{(k)} = \{o_{ful}^{(k)}, o_{ovp}^{(k)}\}$, we include not only the mask o_{ful} of the predicted full region but also record the overlapping region o_{ovp} between the object and other objects in the same image, regardless of whether they are occluded. This process results in an amodal-driven object pool $\mathcal{B}$.

Furthermore, we introduce a novel spatial-aware mixing strategy based on the amodal-driven object pool, enriching the diversity of unlabeled panoramic images by generating amodal-driven mixed samples while respecting panoramic layouts. This spatially consistent mixing ensures semantic coherence, aligning distortions in the panoramic layout even after object insertion. During the adaptation phase, we randomly select R object samples $\mathcal{O} = \{o^{(r)}\}_{r=1}^{R}$ from $\mathcal{B}$ and integrate them into the current training image x_i:

$$\tilde{x}_i = (1 - o_{ful}^{(r)}) \odot x_i + o_{ful}^{(r)} \odot x_r, \tag{9}$$

where x_r represents the image corresponding to the paste object $o^{(r)}$. For ambiguous regions of $o^{(r)}$ (*i.e.*, potentially occluded areas), we assign zero pixels during the mixing process while keeping the complete mask as the label according to the $o_{ovp}^{(r)}$.

$$\tilde{x}_i = (1 - o_{ovp}^{(r)}) \odot \tilde{x}_i. \tag{10}$$

For the three pseudo-labels $\tilde{y}_i^{sem}, \tilde{y}_i^{ins}, \tilde{y}_i^{ains}$ of the mixed image $\tilde{x}_i$, we only use $\{o_{ful}^{(r)}\}_{r=1}^{R}$ to modify them accordingly. This strategy preserves the full shape of objects while

Task	Constraint	Method	Metric	road	sidew.	build.	wall	fence	pole	tr.light	tr.sign	veget.	terrain	sky	pedes.	cyclists	car	truck	ot.veh.	van	tw.whe.
APS	🔒(y^{pan})	DATR [93]	20.3	51.8	09.2	59.9	11.9	12.0	02.0	00.0	03.9	64.6	14.1	70.4	11.1	00.0	39.3	00.0	03.2	**10.1**	01.3
		Trans4PASS [85]	22.9	53.9	14.1	69.4	19.2	11.8	03.8	00.0	05.2	67.6	16.0	77.4	15.3	**04.2**	41.1	06.6	00.0	00.0	07.4
		EDAPS [55]	23.1	54.9	17.0	66.9	18.8	14.5	05.8	04.0	04.6	68.2	16.0	72.8	19.0	00.0	36.7	**05.8**	04.4	00.0	07.2
		UnmaskFormer [7]	**26.6**	61.8	**24.7**	66.8	**20.8**	**15.8**	05.3	**04.3**	03.3	**69.0**	**18.4**	79.4	20.5	03.1	44.6	12.8	**11.3**	00.0	16.6
	🔒(x^{pan}, y^{pan})	Source-only	22.1	57.1	14.2	73.6	15.5	07.6	00.7	00.0	**10.4**	58.3	12.4	83.1	15.2	00.0	40.3	03.8	00.0	00.0	06.1
	🔓$(x^{pin}, y^{pin}, y^{pan})$	360SFUDA++ [95]	24.5	60.3	16.8	71.7	14.1	07.7	00.0	00.0	08.4	55.6	13.6	81.3	21.5	03.2	**46.3**	**19.4**	02.2	00.0	19.4
		UNLOCK (Ours)	26.4 (↑4.3)	**62.7**	14.4	**74.8**	20.2	11.2	00.8	00.0	08.0	62.9	18.3	**84.0**	**21.9**	03.6	45.8	**19.4**	00.0	03.1	**24.4**
PS	🔒(y^{pan})	DATR [93]	19.6	50.4	09.1	59.9	11.9	12.0	02.0	00.0	03.9	64.6	14.1	70.5	12.2	00.0	38.1	00.0	03.4	00.0	01.3
		Trans4PASS [85]	22.7	53.9	14.1	69.4	19.2	11.8	03.8	00.0	05.2	67.6	16.0	77.4	14.6	**04.1**	38.2	06.9	00.0	00.0	07.2
		UniDAPS [86]	22.7	**66.0**	09.5	66.3	17.4	14.3	04.8	00.0	06.1	67.2	16.1	72.7	08.3	00.0	27.3	14.8	**09.2**	00.0	08.9
		EDAPS [55]	23.1	55.0	17.1	66.8	18.7	14.5	**05.8**	04.0	04.7	68.2	16.0	72.8	19.6	00.0	37.8	01.8	04.4	00.0	07.9
		UnmaskFormer [7]	**26.2**	61.7	**24.7**	66.8	**20.8**	**15.8**	05.2	**04.3**	03.3	**69.0**	18.4	79.4	**20.9**	03.5	**43.0**	11.3	07.6	00.0	16.0
	🔒(x^{pan}, y^{pan})	Source-only	22.3	57.8	14.2	73.8	15.5	07.6	00.7	00.0	10.4	58.3	12.4	83.2	14.9	00.0	39.1	06.0	00.0	00.0	07.7
	🔓$(x^{pin}, y^{pin}, y^{pan})$	360SFUDA++ [95]	23.7	61.3	16.8	72.2	14.1	07.6	00.0	00.0	**08.4**	55.9	13.6	81.3	18.4	00.0	40.6	16.4	02.4	00.0	17.6
		UNLOCK (Ours)	25.1 (↑2.8)	60.4	14.4	**75.7**	20.2	11.2	00.7	00.0	08.0	63.6	18.3	**84.0**	16.9	00.0	40.0	**18.5**	00.0	00.0	**19.4**
SS	🔒(y^{pan})	DATR [93]	34.9	71.9	27.2	70.6	22.8	36.0	23.9	00.0	04.5	77.1	37.1	80.1	51.2	02.2	70.2	08.9	06.0	11.0	27.8
		Trans4PASS [85]	40.7	72.8	33.8	78.4	33.5	37.1	26.6	05.3	05.4	77.4	37.9	84.7	**57.5**	04.5	76.6	19.0	17.0	12.7	51.7
		UniDAPS [86]	38.5	72.3	29.2	75.8	33.9	38.9	25.9	11.4	07.5	77.6	37.4	81.4	47.6	03.0	75.6	21.3	03.4	01.3	48.9
		EDAPS [55]	40.2	74.4	35.9	77.0	**36.5**	40.4	28.0	16.1	05.1	78.1	39.5	82.3	55.5	03.3	74.4	06.7	14.1	05.2	50.9
		UnmaskFormer [7]	**43.7**	**76.5**	**37.8**	77.1	34.7	**44.1**	**28.3**	**17.8**	02.8	**78.7**	**41.7**	85.0	57.3	**06.0**	80.6	23.5	**21.7**	**18.8**	53.6
	🔒(x^{pan}, y^{pan})	Source-only	38.7	73.0	29.1	82.0	31.2	31.4	18.0	00.0	16.0	74.1	33.7	88.7	50.2	04.0	80.4	18.0	11.0	04.6	50.2
	🔓$(x^{pin}, y^{pin}, y^{pan})$	360SFUDA++ [95]	39.4	74.6	32.2	81.2	30.7	31.9	18.1	00.0	15.6	72.6	38.3	89.2	50.6	04.6	80.5	20.4	11.6	05.3	51.0
		UNLOCK (Ours)	41.6 (↑2.9)	74.7	32.1	**83.7**	34.7	40.8	18.5	00.0	**18.9**	76.1	39.9	**89.5**	53.7	05.8	**82.2**	**26.1**	00.8	14.0	**57.8**

Table 1. **Scene Segmentation** results on the **K2B** benchmark. Metrics are reported as: mAPQ for Amodal Panoptic Segmentation (**APS**), mPQ for Panoptic Segmentation (**PS**), and mIoU for Semantic Segmentation (**SS**). Per-class results are reported as APQ, PQ, and IoU. ↑ represents the improvement over the baseline of the Source-only method. x^{pin} and x^{pan} represent the images, and y^{pin} and y^{pan} represent the labels, from the pinhole source and panoramic target domains, respectively.

avoiding confusion about the semantic content of other objects. The zero-pixel ambiguous regions act as a masked strategy to enhance the model's ability to reconstruct occluded areas and infer the complete shape of occluded objects. This improves the model's ability to learn contextual knowledge from the target panoramic domain, improving its understanding of both individual objects and their relationships within the scene.

4. Experiments

4.1. Datasets

In this paper, we extend the Real-to-Real adaptation on the *KITTI360-APS→BlendPASS* (**K2B** for short)) benchmark [7] to SFOASS. Moreover, we pioneer Synthetic-to-Real adaptation in both OASS and SFOASS, introducing *AmodalSynthDrive→BlendPASS* benchmark (**A2B** for short) to conduct seamless segmentation research from synthetic to real scenarios.

Source pinhole domain: (1) KITTI360-APS [48]: It extends KITTI360 [39] and includes additional annotations for inmodal and amodal instances. A total of $12,320$ annotated images (1408×376 pixels) are captured using pinhole cameras across 9 cities. It covers 11 *Stuff* classes and 7 *Thing* classes. (2) AmodalSynthDrive [57]: The first synthetic dataset applicable to the OASS task, generated using CARLA [13]. It provides images and annotations from four surround-view cameras with a pinhole perspective, with $10,500$ training images (1920×1080 pixels) for each viewpoint. It cover 11 *Stuff* classes and 7 *Thing* classes.

Target panoramic domain: The BlendPASS dataset [7] consists of $2,000$ unlabeled training panoramic images and 100 labeled test panoramic images (2048×400 pixels), captured in real driving scenes across 40 cities on multiple continents. It covers 11 *Stuff* and 8 *Thing* classes.

Domain gap analysis: Fig. 5 shows the feature distributions of *car* and *pedestrians* between the target domain and two source domains, where there is a large domain gap between different datasets, especially for the **A2B**.

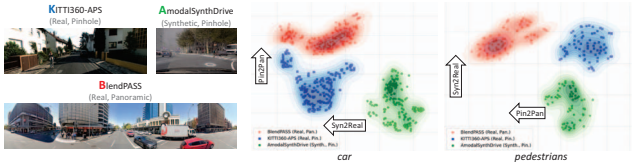


Figure 5. Example images from the datasets [7, 48, 57] and the class distribution visualized via t-SNE [66].

4.2. Experiment Setups

Following [7], we use the proposed architecture as the backbone of our UNLOCK and the reproduced 360SFUDA++ [95]. In the **A2B** benchmark, we reproduce existing methods [55, 85, 86, 93, 95] and train the pinhole source model following the framework in [7]. In the **K2B** benchmark, we directly use the Source-only model from [7] as the pinhole source model. The AdamW optimizer [42] is employed, with a learning rate set to 1×10^{-7}, and a weight decay of 0.01, followed by polynomial decay. The batch size is set to 4, with crop sizes of 376×376 for **K2B** and 400×400 for **A2B**, and both are trained for $10k$ iterations. In UNLOCK, τ^{fix} and τ^{per} for amodal instance predictions are set to 0.3 and 0.5; for instance predictions to 0.5 and 0.3; and for semantic predictions, to 0.5 and 0.8. For ADCL, they are set to 0.95 and 0.1, with the sample object count R

Task		Method	Metric	pedes.	cyclists	car	truck	ot. veh.	van	tw. whe.
AIS		DATR [93]	08.7	13.1	00.0	30.6	06.9	04.7	**01.7**	03.8
		Trans4PASS [85]	09.9	16.0	**00.2**	31.7	08.3	06.0	00.4	06.4
		EDAPS [55]	10.7	15.8	00.1	30.0	09.0	**12.2**	00.4	07.4
		UnmaskFormer [7]	10.5	16.1	00.1	34.1	12.3	02.2	00.6	**08.2**
		Source-only	10.2	15.2	00.0	33.4	12.6	03.3	00.4	06.8
		360SFUDA++ [95]	10.3	15.2	00.0	33.4	13.1	03.3	00.4	06.8
		UNLOCK (Ours)	**10.9** (↑0.7)	**17.1**	00.0	**34.9**	**14.6**	01.7	00.2	07.9
IS		DATR [93]	08.7	14.2	00.0	31.2	07.6	03.7	00.4	03.6
		Trans4PASS [85]	10.0	16.5	00.0	32.2	10.2	05.3	00.2	05.6
		UniDAPS [86]	03.4	02.3	00.0	11.3	06.2	02.8	00.0	01.5
		EDAPS [55]	10.3	16.6	00.0	30.8	06.5	**11.4**	00.4	06.2
		UnmaskFormer [7]	11.1	17.6	00.0	35.2	14.2	02.9	**00.8**	07.1
		Source-only	10.5	15.8	00.0	33.9	13.5	03.6	00.2	06.8
		360SFUDA++ [95]	10.5	15.7	00.0	33.9	13.3	03.6	00.2	06.8
		UNLOCK (Ours)	**11.6** (↑1.1)	**17.7**	00.0	**36.1**	**16.3**	03.4	00.1	**07.4**

Task		Method	Metric	person	rider	car	truck	bus	motor.	bicycle
AIS		DATR [93]	08.1	14.1	**00.4**	23.5	09.6	04.3	04.7	00.0
		Trans4PASS [85]	09.2	13.0	00.2	24.7	**13.0**	07.6	05.7	00.0
		EDAPS [55]	09.5	16.7	00.2	25.2	11.6	06.6	06.1	00.0
		UnmaskFormer [7]	10.2	16.6	00.1	26.5	12.8	08.6	**06.4**	00.0
		Source-only	09.7	16.0	00.0	24.9	11.9	**09.1**	05.6	00.0
		360SFUDA++ [95]	09.6	16.2	00.0	25.3	11.7	08.9	05.2	00.0
		UNLOCK (Ours)	**10.0** (↑0.3)	**17.3**	00.0	**28.9**	09.2	08.3	**06.4**	00.0
IS		DATR [93]	08.0	14.9	**00.4**	25.3	09.2	01.0	04.9	00.0
		Trans4PASS [85]	09.0	15.0	00.1	27.2	11.8	04.1	05.0	00.0
		UniDAPS [86]	00.3	00.8	00.0	00.8	00.1	00.0	00.5	00.0
		EDAPS [55]	09.5	16.4	00.1	28.2	10.7	05.3	05.8	00.0
		UnmaskFormer [7]	09.6	17.0	00.1	28.4	**11.9**	03.8	05.9	00.0
		Source-only	09.3	17.4	00.0	26.9	11.6	04.4	04.9	00.0
		360SFUDA++ [95]	09.4	17.5	00.0	27.1	11.1	05.4	04.8	00.0
		UNLOCK (Ours)	**09.7** (↑0.4)	**17.7**	00.0	**29.3**	09.0	**05.9**	**06.1**	00.0

Table 2. **Instance-level Segmentation** results on the **K2B** (left) and **A2B** (right) benchmark. Metrics are reported as: mAAP for Amodal Instance Segmentation (**AIS**) and mAAP for Instance Segmentation (**IS**). Per-class results are reported as AAP and AP.

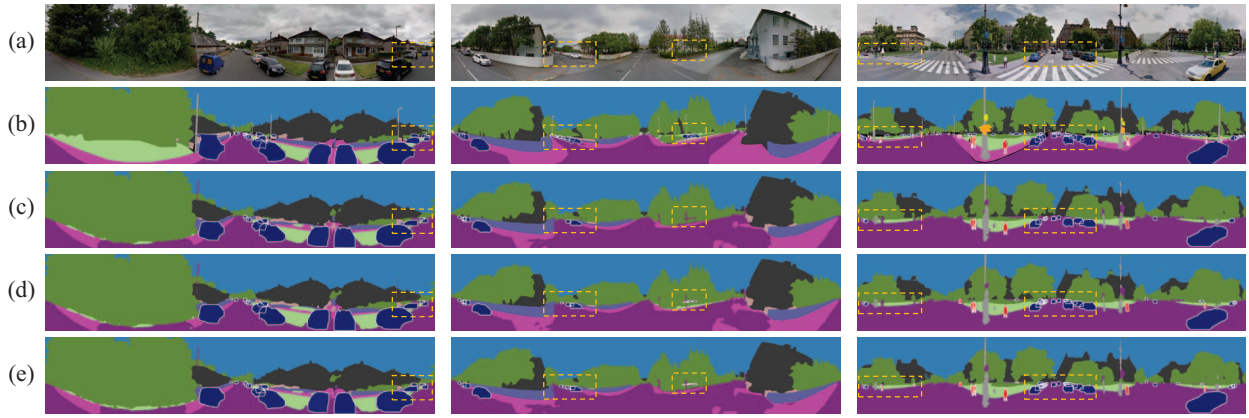


Figure 6. **Visualization results.** From top to bottom are (a) Image, (b) GT, (c) Source-only, and (d) 360SFUDA++ [95], and (e) UNLOCK.

set to 10. The analysis of these thresholds is provided in the supplementary material. Our experiments are conducted on a single NVIDIA GTX 3090 GPU using PyTorch.

4.3. Results of SFOASS

K2B. As shown by the Real-to-Real adaptation results in Table 1 and the left side of Table 2, UNLOCK sets a new benchmark for SFOASS, delivering outstanding performance and effectively addressing the unique challenges of limited FoV, occlusion, and constrained data. Compared to the baseline Source-only, UNLOCK outperforms in key metrics for scene understanding including mAPQ, mPQ, and mIoU, with improvements of $+4.3$, $+2.8$, and $+2.9$, respectively. Even compared to UnmaskFormer [7], which has access to both source domain and target images, UNLOCK achieves comparable results in mAPQ. Notably, in instance-level segmentation, UNLOCK surpasses all existing UDA methods[7, 55, 85, 93]—even without access to source domain data containing $12K$ images-label pairs, reaching the highest mAAP of 10.9 and mAP of 11.6, showcasing its robustness and adaptability. As shown in Fig. 6, UNLOCK successfully and precisely segments numerous *Thing* objects with better shape coherence.

A2B. The results from another Synthetic-to-Real adaptation scenario, shown in Table 3 and the right side of Table 2, further demonstrate UNLOCK's exceptional performance. Compared to 360SFUDA++ [95], which is specifically designed for panoramic data, our method significantly outperforms across all SFOASS metrics. This success is driven by the OPLL pseudo-label generation strategy, which is tailored specifically for SFOASS. The instance-level segmentation results highlight a clear advantage for our method, reflecting its enhanced accuracy in segmenting individual object instances within complex scenes.

4.4. Ablation Study

We conduct a series of ablation studies on the **K2B** scenario to evaluate the effectiveness of our method.

Effectiveness of OPLL. An ablation study was conducted to validate the effectiveness of the OPLL. Table 4 presents the comparison of three different approaches for self-training. Using the final predicted labels from the source model as pseudo-labels or directly using all predictions from the three branches fails to adapt the instance-level branches to the panoramic domain, as evidenced by both mAAP and mAP scores, which are even worse than the no-adaptation Source-only model. With the OPLL designed by us, the knowledge of the source model can be effectively adapted to the target domain by using omni pseudo-labels. Compared with Source-only, OPLL achieves a significant

Task	Constraint	Method	Metric	road	sidew.	build.	wall	fence	pole	lights	sign	veget.	terrain	sky	person	rider	car	truck	bus	motor.	bicycle
APS	🔒 (y^{pan})	DATR [93]	19.4	36.1	00.0	72.4	00.0	01.3	02.4	00.0	07.1	64.8	00.0	88.9	08.3	02.7	31.6	18.4	00.0	14.8	00.0
		Trans4PASS [85]	21.5	38.7	00.0	71.9	00.0	**10.0**	03.8	00.0	**07.4**	65.3	00.0	89.2	11.1	**03.7**	35.3	25.4	07.6	15.9	**00.9**
		EDAPS [55]	21.7	45.5	00.0	**74.9**	00.0	08.5	04.1	00.0	06.0	63.7	00.0	89.5	12.1	02.8	35.9	20.7	09.7	17.0	00.0
		UnmaskFormer [7]	**22.9**	44.1	00.6	67.5	00.0	09.6	**04.3**	00.0	05.0	**67.0**	00.0	**90.2**	16.6	00.0	38.7	**29.2**	**17.1**	**22.2**	00.0
	🔒 (x^{pan}, y^{pan})	Source-only	18.9	47.1	01.9	68.5	00.0	00.9	00.5	00.0	01.6	43.0	00.0	84.7	09.5	00.0	36.3	21.5	07.5	16.7	00.0
		360SFUDA++ [95]	19.4	46.6	01.9	68.1	00.0	01.0	00.5	00.0	01.7	48.2	00.0	86.1	**18.8**	00.0	38.5	18.1	06.4	14.1	00.0
	🔓 $(x^{pin}, y^{pin}, y^{pan})$	UNLOCK (Ours)	21.4 (↑2.5)	**50.8**	**04.7**	73.5	00.0	04.8	00.6	00.0	03.9	49.8	00.0	86.2	17.5	01.2	**41.4**	20.2	09.3	20.9	00.0
PS	🔒 (y^{pan})	DATR [93]	19.5	35.7	00.0	71.5	00.0	01.3	01.8	00.0	07.1	64.8	00.0	88.9	08.3	**05.0**	32.2	15.8	02.4	15.5	00.0
		Trans4PASS [85]	20.9	38.1	00.0	71.9	00.0	**10.0**	03.8	00.0	07.4	65.3	00.0	89.2	09.6	00.0	36.5	25.0	03.3	15.9	00.0
		UniDAPS [86]	18.2	53.1	00.0	72.4	**01.8**	07.8	**05.5**	00.0	**07.7**	67.8	00.0	88.7	07.0	00.0	07.7	01.6	00.0	07.0	00.0
		EDAPS [55]	21.9	46.7	00.0	**75.1**	00.0	08.5	04.1	00.0	06.1	63.6	00.0	89.5	11.8	02.5	36.9	23.4	**08.9**	16.5	00.0
		UnmaskFormer [7]	**20.7**	42.9	00.6	67.0	00.0	09.6	03.6	00.0	05.0	66.9	00.0	**90.2**	**16.6**	00.0	37.3	**25.9**	07.5	00.0	00.0
	🔒 (x^{pan}, y^{pan})	Source-only	19.0	48.8	01.9	68.6	00.0	00.9	00.5	00.0	01.6	42.8	00.0	84.7	12.7	00.0	36.2	20.3	06.4	**16.8**	00.0
		360SFUDA++ [95]	19.3	46.5	01.9	69.9	00.0	01.0	00.5	00.0	01.7	47.9	00.0	86.0	15.6	00.0	36.5	18.9	06.5	15.2	00.0
	🔓 $(x^{pin}, y^{pin}, y^{pan})$	UNLOCK (Ours)	**20.4** (↑1.4)	50.5	**04.2**	74.0	00.0	04.8	00.6	00.0	03.9	49.2	00.0	86.3	14.3	00.0	**38.1**	18.8	05.7	16.8	00.0
SS	🔒 (y^{pan})	DATR [93]	31.5	59.5	06.3	77.7	00.2	09.2	22.9	02.4	10.3	76.6	00.0	93.4	37.7	01.9	68.8	20.8	01.5	54.1	24.2
		Trans4PASS [85]	34.3	58.9	07.6	77.3	00.2	28.7	26.7	05.2	**11.1**	76.1	00.0	93.6	35.8	01.8	**73.6**	28.3	06.8	55.2	30.1
		UniDAPS [86]	36.0	61.5	07.5	76.6	00.8	30.4	**27.7**	10.2	**11.1**	**77.4**	00.0	93.6	**47.3**	02.1	49.4	25.9	**33.0**	**62.3**	**30.6**
		EDAPS [55]	**36.2**	62.2	10.8	**81.1**	00.5	28.7	22.9	10.6	10.5	75.5	**00.1**	93.3	42.7	**02.3**	72.1	33.6	18.8	58.6	26.7
		UnmaskFormer [7]	36.1	60.6	13.3	74.2	00.9	**32.7**	21.6	**14.2**	06.9	76.7	00.0	**93.9**	45.2	01.2	73.3	**34.1**	22.0	56.9	22.7
	🔒 (x^{pan}, y^{pan})	Source-only	31.0	67.1	21.5	76.9	03.2	17.0	17.7	02.7	06.3	64.7	00.0	90.7	40.0	01.8	69.2	19.5	04.6	47.4	08.5
		360SFUDA++ [95]	31.2	64.8	21.3	77.7	**03.4**	16.2	18.9	02.7	06.0	66.6	00.0	91.7	41.9	02.0	67.9	19.1	04.6	49.1	07.9
	🔓 $(x^{pin}, y^{pin}, y^{pan})$	UNLOCK (Ours)	34.3 (↑3.3)	**69.1**	**25.2**	81.5	00.1	25.1	20.0	03.7	10.5	68.0	00.0	91.6	45.9	01.6	71.4	24.0	07.2	50.8	21.4

Table 3. **Scene Segmentation** results on the **A2B** benchmark. Metrics are reported as: mAPQ for Amodal Panoptic Segmentation (**APS**), mPQ for Panoptic Segmentation (**PS**), and mIoU for Semantic Segmentation (**SS**). Per-class results are reported as APQ, PQ, and IoU.

	mAPQ	mPQ	mIoU	mAAP	mAP
Source-only	22.13	22.30	38.65	10.22	**10.54**
Pseudo-Labels	21.90	22.07	**39.12**	10.09	10.37
All Predictions	22.38	22.31	38.90	10.15	10.50
OPLL (Ours)	**24.71**	**24.00**	39.03	**10.52**	10.52

Table 4. Ablation analysis for **OPLL**.

Backbone	mAPQ	mPQ	mIoU	mAAP	mAP
Source-only	22.1	22.3	38.7	10.2	10.5
Trans4PASS [85]	26.1	**25.3**	40.0	**10.9**	11.3
EDAPS [55]	26.3	25.1	40.5	10.7	11.4
UnmaskFormer [7]	**26.4**	25.1	**41.6**	**10.9**	**11.6**

Table 5. Ablation analysis for **backbones**.

OPLL	ADCL	mAPQ	mPQ	mIoU	mAAP	mAP
		22.13	22.30	38.65	10.22	10.54
✓		24.71	24.00	39.03	10.52	10.52
	✓	25.84	24.55	40.31	**10.93**	11.47
✓	✓	**26.41**	**25.07**	**41.64**	10.91	**11.58**

Table 6. Ablation analysis for **components**.

	mAPQ	mPQ	mIoU	mAAP	mAP
Source-only	22.13	22.30	38.65	10.22	10.54
Instance	25.03	24.23	38.93	10.33	10.55
Amodal Instance	25.71	24.50	39.06	10.77	11.22
AoMix [7]	25.32	**24.67**	38.04	10.62	11.26
ADCL-Random	24.81	23.90	38.92	10.69	11.19
ADCL-Zero (Ours)	**25.84**	24.55	**40.31**	**10.93**	**11.47**

Table 7. Ablation analysis for **ADCL**.

Method	mIoU
SFDA [40]	42.70
DATC [73]	43.06
360SFUDA [96]	50.12
360SFUDA++ [95]	52.99
UNLOCK (Ours)	**54.55**

Table 8. **SS** results on the **C2D** benchmark.

improvement of 2.58 in mAPQ.

Effectiveness of ADCL. Another ablation study was conducted to validate the effectiveness of the ADCL. Table 7 presents a comparison of different object mixing strategies. Due to incomplete object shapes, instance-based mixing yields limited improvement in mAPQ and mAAP. Directly using amodal instances causes confusion in contextual information. Unlike AoMix [7], which zeros out full amodal object regions of the input image, our method in UNLOCK sets only the overlapping regions to zero (ADCL-Zero), effectively preserving contextual cues. Replacing overlaps with random values (ADCL-Random) yields inferior results, while ADCL-Zero achieves the best mAPQ of 25.84.

Component Ablation. As shown in Table 6, by combining OPLL and ADCL, our UNLOCK method achieves the best overall performance, outperforming the Source-only trained on the source pinhole domain with +4.28 gains in mAPQ. It shows that our methods can successfully adapt to the target panoramic domain under source-free constraints.

Backbone Ablation. As shown in Table 5, we evaluated the performance of UNLOCK across different backbones to verify the effectiveness of the UNLOCK strategy itself. The results indicate that UNLOCK maintains consistent performance across diverse backbone architectures, confirming its backbone-agnostic nature as an SFDA framework.

4.5. Results of Panoramic Semantic Segmentation

To investigate the generalization capability of UNLOCK, we extend UNLOCK to the Semantic Segmentation (**SS**) task and evaluate it on the Cityscapes-to-DensePASS (**C2D**) benchmark. As shown in Table 8, UNLOCK achieves a significant improvement of +1.56 in mIoU over 360SFUDA++ [95]. Under the pinhole-to-panoramic **C2D** adaptation scenario, our UNLOCK achieves the new state-of-the-art result of 54.55 in mIoU.

5. Conclusion

In this paper, we address key constraints in comprehensive scene understanding, to achieve seamless segmentation with 360° viewpoint coverage and occlusion-aware reasoning, while adapting without relying on source data and target labels. To this end, we introduce a new task, Source-Free Occlusion-Aware Seamless Segmentation (SFOASS), and propose its first solution UNLOCK. We further applied two SFOASS benchmarks for evaluation: Real-to-Real and Synthetic-to-Real. Extensive experimental results demonstrate the effectiveness of the proposed method.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (No. 62473139, 62027810), the Major Research Plan of the National Natural Science Foundation of China (No. 92148204), the Top Ten Technical Research Projects of Hunan Province (No. 2024GK1010), the Key Program of Natural Science Foundation of Hunan Province, China (No. 2025JJ30024), the National Key Research and Development Program of China (No. 2024YFB4708900), the Hunan Provincial Research and Development Project (No. 2025QK3019), in part by the Open Research Project of the State Key Laboratory of Industrial Control Technology, China (No. ICT2025B20), and the Science and Technology Project of State Grid Corporation of China (SGCC) Co., LTD (No. 5700-202423229A-1-1-ZN).

References

[1] Hao Ai, Zidong Cao, and Lin Wang. A survey of representation learning, optimization strategies, and applications for omnidirectional vision. *International Journal of Computer Vision*, 2025. 1

[2] Ibrahim Batuhan Akkaya and Ugur Halici. Self-training via metric learning for source-free domain adaptation of semantic segmentation. *arXiv preprint arXiv:2212.04227*, 2022. 3, 5

[3] Alberto Bacchin, Leonardo Barcellona, Sepideh Shamsizadeh, Emilio Olivastri, Alberto Pretto, and Emanuele Menegatti. PanNote: an automatic tool for panoramic image annotation of people's positions. In *ICRA*, 2024. 1

[4] Seunghyeok Back, Joosoon Lee, Taewon Kim, Sangjun Noh, Raeyoung Kang, Seongho Bak, and Kyoobin Lee. Unseen object amodal instance segmentation via hierarchical occlusion modeling. In *ICRA*, 2022. 2

[5] Jasmin Breitenstein and Tim Fingscheidt. Amodal cityscapes: A new dataset, its generation, and an amodal semantic segmentation challenge baseline. In *IV*, 2022. 2

[6] Yihong Cao, Hui Zhang, Xiao Lu, Zheng Xiao, Kailun Yang, and Yaonan Wang. Towards source-free domain adaptive semantic segmentation via importance-aware and prototype-contrast learning. *IEEE Transactions on Intelligent Vehicles*, 2024. 3

[7] Yihong Cao, Jiaming Zhang, Hao Shi, Kunyu Peng, Yuhongxuan Zhang, Hui Zhang, Rainer Stiefelhagen, and Kailun Yang. Occlusion-aware seamless segmentation. In *ECCV*, 2024. 1, 2, 3, 5, 6, 7, 8

[8] Junjie Chen, Li Niu, Jianfu Zhang, Jianlou Si, Chen Qian, and Liqing Zhang. Amodal instance segmentation via prior-guided expansion. In *AAAI*, 2023. 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 *ECCV*, 2018. 1

[10] Jifeng Dai, Kaiming He, and Jian Sun. Convolutional feature masking for joint object and stuff segmentation. In *CVPR*, 2015. 2

[11] Benjamin Davidson, Mohsan S Alvi, and João F Henriques. 360° camera alignment via segmentation. In *ECCV*, 2020. 1

[12] Liuyuan Deng, Ming Yang, Yeqiang Qian, Chunxiang Wang, and Bing Wang. CNN based semantic segmentation for urban traffic scenes using fisheye camera. In *IV*, 2017. 2

[13] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. CARLA: An open urban driving simulator. In *CoRL*, 2017. 6

[14] Jianshe Duan, Yachao Zhang, and Yanyun Qu. Source-free domain adaptation for point cloud semantic segmentation. In *ICME*, 2024. 3

[15] Ke Fan, Jingshi Lei, Xuelin Qian, Miaopeng Yu, Tianjun Xiao, Tong He, Zheng Zhang, and Yanwei Fu. Rethinking amodal video segmentation from learning supervised signals with object-centric representation. In *ICCV*, 2023. 2

[16] Patrick Follmann, Rebecca König, Philipp Härtinger, Michael Klostermann, and Tobias Böttger. Learning to see the invisible: End-to-end trainable amodal instance segmentation. In *WACV*, 2019. 2

[17] Xiao Fu, Shangzhan Zhang, Tianrun Chen, Yichong Lu, Xiaowei Zhou, Andreas Geiger, and Yiyi Liao. PanopticNeRF-360: Panoramic 3D-to-2D label transfer in urban scenes. *arXiv preprint arXiv:2309.10815*, 2023. 2

[18] Jianxiong Gao, Xuelin Qian, Yikai Wang, Tianjun Xiao, Tong He, Zheng Zhang, and Yanwei Fu. Coarse-to-fine amodal segmentation with shape prior. In *ICCV*, 2023. 2

[19] Shaohua Gao, Kailun Yang, Hao Shi, Kaiwei Wang, and Jian Bai. Review on panoramic imaging and its applications in scene understanding. *IEEE Transactions on Instrumentation and Measurement*, 2022. 1

[20] Vitor Campanholo Guizilini, Jie Li, Rares Ambrus, and Adrien Gaidon. Geometric unsupervised domain adaptation for semantic segmentation. In *ICCV*, 2021. 3

[21] Meng-Hao Guo, Cheng-Ze Lu, Qibin Hou, Zhengning Liu, Ming-Ming Cheng, and Shi-Min Hu. SegNeXt: Rethinking convolutional attention design for semantic segmentation. In *NeurIPS*, 2022. 1

[22] Suresh Guttikonda and Jason Rambach. Single frame semantic segmentation using multi-modal spherical images. In *WACV*, 2024. 2

[23] Bharath Hariharan, Pablo Arbeláez, Ross Girshick, and Jitendra Malik. Simultaneous detection and segmentation. In *ECCV*, 2014. 2

[24] Lukas Hoyer, Dengxin Dai, and Luc Van Gool. DAFormer: Improving network architectures and training strategies for domain-adaptive semantic segmentation. In *CVPR*, 2022. 2

[25] Lukas Hoyer, Dengxin Dai, and Luc Van Gool. HRDA: Context-aware high-resolution domain-adaptive semantic segmentation. In *ECCV*, 2022. 2

[26] Yuan-Ting Hu, Hong-Shuo Chen, Kexin Hui, Jia-Bin Huang, and Alexander G. Schwing. SAIL-VOS: Semantic amodal instance level video object segmentation - A synthetic dataset and baselines. In *CVPR*, 2019. 2

[27] Jiaxing Huang, Dayan Guan, Aoran Xiao, and Shijian Lu. Cross-view regularization for domain adaptive panoptic segmentation. In *CVPR*, 2021. 3

[28] Sujin Jang, Joohan Na, and Dokwan Oh. DaDA: Distortion-aware domain adaptation for unsupervised semantic segmentation. In *NeurIPS*, 2022. 2

[29] Alexander Jaus, Kailun Yang, and Rainer Stiefelhagen. Panoramic panoptic segmentation: Towards complete surrounding understanding via unsupervised contrastive learning. In *IV*, 2021. 3

[30] Alexander Jaus, Kailun Yang, and Rainer Stiefelhagen. Panoramic panoptic segmentation: Insights into surrounding parsing for mobile agents via unsupervised contrastive learning. *IEEE Transactions on Intelligent Transportation Systems*, 2023. 2, 3

[31] Jing Jiang, Sicheng Zhao, Jiankun Zhu, Wenbo Tang, Zhaopan Xu, Jidong Yang, Guoping Liu, Tengfei Xing, Pengfei Xu, and Hongxun Yao. Multi-source domain adaptation for panoramic semantic segmentation. *Information Fusion*, 2025. 3

[32] Xueying Jiang, Jiaxing Huang, Sheng Jin, and Shijian Lu. Domain generalization via balancing training difficulty and model capability. In *ICCV*, 2023. 2

[33] Nazmul Karim, Niluthpol Chowdhury Mithun, Abhinav Rajvanshi, Han-pang Chiu, Supun Samarasekera, and Nazanin Rahnavard. C-SFDA: A curriculum learning aided self-training framework for efficient source free domain adaptation. In *CVPR*, 2023. 3

[34] Ke Li and Jitendra Malik. Amodal instance segmentation. In *ECCV*, 2016. 2

[35] Xuewei Li, Tao Wu, Zhongang Qi, Gaoang Wang, Ying Shan, and Xi Li. SGAT4PASS: Spherical geometry-aware transformer for panoramic semantic segmentation. In *IJCAI*, 2023. 2

[36] Zhixuan Li, Weining Ye, Tingting Jiang, and Tiejun Huang. 2D amodal instance segmentation guided by 3D shape prior. In *ECCV*, 2022. 2

[37] Zhixuan Li, Weining Ye, Tingting Jiang, and Tiejun Huang. GIN: Generative invariant shape prior for amodal instance segmentation. *IEEE Transactions on Multimedia*, 2023. 2

[38] Zhixuan Li, Weining Ye, Juan Terven, Zachary Bennett, Ying Zheng, Tingting Jiang, and Tiejun Huang. MUVA: A new large-scale benchmark for multi-view amodal instance segmentation in the shopping scenario. In *ICCV*, 2023. 2

[39] Yiyi Liao, Jun Xie, and Andreas Geiger. KITTI-360: A novel dataset and benchmarks for urban scene understanding in 2D and 3D. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 6

[40] Yuang Liu, Wei Zhang, and Jun Wang. Source-free domain adaptation for semantic segmentation. In *CVPR*, 2021. 2, 3, 8

[41] Zhaochen Liu, Zhixuan Li, and Tingting Jiang. BLADE: Box-level supervised amodal segmentation through directed expansion. In *AAAI*, 2024. 2

[42] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *ICLR*, 2019. 6

[43] Zhihe Lu, Da Li, Yi-Zhe Song, Tao Xiang, and Timothy M. Hospedales. Uncertainty-aware source-free domain adaptive semantic segmentation. *IEEE Transactions on Image Processing*, 2023. 3

[44] Chaoxiang Ma, Jiaming Zhang, Kailun Yang, Alina Roitberg, and Rainer Stiefelhagen. DensePASS: Dense panoramic semantic segmentation via unsupervised domain adaptation with attention-augmented context exchange. In *ITSC*, 2021. 1

[45] Elham Amin Mansour, Ozan Unal, Suman Saha, Benjamin Bejar, and Luc Van Gool. Language-guided instance-aware domain-adaptive panoptic segmentation. *arXiv preprint arXiv:2404.03799*, 2024. 3

[46] Ivan Martinovic, Josip Saric, and Sinisa Segvic. MC-PanDA: Mask confidence for panoptic domain adaptation. In *ECCV*, 2024. 3

[47] Jieru Mei, Alex Zihao Zhu, Xinchen Yan, Hang Yan, Siyuan Qiao, Liang-Chieh Chen, and Henrik Kretzschmar. Waymo open dataset: Panoramic video panoptic segmentation. In *ECCV*, 2022. 2

[48] Rohit Mohan and Abhinav Valada. Amodal panoptic segmentation. In *CVPR*, 2022. 3, 6

[49] Rohit Mohan and Abhinav Valada. Perceiving the invisible: Proposal-free amodal panoptic segmentation. *IEEE Robotics and Automation Letters*, 2022. 3

[50] Viktor Olsson, Wilhelm Tranheden, Juliano Pinto, and Lennart Svensson. ClassMix: Segmentation-based data augmentation for semi-supervised learning. In *WACV*, 2021. 5

[51] Ege Ozguroglu, Ruoshi Liu, Dídac Surís, Dian Chen, Achal Dave, Pavel Tokmakov, and Carl Vondrick. pix2gestalt: Amodal segmentation by synthesizing wholes. In *CVPR*, 2024. 3

[52] Lorenzo Porzi, Samuel Rota Bulo, Aleksander Colovic, and Peter Kontschieder. Seamless scene segmentation. In *CVPR*, 2019. 2

[53] Lu Qi, Li Jiang, Shu Liu, Xiaoyong Shen, and Jiaya Jia. Amodal instance segmentation with KINS dataset. In *CVPR*, 2019. 2

[54] Giulia Rizzoli, Donald Shenaj, and Pietro Zanuttigh. Source-free domain adaptation for RGB-D semantic segmentation with vision transformers. In *WACV*, 2024. 3

[55] Suman Saha, Lukas Hoyer, Anton Obukhov, Dengxin Dai, and Luc Van Gool. EDAPS: Enhanced domain-adaptive panoptic segmentation. In *ICCV*, 2023. 3, 6, 7, 8

[56] Ahmed Rida Sekkat, Yohan Dupuis, Pascal Vasseur, and Paul Honeine. The omniscape dataset. In *ICRA*, 2020. 2

[57] Ahmed Rida Sekkat, Rohit Mohan, Oliver Sawade, Elmar Matthes, and Abhinav Valada. AmodalSynthDrive: A synthetic amodal perception dataset for autonomous driving. *IEEE Robotics and Automation Letters*, 2024. 6

[58] Ana Serrano, Incheol Kim, Zhili Chen, Stephen DiVerdi, Diego Gutierrez, Aaron Hertzmann, and Belen Masia. Motion parallax for 360° RGBD video. *IEEE Transactions on Visualization and Computer Graphics*, 2019. 1

[59] Hao Shi, Yu Li, Kailun Yang, Jiaming Zhang, Kunyu Peng, Alina Roitberg, Yaozu Ye, Huajian Ni, Kaiwei Wang, and Rainer Stiefelhagen. FishDreamer: Towards fisheye semantic completion via unified image outpainting and segmentation. In *CVPRW*, 2023. 2

[60] Hao Shi, Yifan Zhou, Kailun Yang, Xiaoting Yin, Ze Wang, Yaozu Ye, Zhe Yin, Shi Meng, Peng Li, and Kaiwei Wang.

PanoFlow: Learning 360° optical flow for surrounding temporal understanding. *IEEE Transactions on Intelligent Transportation Systems*, 2023. 1

[61] Yihong Sun, Adam Kortylewski, and Alan Yuille. Amodal segmentation through out-of-task and out-of-distribution generalization with a Bayesian model. In *CVPR*, 2022. 2

[62] Onur Tasar, S. L. Happy, Yuliya Tarabalka, and Pierre Alliez. ColorMapGAN: Unsupervised domain adaptation for semantic segmentation using color mapping generative adversarial networks. *IEEE Transactions on Geoscience and Remote Sensing*, 2020. 3

[63] Minh Tran, Khoa Vo, Kashu Yamazaki, Arthur Fernandes, Michael Kidd, and Ngan Le. AISFormer: Amodal instance segmentation with transformer. In *BMVC*, 2022. 2

[64] Minh Tran, Khoa Vo, Tri Nguyen, and Ngan Le. Amodal instance segmentation with diffusion shape prior estimation. In *ACCV*, 2024. 3

[65] Wilhelm Tranheden, Viktor Olsson, Juliano Pinto, and Lennart Svensson. DACS: Domain adaptation via cross-domain mixed sampling. In *WACV*, 2021. 2

[66] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 2008. 6

[67] Yan Wang, Jian Cheng, Yixin Chen, Shuai Shao, Lanyun Zhu, Zhenzhou Wu, Tao Liu, and Haogang Zhu. FVP: Fourier visual prompting for source-free unsupervised domain adaptation of medical image segmentation. *IEEE Transactions on Medical Imaging*, 2023. 3

[68] Zhonghao Wang, Mo Yu, Yunchao Wei, Rogério Feris, Jinjun Xiong, Wen-Mei Hwu, Thomas S. Huang, and Honghui Shi. Differential treatment for stuff and things: A simple unsupervised domain adaptation method for semantic segmentation. In *CVPR*, 2020. 3

[69] Yuting Xiao, Yanyu Xu, Ziming Zhong, Weixin Luo, Jiawei Li, and Shenghua Gao. Amodal segmentation based on visible region segmentation and shape prior. In *AAAI*, 2021. 2

[70] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, José M. Álvarez, and Ping Luo. SegFormer: Simple and efficient design for semantic segmentation with transformers. In *NeurIPS*, 2021. 1

[71] Hanqing Xu, Ming Yang, Liuyuan Deng, Yeqiang Qian, and Chunxiang Wang. Neutral cross-entropy loss based unsupervised domain adaptation for semantic segmentation. *IEEE Transactions on Image Processing*, 2021. 3

[72] Katherine Xu, Lingzhi Zhang, and Jianbo Shi. Amodal completion via progressive mixed context diffusion. In *CVPR*, 2024. 3

[73] Cheng-Yu Yang, Yuan-Jhe Kuo, and Chiou-Ting Hsu. Source free domain adaptation for semantic segmentation via distribution transfer and adaptive class-balanced self-training. In *ICME*, 2022. 8

[74] Kailun Yang, Xinxin Hu, Luis Miguel Bergasa, Eduardo Romera, Xiao Huang, Dongming Sun, and Kaiwei Wang. Can we PASS beyond the field of view? Panoramic annular semantic segmentation for real-world surrounding perception. In *IV*, 2019. 2

[75] Kailun Yang, Xinxin Hu, Luis Miguel Bergasa, Eduardo Romera, and Kaiwei Wang. PASS: Panoramic annular semantic segmentation. *IEEE Transactions on Intelligent Transportation Systems*, 2020. 2

[76] Kailun Yang, Xinxin Hu, and Rainer Stiefelhagen. Is context-aware CNN ready for the surroundings? Panoramic semantic segmentation in the wild. *IEEE Transactions on Image Processing*, 2021. 2

[77] Kailun Yang, Jiaming Zhang, Simon Reiß, Xinxin Hu, and Rainer Stiefelhagen. Capturing omni-range context for omnidirectional segmentation. In *CVPR*, 2021. 1

[78] Yaozu Ye, Kailun Yang, Kaite Xiang, Juan Wang, and Kaiwei Wang. Universal semantic segmentation for fisheye urban driving images. In *SMC*, 2020. 2

[79] Li Yi, Gezheng Xu, Pengcheng Xu, Jiaqi Li, Ruizhi Pu, Charles Ling, Allan Mcleod, and Boyu Wang. When source-free domain adaptation meets learning with noisy labels. In *ICLR*, 2023. 3

[80] Senthil Kumar Yogamani, Christian Witt, Hazem Rashed, Sanjaya Nayak, Saquib Mansoor, Padraig Varley, Xavier Perrotton, Derek O'Dea, Patrick Pérez, Ciarán Hughes, Jonathan Horgan, Ganesh Sistu, Sumanth Chennupati, Michal Uricár, Stefan Milz, Martin Simon, and Karl Amende. WoodScape: A multi-task, multi-camera fisheye dataset for autonomous driving. In *ICCV*, 2019. 2

[81] Zhiqi Yu, Jingjing Li, Zhekai Du, Lei Zhu, and Heng Tao Shen. A comprehensive survey on source-free domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 3

[82] Hongyi Zhang, Moustapha Cissé, Yann N. Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. In *ICLR*, 2018. 5

[83] Jiaming Zhang, Kailun Yang, and Rainer Stiefelhagen. ISSAFE: Improving semantic segmentation in accidents by fusing event-based data. In *IROS*, 2021. 2

[84] Jiaming Zhang, Chaoxiang Ma, Kailun Yang, Alina Roitberg, Kunyu Peng, and Rainer Stiefelhagen. Transfer beyond the field of view: Dense panoramic semantic segmentation via unsupervised domain adaptation. *IEEE Transactions on Intelligent Transportation Systems*, 2022. 2, 3

[85] Jiaming Zhang, Kailun Yang, Chaoxiang Ma, Simon Reiß, Kunyu Peng, and Rainer Stiefelhagen. Bending reality: Distortion-aware transformers for adapting to panoramic semantic segmentation. In *CVPR*, 2022. 2, 3, 6, 7, 8

[86] Jingyi Zhang, Jiaxing Huang, and Shijian Lu. Hierarchical mask calibration for unified domain adaptive panoptic segmentation. In *CVPR*, 2023. 6, 7, 8

[87] Jingyi Zhang, Jiaxing Huang, Xiaoqin Zhang, and Shijian Lu. UniDAformer: Unified domain adaptive panoptic segmentation transformer via hierarchical mask calibration. In *CVPR*, 2023. 3

[88] Jiaming Zhang, Kailun Yang, Hao Shi, Simon Reiß, Kunyu Peng, Chaoxiang Ma, Haodong Fu, Philip H. S. Torr, Kaiwei Wang, and Rainer Stiefelhagen. Behind every domain there is a shift: Adapting distortion-aware vision transformers for panoramic semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 3

[89] Weiming Zhang, Yexin Liu, Xu Zheng, and Lin Wang. GoodSAM: Bridging domain and capacity gaps via segment anything model for distortion-aware panoramic semantic segmentation. In *CVPR*, 2024.

[90] Weiming Zhang, Yexin Liu, Xu Zheng, and Lin Wang. GoodSAM++: Bridging domain and capacity gaps via segment anything model for panoramic semantic segmentation. *arXiv preprint arXiv:2408.09115*, 2024. 3

[91] Xingchen Zhao, Niluthpol Chowdhury Mithun, Abhinav Rajvanshi, Han-Pang Chiu, and Supun Samarasekera. Unsupervised domain adaptation for semantic segmentation with pseudo label self-refinement. In *WACV*, 2023. 3

[92] Junwei Zheng, Ruiping Liu, Yufan Chen, Kunyu Peng, Chengzhi Wu, Kailun Yang, Jiaming Zhang, and Rainer Stiefelhagen. Open panoramic segmentation. In *ECCV*, 2024. 2

[93] Xu Zheng, Tianbo Pan, Yunhao Luo, and Lin Wang. Look at the neighbor: Distortion-aware unsupervised domain adaptation for panoramic semantic segmentation. In *ICCV*, 2023. 3, 6, 7, 8

[94] Xu Zheng, Jinjing Zhu, Yexin Liu, Zidong Cao, Chong Fu, and Lin Wang. Both style and distortion matter: Dual-path unsupervised domain adaptation for panoramic semantic segmentation. In *CVPR*, 2023. 3

[95] Xu Zheng, Pengyuan Zhou, Athanasios V. Vasilakos, and Lin Wang. 360SFUDA++: Towards source-free UDA for panoramic segmentation by learning reliable category prototypes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 2, 3, 6, 7, 8

[96] Xu Zheng, Pengyuan Zhou, Athanasios V. Vasilakos, and Lin Wang. Semantics distortion and style matter: Towards source-free UDA for panoramic segmentation. In *CVPR*, 2024. 2, 3, 8

[97] Zishuo Zheng, Chunyu Lin, Lang Nie, Kang Liao, Zhijie Shen, and Yao Zhao. Complementary bi-directional feature compression for indoor 360° semantic segmentation with self-distillation. In *WACV*, 2023. 2

[98] Yan Zhu, Yuandong Tian, Dimitris Metaxas, and Piotr Dollár. Semantic amodal segmentation. In *CVPR*, 2017. 2

[99] Yang Zou, Zhiding Yu, B. V. K. Vijaya Kumar, and Jinsong Wang. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In *ECCV*, 2018. 3