

AH-OCDA: Amplitude-based Curriculum Learning and Hopfield Segmentation Model for Open Compound Domain Adaptation

Jaehyun Choi*

Junwon Ko*

Dong-Jae Lee*

Junmo Kim

School of Electrical Engineering, KAIST, South Korea

{chlwogus, kojunewon, jhtwosun, junmo.kim}@kaist.ac.kr

Abstract

Open compound domain adaptation (OCDA) is a practical domain adaptation problem that consists of a source domain, target compound domain, and unseen open domain. In this problem, the absence of domain labels and pixel-level segmentation labels for both compound and open domains poses challenges to the direct application of existing domain adaptation and generalization methods. To address this issue, we propose Amplitude-based curriculum learning and a Hopfield segmentation model for Open Compound Domain Adaptation (AH-OCDA). Our method comprises two complementary components: 1) amplitude-based curriculum learning and 2) Hopfield segmentation model. Without prior knowledge of target domains within the compound domains, amplitude-based curriculum learning gradually induces the semantic segmentation model to adapt from the near-source compound domain to the far-source compound domain by ranking unlabeled compound domain images through Fast Fourier Transform (FFT). Additionally, the Hopfield segmentation model maps segmentation feature distributions from arbitrary domains to the feature distributions of the source domain. AH-OCDA achieves state-of-the-art performance on two OCDA benchmarks and extended open domains, demonstrating its adaptability to continuously changing compound domains and unseen open domains.

1. Introduction

Semantic segmentation has been explored in various areas, including medical image processing, autonomous driving, and robot scene understanding. However, the need for meticulous pixel-level labels in training presents a significant challenge for direct model deployment. To mitigate the burden of acquiring such labels, domain adaptation and domain generalization have been investigated. These tasks

primarily focus on minimizing the distribution difference between the source and target domains, also known as the domain gap. Although they have demonstrated remarkable performance, each approach has certain limitations: domain adaptation experiences performance degradation in novel target domains, while domain generalization requires obtaining pixel-level labels across multiple domains. This challenge has driven the development of approaches such as multi-target domain adaptation, single-source domain generalization, and recently, open compound domain adaptation (OCDA).

OCDA, first proposed by *Liu et al.* [18], leverages the advantages of aforementioned tasks and adapts to multiple unlabeled target domains (a compound domain) and novel domains (an open domain) with a single labeled source domain. Specifically, OCDA treats the compound domain as one continuous spectrum of potential target domains, enhancing its practicality even when the domain labels within the real-world target domains are vague. Additionally, by addressing both the compound and open domains simultaneously, OCDA further enhances its practicality, aligning with the ideal scenario in which semantic segmentation models are trained on labeled synthetic source domains but deployed in environments where domains are unknown. Previous works [8, 18, 21] solve this complicated problem by dividing the compound domain into multiple discrete domains, narrowing the task into several unsupervised domain adaptation problems. While these approaches have demonstrated promising performance, they are limited by training separate segmentation models, requiring sophisticated hyperparameter tuning for each domain and model while also not fully accounting for the open domain.

To overcome the limitations, we propose a novel method, amplitude-based curriculum learning with the Hopfield segmentation model for OCDA (AH-OCDA). AH-OCDA integrates two key components: amplitude-based curriculum learning and the Hopfield segmentation model. The amplitude-based curriculum learning is a ranking-based approach that leverages the frequency spectrum of each im-

*Equal contribution.

age to treat the compound domain as a continuous spectrum of target domains. Moreover, the Hopfield segmentation model maps distorted features to source features using a modern continuous Hopfield network, aiming to overcome the challenge of unseen open domains. It is noteworthy that these two components are orthogonal, complementing each other to achieve better performance in OCDA.

Amplitude-based curriculum learning leverages domain-specific information of the image, drawing upon the insight that amplitude in the low-frequency spectrum contains the domain-specific information [35, 38]. Specifically, we extract the low-frequency component from the amplitude obtained by the Fast Fourier Transform (FFT) of the target image and compute mean squared error (MSE) with the mean amplitude of source domain images. We continuously adapt the model trained in the source domain to the near-source target domain, while utilizing the previously adapted target domain as *fake source* with pseudo labels. The proposed curriculum learning only requires knowledge of the individual gap between target images and the source domain, maintaining robustness against dynamic domain shifts. Meanwhile, the Hopfield segmentation model maps segmentation feature distributions from arbitrary domains to the feature distributions of the source domain by learning and retrieving feature patterns. Modern continuous Hopfield network [19, 25] learns relational feature mappings, enabling it to restore corrupted or noisy patterns to their original pattern by leveraging associative memory. Building on this capability, the Hopfield segmentation model implicitly stores representative information within its parameters, including intra-class and inter-class information. This allows the network to accurately predict distorted feature vectors from compound or open domains previously encountered. Unlike approaches that rely on manually modeling feature distribution statistics, such as memorizing the class centroids or statistical information, our Hopfield segmentation model requires only the source domain to learn raw patterns and map the target domain features to their corresponding source domain features.

While the Hopfield segmentation model is capable of mapping arbitrary domain features to source domain features, it cannot guarantee the successful mapping of features from significantly distant domains to those in the known domain. At this point, amplitude-based curriculum learning trains the Hopfield segmentation model gradually, starting from the near-source target domain to the far-source target domain, enabling the model to map target domain features to the source domain features. By gradually adapting to target domains through adversarial learning and using *fake source*, the network can fully leverage the memory retrieval process to progressively enlarge the boundary of the known domain. Conversely, the Hopfield network within the segmentation model enables the model to pre-

dict domains beyond the known boundary established by curriculum learning by restoring distorted features to the corresponding source features. Our experiments confirmed that these components function orthogonally, significantly enhancing the model’s performance when combined. Furthermore, we validated effectiveness of AH-OCDA on two OCDA benchmarks and additional extended open domains, where we achieved state-of-the-art performance.

2. Related works

2.1. Domain Adaptation and Generalization

Domain adaptation (DA) is a key technique in machine learning where a model trained on one domain is adapted to perform well on a different, but related domain. Unsupervised domain adaptation (UDA) is a specific approach in which the model is adapted to an unlabeled target domain. They have advanced to scenarios with single target domain [22, 28, 31] or clearly distinct multi-target domains [6, 7, 40] and also to scenario with multiple source domain [23, 36]. In parallel, the concept of domain generalization (DG) aims to train a model on multiple source domains to be generalized on unseen target domains [1, 2, 4, 12, 35] or train a model on single source domains to be generalized on unseen target domains [15, 24]. However, these methods may not always capture the full complexity of real-world scenarios. In practical applications, data collection often results in mixed, continuous, and unseen domains, highlighting the need for new problem settings that better reflect these diverse and dynamic real-world conditions.

2.2. Open Compound Domain Adaptation

Open compound domain adaptation (OCDA) [18] suggested a problem setting that has a labeled source domain and a compound of multiple unlabeled target domains for training phase and unseen target domains for testing phase. They proposed a curriculum learning framework for semantic segmentation tasks, which schedules the curriculum based on the prediction confidence of a model that is only trained on the source domain. On the other hand, the following studies did not follow OCDA’s approach but proposed their own unique methods. DHA [21] and MOCDA [8] assigned pseudo domain labels with K-means clustering to break down OCDA into multiple UDA settings. Specifically, DHA [21] utilizes K-means clustering on the convolutional feature statistics, which are mean and standard deviations. They stylize the source domain into each target latent domain to reduce the domain gap and adversarially align a shared segmentation model employing individual discriminators for each pair of stylized source and target latent domains. MOCDA [8] utilizes K-means clustering on style code, which is low-dimensional space generated by an additional encoder. They adopt a model-agnostic

meta-learning approach, where a hyper-network combines domain-specific segmentation models and discriminators to consider the continuous compound target domain. However, K-means clustering may not be a suitable approach for open compound domain adaptation, as it may not effectively represent continuous target domains and accommodate expandable unseen domains. To address this limitation, we explore a more dynamic approach.

2.3. Fourier Transform for Domain Adaptation

Several works have discussed how the Fourier transform represents the style information of an image in its magnitude spectrum and have applied this concept to domain adaptation [11, 16, 29, 33, 37, 38, 42]. In this perspective, FDA [38] proposed a Fast Fourier Transform (FFT)-based style transfer that reduces the style gap between two domains by replacing the low-frequency spectrum of the images. Building on FDA, some methods have further developed FFT-based style transfer for DA by incorporating the ordering of target data during the training phase, similar to approaches used in continual learning [29, 33]. BAT [42] further applied FFT-based style transfer for multi-target domain adaptation instead of network-based style transfer. To cope better with OCDA task, AST-OCDA [13] suggested applying the FT-based style transfer on the feature space of the segmentation model instead of the input RGB image. For single-domain or discrete multiple-domain adaptation, applying Fourier-based style information at a particular feature level has proven effective. However, in OCDA, the direct application of Fourier-based style transfer is less feasible for adapting the source domain to diverse, continuous target domains, and unseen domains. To overcome this, we considered the flexible utilization of domain-specific information in OCDA scenarios.

2.4. Associative Memory

Associative memory systems function by retrieving data not through specific addresses but via queries that resemble the type of data stored within them [9]. When a query is made, the system identifies and returns a data point that most closely matches the query. For example, presenting an image as a query would yield other images that are deemed similar. This retrieval process is often compared to how the human brain operates, where memories can be recalled with partial cues, such as remembering a song from a few notes [19]. Two classical and influential models are the Hopfield network (HN) and the sparse distributed memory (SDM). More recently, they have been generalized to the modern continuous Hopfield network (MCHN), which substantially improves performance and can handle continuous inputs. Specifically, MCHN uses the energy function $E = q^\top q + \log \sum \exp(Wq)$, which can be minimized using the convex-concave method, giving the up-

date rule $z = W^\top \sigma(Wq)$, where $\sigma(\cdot)$, W , q represent the softmax function, the memory matrix and the query vector, respectively. This is similar to the attention operation $z = V\sigma(Wq)$ with key matrix K and value matrix V , which is widely used in recent neural networks [10, 14, 34], where we can associate $Q = q$, $K = W$, and $V = W$ [25].

3. Method

Our approach for OCDA in semantic segmentation consists of two components: 1) amplitude-based curriculum learning and 2) the Hopfield segmentation model. Figure 1 illustrates our whole pipeline.

3.1. Preliminaries

Let spatial domain space $\mathcal{X} \subset \mathbb{R}^{H \times W \times 3}$ and frequency domain space $\mathcal{A} \subset \mathbb{C}^{H \times W \times 3}$, where H and W are height and width of the input image, respectively. Through FFT $\mathcal{F} : \mathcal{X} \rightarrow \mathcal{A}$, we transform the image $x \in \mathcal{X}$ to the spectrum $\mathcal{F}(x)$. Then, we obtain amplitude $|\mathcal{F}(x)| \in \mathbb{R}^{H \times W \times 3}$ and phase $\text{Arg}(\mathcal{F}(x)) \in \mathbb{R}^{H \times W \times 3}$. The semantic segmentation model Seg consists of an encoder f and classifier g , such that $Seg : x \mapsto y = g(f(x))$, where $y \in \mathbb{R}^{H \times W}$ represents semantic segmentation prediction.

Following OCDA, the labeled source domain consists of synthetic RGB images x_s and corresponding semantic segmentation label y_s , such that

$$\mathcal{S} = \{(x_s^i, y_s^i) \mid x_s^i \in \mathcal{X}, y_s^i\}, i = 1, \dots, N_s \quad (1)$$

where N_s is the number of images in the labeled source domain. The unlabeled compound domain consists of real RGB input images x_c , such that

$$\mathcal{C} = \{x_c^i \mid x_c^i \in \mathcal{X}\}, i = 1, \dots, N_c \quad (2)$$

where N_c is the number of images in the unlabeled compound domain. The unseen domain consists of real RGB images x_u , such that

$$\mathcal{U} = \{x_u^i \mid x_u^i \in \mathcal{X}\}, i = 1, \dots, N_u \quad (3)$$

where N_u is the number of images in the unseen domain.

3.2. Amplitude-based Curriculum Learning

The amplitude-based curriculum learning is a ranking-based approach that only requires knowledge of the individual gap between the target image and the source domain images. First, we transform the image x_i to the spectrum through the FFT. Then we shift values making the center of the spectrum to represent the low-frequency region. Finally, we obtain the domain-specific information by utilizing only the amplitude from the spectrum and cropping the central region, a region that is known to contain the domain-specific information.

$$F_\beta^i = \text{Crop}_\beta \{|\mathcal{F}(x_i)|\} \in \mathbb{R}^{\beta H \times \beta W \times 3} \quad (4)$$

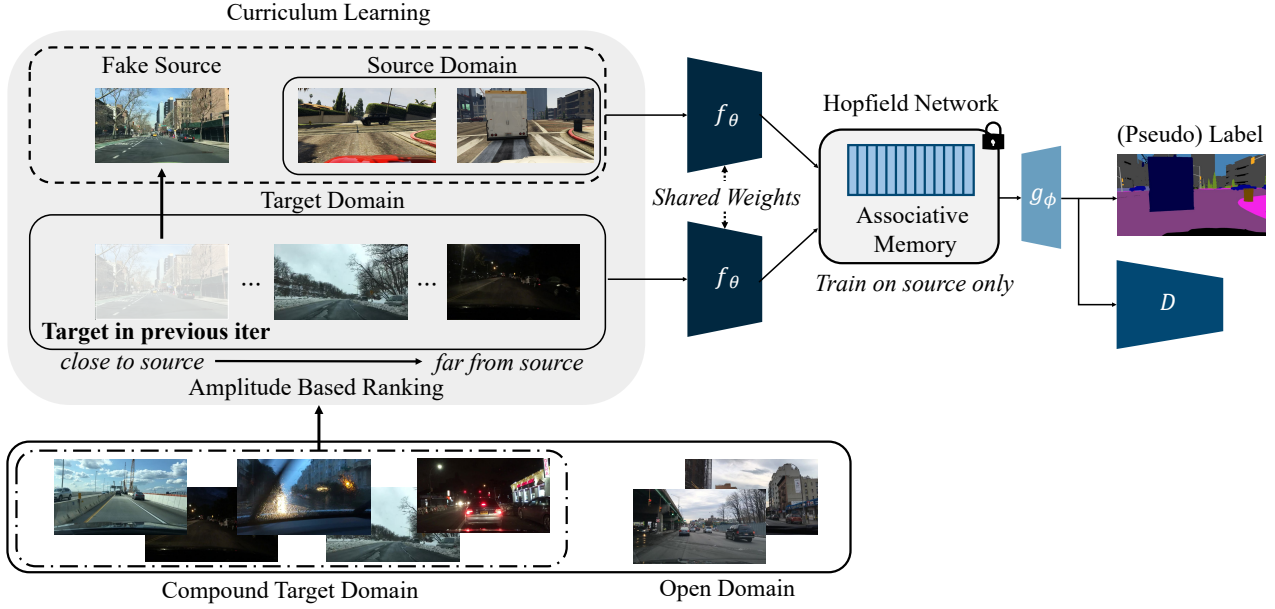


Figure 1. Overview of the proposed pipeline. For the amplitude-based curriculum learning, we extract the amplitudes with the Fast Fourier Transform and measure the distance between each target image and the source domain images. The Hopfield segmentation model is frozen when training.

where β is the cropping ratio of height and width in the spectrum. The average domain information of the source domain is defined as below.

$$\mathcal{M}_s = \frac{1}{N_s} \sum_{x_s \in \mathcal{S}} \text{Crop}_\beta \{|\mathcal{F}(x_s)|\} \quad (5)$$

Once \mathcal{M}_s is obtained, we can define the domain distance δ of the compound image x_c like below.

$$\delta_c = (\mathcal{M}_s - F_\beta^c)^2 \quad (6)$$

We rank $x_c \in \mathcal{C}$ to define curriculum G based on their δ_c , arranging them in ascending order.

$$G = \{x_c \mid x_c \text{ is sorted in ascending order for } \delta_c\} \quad (7)$$

For the adversarial alignment, we divide G into K groups where K is a factor of size of G .

$$G = \bigcup_{j=1}^K \mathcal{T}_j \quad (8)$$

where each \mathcal{T}_j contains N_c/K number of x_c . The learning process gradually includes \mathcal{T}_j to perform near-source to far-source curriculum learning.

The segmentation model gets x_s and x_c from the curriculum and predicts the softmax output \hat{y}_s and \hat{y}_c . Then, the discriminator takes \hat{y}_s and \hat{y}_c to differentiate whether

the input originates from the source or the target domain. In this process, adversarial learning induces the segmentation model to be unable to distinguish between source and target domain images. However, naively applying curriculum learning has a limitation as the previously leveraged target images \mathcal{T}_{j-1} will no longer be utilized for training and thus may result in catastrophic forgetting. For example, when the pipeline is on \mathcal{T}_j , the previous x_c in \mathcal{T}_{j-1} is no longer being used for training. In this case, the performance on x_c will get deteriorated reducing overall domain adaptation performance on \mathcal{C} . To address this issue, our framework assigns half of previously adapted target domain images as a fake source with the pseudo-label from prediction along with source domain images as $\hat{\mathcal{S}}_j$ where j is the step in which curriculum learning is currently in.

$$\hat{\mathcal{S}}_j = \mathcal{S} \cup \mathcal{P}_j \quad (9)$$

where $\mathcal{P}_j = \{x_c^1, x_c^2, \dots, x_c^{\lfloor \frac{|G| \times (j-1)}{2K} \rfloor}\}$. The fake-source approach ensures that the framework continuously adapts with the combined set of \mathcal{S} and the fraction of x_c .

3.3. Hopfield Segmentation Model

In contrast to traditional domain generalization scenarios, where multiple labeled domains are typically available, OCDA contains labels only in the source domain, while both the compound and open domains do not have segmentation and domain labels during training. This challenge

results in suboptimal performance of previous methods on target domains that significantly deviate from the source domain, as well as on unseen open domains. To address this challenge, we devised the Hopfield segmentation model to empower the semantic segmentation pipeline with the capability to align features in the target domain with their corresponding source domain, even on the novel unseen domains. Intuitively, adaptation for compound and open domains can be easily solved by mapping the arbitrary target features to the source feature distribution. The Hopfield network in the segmentation model correct distorted feature patterns or noisy inputs from unfamiliar domains by retrieving stored source patterns. In other words, given the encoder features z , our objective is to learn the mapping function $H(\cdot)$ that maps the arbitrary feature to the source feature distribution.

Mathematically, the associative memory is a function $g(\cdot) : \mathbb{R}^{C_i} \rightarrow \mathbb{R}^{C_o}$ mapping a vector in an input space of C_i to a vector in an output space of dimension C_o , with two additional inputs of a memory matrix $M \in \mathbb{R}^{M_N \times C_i}$ consisting of a set of M_N stored patterns, and projection matrices $P \in \mathbb{R}^{C_i \times C_o}$, consisting of a potentially different set of stored patterns with dimension C_o for hetero-association. Similarly, our Hopfield network consists of a memory matrix $M \in \mathbb{R}^{M_N \times C_l}$ and query, key, and value projection matrices $W_q, W_k \in \mathbb{R}^{C_l \times C_s}$ and $W_v \in \mathbb{R}^{C_s \times C_s}$, where M_N, C_l, C_s represents the memory size, dimension of the feature vector, and dimension of the projected feature vector for similarity measurement, respectively.

We first measure the similarity of the feature embedding z and each raw stored pattern vector $m_i \in M_N$:

$$\text{sim}(z, m_i) = \frac{\exp(\tau \cdot W_q z \cdot (W_k m_i)^\top)}{\sum_{j \in M_N} \exp(\tau \cdot W_q z \cdot (W_k m_j)^\top)}. \quad (10)$$

where τ represents the scaling factor. The role of each projection matrix is to project the raw state patterns z and stored patterns m_j to associative space and reduce computational complexity through dimension reduction ($C_s < C_l$). Then, we project the z to the stored pattern with the weighted sum of the similarity and the stored patterns:

$$\hat{z} = \sum_{j \in M_N} \text{sim}(z, m_j) W_v m_j. \quad (11)$$

The projected feature \hat{z} is fed to the classifier to predict the segmentation map and update its memory with the segmentation loss.

For training the Hopfield segmentation model, we pre-train the Hopfield network in the segmentation model for the source dataset. Then, in the following adaptation phase, we freeze the associative memory. Specifically, we freeze the memory weights with key and value projection weights. Therefore, the memory network maps the feature from tar-

get domains to the trained source domain feature distribution. The pretrained Hopfield network in the segmentation model is likely to be robust even on far-source target domains, as the amplitude-based curriculum learning gradually trains the entire segmentation model to obtain a well-structured target feature distribution.

3.4. Loss Function

Our framework utilizes the following loss to optimize the segmentation model Seg and the discriminator D :

$$\mathcal{L}(\hat{S}_j, \mathcal{T}_j) = \mathcal{L}_{ce}^{Seg}(\hat{S}_j) + \lambda_{adv}(\mathcal{L}_{adv}^{Seg}(\mathcal{T}_j) + \mathcal{L}_{adv}^D(\hat{S}_j, \mathcal{T}_j)). \quad (12)$$

This loss consists of two parts: \mathcal{L}_{ce}^S is the cross entropy loss that updates the segmentation model to make accurate predictions in the source domain and the *fake-source*. Formally, \mathcal{L}_{ce}^S is defined as

$$\mathcal{L}_{ce}^{Seg}(\hat{S}_j) = - \sum_{h,w} \sum_{cls} Y_{\hat{S}}^{(h,w,cls)} \log(Seg(\hat{S}_j)^{(h,w,cls)}), \quad (13)$$

where cls is the number of classes, and $Y_{\hat{S}}$ is the (pseudo) label of the (fake-) source image. \mathcal{L}_{adv} is applied separately to the segmentation model \mathcal{L}_{adv}^S and the discriminator \mathcal{L}_{adv}^D , which are

$$\mathcal{L}_{adv}^{Seg}(\mathcal{T}_j) = - \sum_{h,w} \log(D(Seg(\mathcal{T}_j))^{(h,w,1)}), \quad (14)$$

$$\mathcal{L}_{adv}^D(\mathcal{T}_j, \hat{S}_j) = - \sum_{h,w} \log(D(Seg(\mathcal{T}_j))^{(h,w,0)}) + \log(D(Seg(\hat{S}_j))^{(h,w,1)}). \quad (15)$$

\mathcal{L}_{adv} induces the adversarial learning for the segmentation model and discriminator. λ_{adv} adjusts the importance between the two losses.

4. Experiments

4.1. Experimental settings

For the labeled source domain we leverage GTA5 [26] and SYNTHIA [27] while BDD100K (C-Driving) [39] is utilized for the unlabeled compound and open domains. For extended open domain, we evaluate with Cityscapes [3] and KITTI [5]. The details for each dataset can be found in the supplementary material. For all experiments, We used SGD with a polynomial decaying learning rate, starting at 0.00025 for the segmentation model and 0.0001 for the discriminator while others such as batch size follow [18] Hyperparameters introduced for AH-OCDA such as λ , β , K , τ , and the Hopfield segmentation model memory size are set to 0.001, 0.09, 3, 1, and 64, respectively.

Table 1. Experiment results on GTA5 \rightarrow C-Driving benchmarks in mean IoU. Methods in the middle and lower sections list DA and OCDA methods in semantic segmentation, respectively.

Method	Compound (C)			Open (O)	Total	
	Rainy	Snowy	Cloudy	Overcast	C	C+O
Source Only	16.2	18.0	20.9	21.2	18.9	19.1
AdaptSegNet [30]	20.2	21.2	23.8	25.1	22.1	22.5
CBST [43]	21.3	20.6	23.9	24.7	22.2	22.6
IBN-Net [20]	20.6	21.9	26.1	25.5	22.8	23.5
PyCDA [17]	21.7	22.3	25.9	25.4	23.3	23.8
OCDA [18]	22.0	22.9	27.0	27.9	24.5	25.0
DHA [21]	27.0	26.3	30.7	32.8	28.5	29.2
MOCDA [8]	24.4	27.5	30.1	31.4	27.7	29.4
AST-OCDA [13]	28.2	27.8	31.6	34.0	29.2	30.4
AH-OCDA (Ours)	28.1	28.1	32.0	33.6	29.7	31.4

Table 2. Experiment results on GTA5 \rightarrow extended open domains, Cityscape and KITTI benchmarks, reported as mean IoU. “w/o Curr” and “w/o Hopf” denotes AH-OCDA with the Hopfield segmentation model and amplitude-based curriculum learning, respectively.

Method	Extended Open		Total
	Citys.	KITTI	
Source only	19.3	24.1	21.7
AdaptSegNet [30]	22.0	23.4	22.7
MOCDA [8]	31.1	30.9	31.0
AST-OCDA [13]	32.6	31.8	32.2
w/o Curr	30.2	32.7	31.5
w/o Hopf	28.3	31.0	29.7
AH-OCDA (Ours)	32.6	34.0	33.3

4.2. Baselines

The efficacy of our framework is assessed through a comprehensive comparison with a selection of established baselines. **Source Only** refers to a VGG segmentation model trained on the source domain. **UDA methods** are unsupervised single-source to single-target domain adaptation frameworks. These include AdaptSegNet [30], CBST [43], IBN-Net [20], PyCDA [17], CRST [44] and AdvEnt [32]. For **OCDA frameworks**, OCDA [18] is the baseline framework for open compound domain adaptation. DHA [21] and MOCDA [8] leverage clustering to decompose OCDA into multiple UDA. AST-OCDA [13] applies FFT-based style transfer on the feature space of the segmentation model.

4.3. Results

GTA5 as Source The semantic segmentation performance on GTA5 \rightarrow C-Driving is reported in Table 1. Com-

Table 3. Experiment results on SYNTHIA \rightarrow C-Driving benchmarks in mean IoU. Methods in the middle and lower sections list DA and OCDA methods in semantic segmentation, respectively.

Method	Compound (C)			Open (O)	Total	
	Rainy	Snowy	Cloudy	Overcast	C	C+O
Source only	16.3	18.8	19.4	19.5	18.4	18.5
CBST [43]	16.2	19.6	20.1	20.3	18.9	19.1
CRST [44]	16.3	19.9	20.3	20.5	19.1	19.3
AdaptSegNet [30]	17.0	20.5	21.6	21.6	20.0	20.2
AdvEnt [32]	17.7	19.9	20.2	20.5	19.3	19.6
DHA [21]	18.8	21.2	23.6	23.6	21.5	21.8
AST-OCDA [13]	20.3	22.6	24.9	25.4	22.6	23.3
AH-OCDA (Ours)	20.7	22.9	25.8	26.2	23.6	24.8

Table 4. Experiment results on SYNTHIA \rightarrow extended open domains, Cityscape and KITTI benchmarks, reported as mean IoU.

Method	Extended Open		Total
	Citys.	KITTI	
Source only	24.7	20.7	22.7
AdaptSegNet [30]	35.9	24.7	30.3
MOCDA [8]	32.2	34.2	33.2
AST-OCDA [13]	37.2	35.7	36.5
AH-OCDA (Ours)	36.8	37.2	37.0

paring performance on both Compound and Compound + Open, hereinafter denoted as C and C + O respectively, AH-OCDA achieves the best mIoU. In Table 2, we report the semantic segmentation performance on GTA5 \rightarrow extended open domains, Cityscape and KITTI. AH-OCDA outperforms the SOTA method AST-OCDA [13] by 1.1%p on total mIoU, which verifies that it has better domain generalization ability. It is because the Hopfield segmentation model helps to predict segmentation features based on source information leading the performance improvement on the extended open domains.

SYNTHIA as Source The semantic segmentation performance on SYNTHIA \rightarrow C-Driving is shown in Table 3. AH-OCDA achieves the best semantic segmentation performance among all the baselines on all domains. In comparison with AST-OCDA, AH-OCDA outperforms 1.0 %p on C and 1.5 %p on C + O, outperforming on all datasets, including one of the extended open dataset (Table 4).

Table 5. Ablation study results by varying hyperparameter K on GTA5 \rightarrow C-Driving benchmark, reported in mean IoU. Following DHA, compound domains are further divided for “Night”.

Method	Compound (C)					Open (O)		Total
	Rainy	Snowy	Cloudy	Night	Overcast	C	C+O	
Source only	23.3	24.0	28.2	8.1	30.2	25.7	26.4	
DHA($K=2$)	26.0	26.6	32.4	11.1	33.6	29.3 (-0.5)	29.7 (-0.7)	
DHA($K=3$)	26.4	27.5	33.3	11.8	34.3	29.8	30.4	
DHA($K=4$)	25.2	26.4	32.7	12.1	33.8	29.1 (-0.7)	29.5 (-0.9)	
DHA($K=5$)	25.4	27.0	32.5	13.3	33.1	29.2 (-0.6)	29.5 (-0.9)	
Ours($K=2$)	30.6	28.4	31.7	18.4	33.1	29.8 (-0.4)	31.3 (-0.7)	
Ours($K=3$)	30.1	28.6	32.4	18.5	34.0	30.1	32.0	
Ours($K=4$)	29.9	28.5	32.5	19.8	34.1	29.7 (-0.1)	31.5 (+0.0)	
Ours($K=5$)	29.2	28.4	31.9	20.0	33.6	29.7 (-0.5)	31.5 (-0.5)	

4.4. Ablation

4.4.1 Number of splits K in curriculum learning

The hyperparameter K in AH-OCDA determines the number of splits applied to the compound domain to gradually expand known distributions during curriculum learning. Similarly, DHA [21] employs a comparable hyperparameter to divide compound domains into sub-latent domains through k-means clustering. To evaluate the robustness with K , we compare the hyperparameter K used in AH-OCDA and DHA and report the results in Table 5. In both methods, the best mIoU is achieved when the hyperparameter is set to 3. However, when considering the robustness to K , a critical aspect that must be considered since the number of domains within a compound domain is unknown in advance, AH-OCDA demonstrates a much more stable performance compared to DHA. Specifically, AH-OCDA shows no performance drop when K is set to 4 and exhibits less or comparable performance drop for other values of K .

4.4.2 Sub-domains visualization

Figure 3 demonstrates randomly selected images from each curriculum domain when K is set to 3. Through amplitude-based ranking, the night or darker images are ranked later in the curriculum learning. We also visualize the curriculum domains in the feature space by extracting features using VGG-16, the segmentation network used in all experiments. Figure 2 validates that curriculum learning ranks the images within a compound domain as intended, ranking the images from near-source to far-source.

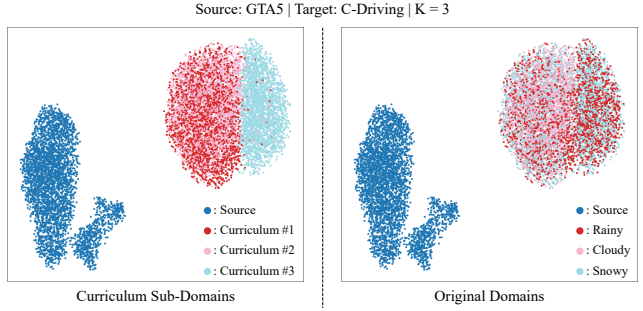


Figure 2. UMAP results of GTA5 and C-Driving when $K = 3$.



Figure 3. Sample images from each curriculum when $K = 3$.

4.4.3 Hopfield segmentation model settings

We conduct an ablation study on the hyperparameters for the Hopfield segmentation model, memory size, and training strategy and report it in Table 6. Experimentally, the size of memory does not have a crucial effect on segmentation performance when above 64. As the memory size of 64 provides the best trade-off between performance and computational complexity, we used 64 for all experiments as aforementioned. For the training strategy, we experiment with not freezing and freezing the Hopfield segmentation model when training on the compound domain. The performance drops when the memory is not frozen as the memory gradually forgets the segmentation feature of the source domain thus we leverage the frozen model.

Table 6. Ablation study results by varying hyperparameters of the Hopfield segmentation model on GTA5 \rightarrow C-Driving benchmarks. M_N represents memory size, and ‘w/o Freeze’ and ‘Freeze’ denote training and freezing the memory on target data, respectively.

Method	Compound	Open	Total
$M_N = 32$	28.7	32.0	30.2
$M_N = 64$	29.5	33.3	31.2
$M_N = 96$	29.5	33.2	31.1
w/o Freeze	21.1	23.4	22.1
Freeze	29.5	33.3	31.2

Table 7. Ablation study results of each component in AH-OCDA on GTA5 \rightarrow C-Driving benchmarks. “w/o Curr” denotes AH-OCDA with Hopfield segmentation model and “w/o Hop” denotes AH-OCDA with amplitude-based curriculum learning.

Method	Compound (C)				Open (O)		Total	
	Rainy	Snowy	Cloudy	Night	Overcast	C	C+O	
Source Only	23.3	24.0	28.2	8.1	30.2	25.7	26.4	
w/o Curr	26.7	25.3	29.4	12.2	30.5	26.9 (+1.2)	28.5 (+2.1)	
w/o Hopf	25.7	27.8	29.8	15.5	31.4	27.7 (+2.0)	29.4 (+3.0)	
AH-OCDA	30.1	28.6	32.4	18.5	34.0	30.2 (+4.5)	32.0 (+5.6)	

4.4.4 Complementary components of AH-OCDA

Table 7 shows the effect of each component in AH-OCDA. AH-OCDA outperforms ‘w/o Hopf’ (Ours without the Hopfield segmentation model) and ‘w/o Curr’ (Ours without amplitude-based curriculum learning). In Table 7, there is a significant performance drop in the ‘Night’ category for ‘w/o Curr’ model aligning with the statement in Sec. 3.3 that the Hopfield segmentation model may not function well when adapting to a far-source target domain in the compound domain. The amplitude-based curriculum learning complements this issue by gradually training the model from the near-source target domain so that the Hopfield segmentation model can be trained stably.

4.4.5 Other backbone networks

To validate that AH-OCDA is not limited to specific backbone networks, we additionally experiment using ResNet and Transformer-based backbone and compare it with unsupervised domain adaptation methods in semantic segmentation, ProDA [41] (ResNet-101) and DAFormer [10] (Segformer [34]), respectively. Table 8 shows the performance of VGG-16, ResNet-101, and Transformer-based architec-

Table 8. Ablation study results on GTA5 \rightarrow C-Driving benchmarks, varying the backbone network from VGG-16, ResNet-101, and Transformer-based (top to bottom). Each backbone network is compared with the well-known DA methods in semantic segmentation using the corresponding backbone network.

Method	Compound (C)			Open (O)	Total	
	Rainy	Snowy	Cloudy	Overcast	C	C+O
AdaptSegNet	20.2	21.2	23.8	25.1	22.1	22.5
AH-OCDA	28.1	28.1	32.0	33.6	29.7	31.4
ProDA	31.61	30.32	36.29	34.63	33.36	33.93
AH-OCDA	37.07	31.82	30.74	40.68	33.74	36.70
DAFormer	34.04	31.39	31.14	34.60	32.10	33.18
AH-OCDA	31.43	28.85	39.02	37.88	34.42	35.90

tures in the top, middle, and bottom rows, respectively. Although the performances cannot be directly compared due to the comparison methods being proposed for another task, as aforementioned, the performance gain indirectly demonstrates that AH-OCDA is not restricted to certain backbone networks. Notably, AH-OCDA shows better performance on a few of the target domains in the compound domain than the unsupervised domain adaptation methods, demonstrating the need for separate methods for OCDA setting.

5. Conclusion

In this paper, we propose a novel method, AH-OCDA, considering the differences between OCDA with conventional domain adaptation and generalization settings. Specifically, we propose two complementary components: 1) amplitude-based curriculum learning and 2) the Hopfield segmentation model. First, by utilizing the FFT, we obtain the amplitude of each target image and measure the difference between the mean amplitude of the source domain images. With the obtained distance, we rank them to formulate the OCDA training in a curriculum-learning manner, without any prior knowledge of domains in the compound domain. Second, the Hopfield segmentation model learns the segmentation feature distribution on the source domain with memory and aligns features from unknown domains to the source domain. We verify the effectiveness of the AH-OCDA with experiments on two datasets, GTA5 \rightarrow C-Driving and SYNTHIA \rightarrow C-Driving. Additionally, we present the experiment results on two novel unseen domains, Cityscape and KITTI, and achieve comparable or state-of-the-art performance.

Acknowledgements. This work was supported by Artificial intelligence industrial convergence cluster development project funded by the Ministry of Science and ICT(MSIT, Korea) & Gwangju Metropolitan City.

References

- [1] Blanchard, G., Deshmukh, A.A., Dogan, Ü., Lee, G., Scott, C.: Domain generalization by marginal transfer learning. *The Journal of Machine Learning Research* **22**(1), 46–100 (2021) [2](#)
- [2] Choi, J., Seong, H.S., Park, S., Heo, J.P.: Tcx: Texture and channel swappings for domain generalization. *Pattern Recognition Letters* **175**, 74–80 (2023) [2](#)
- [3] Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B.: The cityscapes dataset for semantic urban scene understanding. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 3213–3223 (2016) [5](#)
- [4] Dou, Q., Coelho de Castro, D., Kamnitsas, K., Glocker, B.: Domain generalization via model-agnostic learning of semantic features. *Advances in neural information processing systems* **32** (2019) [2](#)
- [5] Geiger, A., Lenz, P., Stiller, C., Urtasun, R.: Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research* **32**(11), 1231–1237 (2013) [5](#)
- [6] Gholami, B., Sahu, P., Rudovic, O., Bousmalis, K., Pavlovic, V.: Unsupervised multi-target domain adaptation: An information theoretic approach. *IEEE Transactions on Image Processing* **29**, 3993–4002 (2020) [2](#)
- [7] Gong, B., Grauman, K., Sha, F.: Reshaping visual datasets for domain adaptation. *Advances in Neural Information Processing Systems* **26** (2013) [2](#)
- [8] Gong, R., Chen, Y., Paudel, D.P., Li, Y., Chhatkuli, A., Li, W., Dai, D., Van Gool, L.: Cluster, split, fuse, and update: Meta-learning for open compound domain adaptive semantic segmentation. In: *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*. pp. 8344–8354 (2021) [1](#), [2](#), [6](#)
- [9] Hopfield, J.J.: Hopfield network. *Scholarpedia* **2**(5), 1977 (2007) [3](#)
- [10] Hoyer, L., Dai, D., Van Gool, L.: Daformer: Improving network architectures and training strategies for domain-adaptive semantic segmentation. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 9924–9935 (2022) [3](#), [8](#)
- [11] Huang, J., Guan, D., Xiao, A., Lu, S.: Rda: Robust domain adaptation via fourier adversarial attacking. In: *Proceedings of the IEEE/CVF international conference on computer vision*. pp. 8988–8999 (2021) [3](#)
- [12] Huang, Z., Wang, H., Xing, E.P., Huang, D.: Self-challenging improves cross-domain generalization. In: *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II* 16. pp. 124–140. Springer (2020) [2](#)
- [13] Kundu, J.N., Kulkarni, A.R., Bhambri, S., Jampani, V., Radhakrishnan, V.B.: Amplitude spectrum transformation for open compound domain adaptive semantic segmentation. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 36, pp. 1220–1227 (2022) [3](#), [6](#)
- [14] Lee, D.J., Lee, J.Y., Shon, H., Yi, E., Park, Y.H., Cho, S.S., Kim, J.: Lightweight monocular depth estimation via token-sharing transformer. In: *2023 IEEE International Conference on Robotics and Automation (ICRA)*. pp. 4895–4901. IEEE (2023) [3](#)
- [15] Li, L., Gao, K., Cao, J., Huang, Z., Weng, Y., Mi, X., Yu, Z., Li, X., Xia, B.: Progressive domain expansion network for single domain generalization. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 224–233 (2021) [2](#)
- [16] Li, T.B., Su, Y.T., Song, D., Li, W.H., Wei, Z.Q., Liu, A.A.: Progressive fourier adversarial domain adaptation for object classification and retrieval. *IEEE Transactions on Multimedia* (2023) [3](#)
- [17] Lian, Q., Lv, F., Duan, L., Gong, B.: Constructing self-motivated pyramid curriculums for cross-domain semantic segmentation: A non-adversarial approach. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 6758–6767 (2019) [6](#)
- [18] Liu, Z., Miao, Z., Pan, X., Zhan, X., Lin, D., Yu, S.X., Gong, B.: Open compound domain adaptation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 12406–12415 (2020) [1](#), [2](#), [5](#), [6](#)
- [19] Millidge, B., Salvatori, T., Song, Y., Lukasiewicz, T., Bogacz, R.: Universal hopfield networks: A general framework for single-shot associative memory models. In: *International Conference on Machine Learning*. pp. 15561–15583. PMLR (2022) [2](#), [3](#)
- [20] Pan, X., Luo, P., Shi, J., Tang, X.: Two at once: Enhancing learning and generalization capacities via ibn-net. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. pp. 464–479 (2018) [6](#)
- [21] Park, K., Woo, S., Shin, I., Kweon, I.S.: Discover, hallucinate, and adapt: Open compound domain adaptation for semantic segmentation. *Advances in Neural Information Processing Systems* **33**, 10869–10880 (2020) [1](#), [2](#), [6](#), [7](#)

- [22] Peng, X., Bai, Q., Xia, X., Huang, Z., Saenko, K., Wang, B.: Moment matching for multi-source domain adaptation. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 1406–1415 (2019) [2](#)
- [23] Peng, X., Bai, Q., Xia, X., Huang, Z., Saenko, K., Wang, B.: Moment matching for multi-source domain adaptation. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 1406–1415 (2019) [2](#)
- [24] Qiao, F., Zhao, L., Peng, X.: Learning to learn single domain generalization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12556–12565 (2020) [2](#)
- [25] Ramsauer, H., Schäfl, B., Lehner, J., Seidl, P., Widrich, M., Adler, T., Gruber, L., Holzleitner, M., Pavlović, M., Sandve, G.K., et al.: Hopfield networks is all you need. arXiv preprint arXiv:2008.02217 (2020) [2](#), [3](#)
- [26] Richter, S.R., Vineet, V., Roth, S., Koltun, V.: Playing for data: Ground truth from computer games. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part II 14. pp. 102–118. Springer (2016) [5](#)
- [27] Ros, G., Sellart, L., Materzynska, J., Vazquez, D., Lopez, A.M.: The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3234–3243 (2016) [5](#)
- [28] Saenko, K., Kulis, B., Fritz, M., Darrell, T.: Adapting visual category models to new domains. In: Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5–11, 2010, Proceedings, Part IV 11. pp. 213–226. Springer (2010) [2](#)
- [29] Termöhlen, J.A., Klingner, M., Brettin, L.J., Schmidt, N.M., Fingscheidt, T.: Continual unsupervised domain adaptation for semantic segmentation by online frequency domain style transfer. In: 2021 IEEE International Intelligent Transportation Systems Conference (ITSC). pp. 2881–2888. IEEE (2021) [3](#)
- [30] Tsai, Y.H., Hung, W.C., Schuster, S., Sohn, K., Yang, M.H., Chandraker, M.: Learning to adapt structured output space for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7472–7481 (2018) [6](#)
- [31] Venkateswara, H., Eusebio, J., Chakraborty, S., Panchanathan, S.: Deep hashing network for unsupervised domain adaptation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5018–5027 (2017) [2](#)
- [32] Vu, T.H., Jain, H., Bucher, M., Cord, M., Pérez, P.: Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2517–2526 (2019) [6](#)
- [33] Wang, A., Islam, M., Xu, M., Ren, H.: Curriculum-based augmented fourier domain adaptation for robust medical image segmentation. IEEE Transactions on Automation Science and Engineering (2023) [3](#)
- [34] Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P.: Segformer: Simple and efficient design for semantic segmentation with transformers. Advances in neural information processing systems **34**, 12077–12090 (2021) [3](#), [8](#)
- [35] Xu, Q., Zhang, R., Zhang, Y., Wang, Y., Tian, Q.: A fourier-based framework for domain generalization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 14383–14392 (2021) [2](#)
- [36] Xu, R., Chen, Z., Zuo, W., Yan, J., Lin, L.: Deep cocktail network: Multi-source unsupervised domain adaptation with category shift. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3964–3973 (2018) [2](#)
- [37] Yang, Y., Lao, D., Sundaramoorthi, G., Soatto, S.: Phase consistent ecological domain adaptation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9011–9020 (2020) [3](#)
- [38] Yang, Y., Soatto, S.: Fda: Fourier domain adaptation for semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4085–4095 (2020) [2](#), [3](#)
- [39] Yu, F., Chen, H., Wang, X., Xian, W., Chen, Y., Liu, F., Madhavan, V., Darrell, T.: Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2636–2645 (2020) [5](#)
- [40] Yu, H., Hu, M., Chen, S.: Multi-target unsupervised domain adaptation without exactly shared categories. arXiv preprint arXiv:1809.00852 (2018) [2](#)

- [41] Zhang, P., Zhang, B., Zhang, T., Chen, D., Wang, Y., Wen, F.: Prototypical pseudo label denoising and target structure learning for domain adaptive semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12414–12424 (2021) [8](#)
- [42] Zhong, X., Li, W., Liao, L., Xiao, J., Liu, W., Huang, W., Wang, Z.: Bat: Bi-alignment based on transformation in multi-target domain adaptation for semantic segmentation. In: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1–5. IEEE (2023) [3](#)
- [43] Zou, Y., Yu, Z., Kumar, B., Wang, J.: Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In: Proceedings of the European conference on computer vision (ECCV). pp. 289–305 (2018) [6](#)
- [44] Zou, Y., Yu, Z., Liu, X., Kumar, B., Wang, J.: Confidence regularized self-training. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 5982–5991 (2019) [6](#)