# Fourier-Based Instance Selective Whitening for Domain Generalized Lane Detection

Weiyan Xu[1,2], Shikui Wei[1,2](✉), Sen Xu[1,2], Chuangchuang Tan[1,2], Shengli Zhang[1,2], and Yao Zhao[1,2]

[1] Institute of Information Science, Beijing Jiaotong University, Beijing, China
{21120310,shkwei,19112009,tanchuangchuang,victory_zhang,yzhao}@bjtu.edu.cn
[2] Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing, China

**Abstract.** Lane detection represents a fundamental task within autonomous driving. While deep learning has made remarkable advancements in the source domain, its ability to generalize to unseen target domains still poses a challenge. To address this issue, we present a Fourier-based instance selective whitening framework. This framework utilizes the distinct frequencies within the Fourier spectrum to decompose data style into environment and texture styles. Our method preserves semantic features by stabilizing the phase component, while also extending the style through perturbing and amalgamating the amplitude component. Further, we propose a standardized instance selective whitening strategy to analyze overall distributional changes, emphasizing general features and reducing domain-specific information. Our approach is validated through extensive experiments across multiple challenging datasets, such as Tusimple, CULane, and LLAMAS, which demonstrates significant effectiveness when compared to existing methods.

**Keywords:** Lane detection · Domain generalization · Fourier transform

## 1 Introduction

Lane detection is a key part of the autonomous driving task, and it is the front module of systems such as Lane Keeping Assist (LKA) and Lane Departure Warning (LDW). With the development of deep learning networks, lane detection methods have made great breakthroughs [36,41]. These methods have achieved satisfactory results in the source domain, but they perform very poorly when faced with unknown domains, especially those that we cannot access. This is due to the model degradation caused by the distribution shift between the source domain and the target domain. However, the source domain cannot contain all the data distribution, so it is particularly important to reduce the domain shift between the source domain and the target domain.

Domain Adaptation (DA) [2,8,15,34,38,40] extracts data information from the target domain to alleviate the difference in features or distributions between


W. Hong and G. Kanaparan (Eds.): ICCSE 2023, CCIS 2023, pp. 415–428, 2024.
https://doi.org/10.1007/978-981-97-0730-0_37

the source domain and the target domain. However, these approaches essentially rely on the availability of target domain data, limiting the generalization to unseen target domains. Additionally, there is no single ideal target domain that encompasses all possible distributions. In contrast, domain generalization (DG) focuses on acquiring generalized content features from the domain, preventing the model from overfitting to the unique style features of the source domain. It achieves the generalization of previously unseen target domains without requiring additional target data.

In recent researches [37,38], Fourier-based data augmentation has been proposed as a method for addressing the domain shift problem, showing remarkable performance improvements. The Fourier transformation has a particular advantage: the phase component retains high-level semantics, while the amplitude component encapsulates low-level statistics [13,22,23,27]. This property contributes to generate data that more closely aligns with the real data distribution. Nevertheless, learning basic features through gradient backward propagation remains challenging due to the difficulty networks face in distinguishing style differences. Concurrently, several studies [9,10,24,26,30,32,33] have explored the combination of instance normalization and instance whitening to constrain models by removing domain-specific style information from the data, enabling the learning of general domain content features. A recent study [6] has focused on the correlation between style and feature covariance matrices to identify style-sensitive elements within features for whitening. However, the single style of source domain and the random enhancement bring difficulties in distinguishing these sensitive elements. This approach leads to elements whitened, which contains critical semantic features. It also results in a decrease in performance when applied to an unseen target domain.

In this paper, we propose two modules: Fourier-based Style Extension (FSE) and Standardized Instance Selective Whitening (SISW). FSE utilizes different frequencies from the Fourier spectrum to decompose semantic features into environmental and texture features. Then the amplitude component is disturbed and mixed to extend the distribution of the source domain with diverse styles, as well as ensure the preservation of semantics by fixing the phase component. This effectively emphasizes domain-independent information. Based on FSE, SISW standardizes and analyzes the overall feature covariance matrices to identify elements responsible for domain shifts. Then whitening these elements to enhance the discrimination ability of essential features. Our method alleviates the lack of semantic information during style expansion and improves the accuracy in selecting style-sensitive elements. Unlike methods that let the network learn by itself, such as data augmentation, our approach explicitly guides the network to learn the domain-independent features. Thus, our method leads to effective generalization across various domains.

Our main contributions are as follows:

- We propose a novel domain generalization framework for lane segmentation, which extends the source styles with Fourier transform before catching and removing domain-specific information to learn content common features.

- The proposed domain-specific style whitening modules, FSE and SISW, are plug-and-play and convenient to migrate to different backbones. It effectively improves the generalization with ignorable time consumption.
- Our approach aims at the lane detection domain generalization. Through qualitative and quantitative evaluations, the proposed method shows state-of-the-art efficacy across different domains.

## 2 Relate Work

**Semantic Segmentation on Lane Detection.** Semantic segmentation is a task of pixel-level classification. Lane segmentation is an important application of it. With the development of deep learning, many advanced methods of lane segmentation have been proposed, such as [25] generalizes traditional deep convolutions to slice-by-slice convolutions on the four directions to enhance feature extraction, [35] adds an embedding model for pixel and combines binary segmentation to cluster lane instances, [14] proposes self-attention distillation to learn from itself from contextual information, [36] adopts multitask framework contains lane segmentation, road segmentation and object detection to mutual restraint. Although these methods are effective within the source domain, they perform poorly when tested on unseen target domains. As far as we know, there have been no existing methods to address domain generalization within lane segmentation.

**Domain Generalization.** Domain Generalization (DG) aims to acquire domain-independent common features to effectively generalize to unseen target domains. Unlike Domain Adaptation (DA), DG has no access to the target domain information during training, posing significant challenges to the task. Previous research in DG has been mainly focused on image classification tasks, such as meta-learning [3,17], adversarial training [18,28], auto encoder [11,18], metric learning [7,20], data augmentation [12,39,42]. There are a few recent studies on domain generalization for semantic segmentation, which research from two aspects: domain style diversity, normalization and whitening. The methods of domain style diversity adopt data augmentation and extension to enrich the data style of the source domain. [16] augments the source domain features to resemble wild styles and stylizes the data, subsequently enhancing the capability to distinguish category semantics within the feature space. The methods of normalization and whitening enhance the generalization ability of network by eliminating distribution noise and reducing feature redundancy. [33] proves instance normalization will reserve content information to avoid overfitting the data distribution. Further, [24] adds batch normalization to improve the feature discriminability. [26] combines IN with other whitening methods to learn essential features. [6] proposes an instance selective whitening to extract domain-specific features to normalize and whiten. These two aspects complement each other: domain style diversity highlights domain-independent general features, while normalization and whitening capture general features and eliminate redundant information. The effective combination of these two aspects opens up new possibilities for domain generalization, and our aim is to explore these possibilities to enhance model generalization ability.

(a) (b) (c)

**Fig. 1.** Visualization of the low-frequency component and high-frequency component. (a) original image; (b) reconstructed image with low-frequency component; (c) reconstructed image with high-frequency component.

## 3 Methods

In this section, we apply the Fourier transform to break down the style of the data from the source domain into two different styles: environmental and textural. We extend the new data style by changing the amplitude component. Subsequently, we combine this process with standardized instance selective whitening to selectively remove domain-specific style information, thereby enhancing the generalization of model. The overall learning process is outlined in Fig. 3.

### 3.1 Fourier-Based Source Domain Style Extension

We consider that overfitting to the data style of the source domain is the primary factor leading to the decline in the generalization of model. While common data extension methods attempt to address this issue, they often present limitations. Moreover, the generated data does not consistently align with the real-world distribution, resulting in the loss of certain aspects of semantic information.

To address this challenge, we propose a Fourier-based data style extension method, which introduces random variations or mixes the amplitude information to extend the data styles. This method can diversify data styles while preserving semantic features, effectively alleviating incomplete semantic cases. Specifically, for a single-channel image, we obtain its Fourier transform:

$$\mathcal{F}(x)(u,v) = \sum_{h,w} x(h,w) e^{-j2\pi\left(\frac{h}{H}u + \frac{w}{W}v\right)} \tag{1}$$

$$\begin{aligned} \mathcal{F}^A(x)(u,v) &= \left[R^2(x)(u,v) + I^2(x)(u,v)\right]^{1/2} \\ \mathcal{F}^P(x)(u,v) &= \arctan\left[\frac{I(x)(u,v)}{R(x)(u,v)}\right] \end{aligned} \tag{2}$$

where $R(x)$ and $I(x)$ represent the real and imaginary part of $\mathcal{F}(x)$, and $\mathcal{F}^{-1}(x)$ represents the inverse transform of Fourier, and the two processes can be efficiently calculated by [21].

Utilizing the diverse frequency ranges of the Fourier transform to represent various degrees of grayscale change, we attempt to break down the data style,

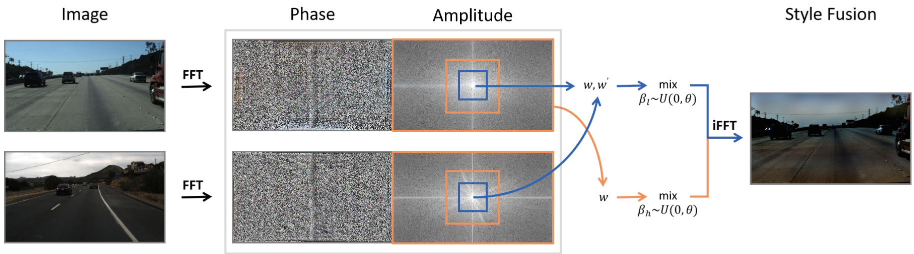


**Fig. 2.** Fourier-based style extension (FSE). We decompose the amplitude and phase components, dividing them into low-frequency and high-frequency regions. The semantic information in regions where they intersect is difficult to distinguish, so we don't address them here. Subsequently, we use different enhancement strategies for low-frequency and high-frequency amplitudes to extend styles.

as the Fig. 1. We consider the low-frequency signal in the Fourier domain as representing relatively smooth environmental features, while the high-frequency signal embodies the significantly changeable texture features. To represent these distinct types of signals separately, we define masks $\mathrm{M}_{\beta_{\text{low}}}$ and $\mathrm{M}_{\beta_{\text{high}}}$:

$$\begin{aligned} \mathrm{M}_{\beta_l}(h,w) &= \mathbb{1}_{(h,w)\in[-\beta_l H:\beta_l H, -\beta_l W:\beta_l W]} \\ \mathrm{M}_{\beta_h}(h,w) &= 1 - \mathbb{1}_{(h,w)\in[-\beta_h H:\beta_h H, -\beta_h W:\beta_h W]} \end{aligned} \tag{3}$$

where $\mathbb{1}$ denotes value in the range is 1 and outside is 0. $\beta_l, \beta_h \in (0,1)$ and $\beta_l + \beta_h < 1$, as the Fig. 2.

In recent studies [13,22,23,27], it has been proved that the Fourier phase component can effectively preserve semantic information. Based on this, we highlight the semantic feature by fixing the phase component of Fourier space. Then, we introduce random variations in the amplitude to extend the data styles. Inspired by [37,38], for the low-frequency component, we engage in random linear interpolation of the amplitudes from two arbitrary images. Considering the similarity in style information from the source domain image, we randomly adjust the value of the amplitude component, either by enlarging or reducing it, before the linear interpolation of the amplitudes. Combined with the random transform, the extension of environmental styles can be expressed as:

$$\mathcal{F}^A_{env}(x_i) = \mathrm{M}_{\beta_l} \circ ((1-\lambda)w_i\mathcal{F}^A(x_i) + \lambda w_{i'}\mathcal{F}^A(x_{i'})) + (1-\mathrm{M}_{\beta_l}) \circ \mathcal{F}^A(x_i) \tag{4}$$

where $\lambda \in (0,1)$ is the fusion factor, $w_i$ and $w_{i'}$ are the disturbance factors of low frequency about different samples $x_i$ and $x_{i'}$.

We attempted to apply a similar method to the high-frequency component, but it did not yield an improved outcome. This was primarily due to the high-frequency component containing a portion of semantic information. When combined with another image, the mix resulted in the destruction of semantic information. Consequently, we only and randomly reduce the amplitude, proportionally, in order to preserve the original semantics and prevent the loss of semantic details.

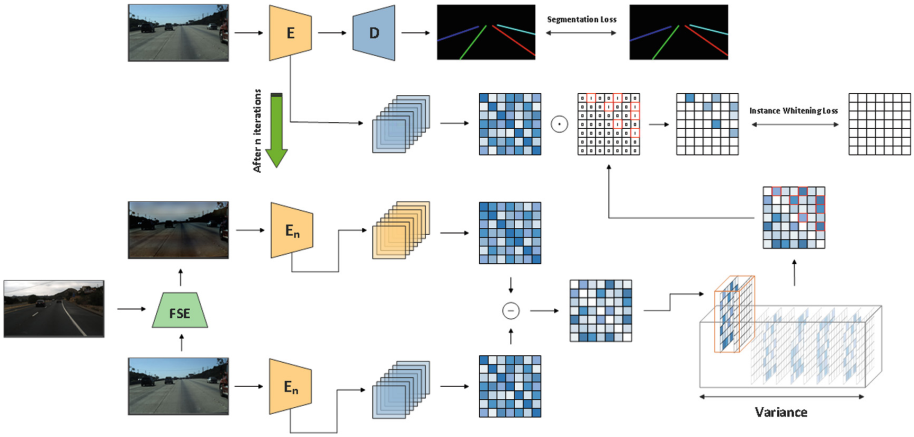


**Fig. 3.** The framework of the proposed FSW. $E_n$ denotes our model is trained after $n$ iterations. It stops training to generate a selective matrix, and then adds standardized instance selective whitening loss during subsequent training.

$$\mathcal{F}^A_{tex}(x_i) = \mathrm{M}_{\beta_h} \circ b_i \mathcal{F}^A(x_i) + (1 - \mathrm{M}_{\beta_h}) \circ \mathcal{F}^A(x_i) \tag{5}$$

where $b_i$ is the disturbance factor of high frequency. Thus, for the source domain image $x_i$ and arbitrary sample $x_{i'}$. We extend environmental features and texture features in turn. The style extension can be formalized as:

$$(\mathcal{F}^A_{env}(x_i), \mathcal{F}^A_{tex}(x_i)) \rightarrow \hat{\mathcal{F}}^A(x_i) \tag{6}$$

$$x_{ext} = \mathcal{F}^{-1}([\hat{\mathcal{F}}^A(x_i), \mathcal{F}^P(x_i)]) \tag{7}$$

where $\mathcal{F}^A_{env}(x_i)$ and $\mathcal{F}^A_{tex}(x_i)$ are two relatively independent processes that need to be executed sequentially, with no strict order between them.

### 3.2 Standardized Instance Selective Whitening

The previously mentioned Fourier-based style extension method enriches the data distribution of the source domain. However, merely constraining the model through the comparison loss results in the model overfitting the new distribution. This approach makes it challenging to learn domain-generalized content features. Recent studies [9,10] have highlighted the relationship between style and feature statistics. The adjustment of feature statistics using the IN layer proves beneficial in enhancing feature diversity. Inspired by [6], we propose standardized instance selective whitening, a method that standardizes the covariance matrix and selects the style-sensitive feature elements for whitening. We determine the feature sensitivity to style through overall distribution changes in covariance matrices. This enables us to improve the discernment of domain-specific features and accurately remove redundant style information.

Following [6,24], we add the instance standardization layer [33] into shallow networks. Initially, we train $n$ epochs without containing whitening loss to enable the networks to learn basic semantic features. Subsequently, we infer both the original image and its stylistically extended image, calculating the covariance matrix for each:

$$\mu = \tfrac{1}{HW}\mathrm{X} \cdot 1 \in \mathbb{R}^{C\times 1} \tag{8}$$

$$\Sigma = \tfrac{1}{HW}(\mathrm{X} - \mu \cdot 1^{\top})(\mathrm{X} - \mu \cdot 1^{\top})^{\top} \in \mathbb{R}^{C\times C} \tag{9}$$

where $\mathrm{X} \in \mathbb{R}^{C\times HW}$ denotes the feature map with the dimensions of C, H and W, expanded along the channel dimension. $1 \in \mathbb{R}^{HW}$ is a column vector of ones, and $\mu$ and $\Sigma$ are the mean vector and the covariance matrix.

[6] uses the mean of instance variance to describe the style sensitivity of feature elements. Nevertheless, the variance of individual instances has limitations. We consider the changes in the overall distribution might provide more accurate representations. As there are inherent differences among the feature covariance matrices of distinct images, our method standardizes covariance matrices initially. Then, we represent the style sensitivity of feature elements by the variance of overall distribution, thereby obtaining the variance matrix:

$$\mathrm{V} = \frac{1}{N}\sum_{i=1}^{N} \sigma_i^2 \tag{10}$$

$$\mu_{\Sigma} = \frac{1}{2N}\sum_{i=1}^{N}(\Sigma_{\mathrm{s}}(\mathrm{x_i}) + \Sigma_{\mathrm{s}}(\tau(\mathrm{x_i})) - 2\Sigma_{\mathrm{sta}}(\mathrm{x_i})) \tag{11}$$

$$\sigma_i^2 = \frac{1}{2}((\Sigma_{\mathrm{s}}(\mathrm{x_i}) - \Sigma_{\mathrm{sta}}(\mathrm{x_i}) - \mu_{\Sigma})^2 + (\Sigma_{\mathrm{s}}(\tau(\mathrm{x_i})) - \Sigma_{\mathrm{sta}}(\mathrm{x_i}) - \mu_{\Sigma})^2) \tag{12}$$

where $N$ is the number of samples, $x_i$ is the $i$-th image sample, $\tau$ is the Fourier-based style extension, and $\Sigma_{\mathrm{sta}}(\cdot)$ denotes the regularized covariance matrix, here we simply use the covariance matrix of the original image, $\Sigma_{\mathrm{sta}}(\cdot) = \Sigma_{\mathrm{s}}(\cdot)$.

The above studies [9,10] indicate that the feature covariance of a model with poor generalization is sensitive to style shifts. The variance matrix V accurately reflects this sensitivity. Consequently, we assume that the elements with higher variance in the covariance matrix may contain more domain-specific styles. To select these elements, we follow the method of [6], utilizing k-means clustering to assign the elements $\mathrm{V}_{i,j}(i < j)$ into k clusters. Subsequently, we select the elements in clusters with the top m high variance values as $G_{high}$. We can get the mask matrix M from $G_{high}$:

$$\mathrm{M}_{i,j} = \begin{cases} 1, & if\ \mathrm{V}_{i,j} \in G_{high} \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

The mask represents the domain-specific style elements within the features. Consequently, we propose standardized instance selective whitening (SISW) loss and implement it within the instance normalization layer to optimize the network:

$$\mathcal{L}_{\text{FISW}} = E\left[\left\|\Sigma_{\text{s}} \odot \text{M}\right\|_1\right] \tag{14}$$

Combining with lane segmentation tasks, our total loss can be described as

$$\mathcal{L} = \mathcal{L}_{\text{seg}} + \mathcal{L}_{\text{exist}} + \frac{1}{L}\sum_{i}^{L} \mathcal{L}_{\text{FISW}}^{i} \tag{15}$$

where $\mathcal{L}_{\text{seg}}$ is the cross-entropy loss for lane segmentation, $\mathcal{L}_{\text{exist}}$ is the lane exist loss, $i$ and $L$ denotes separately the layer and the total number which applies SISW.

## 4 Experiments

In this section, we conduct many experiments between simple and complex scenes to prove the efficiency of our method. Besides, we compare our method with multiple representative methods of feature normalization and whitening in the field of domain generalization for semantic segmentation.

### 4.1 Dataset

We focus on the domain generalization of lane detection and design two standard benchmark settings: "Tusimple to CULane" and "Tusimple to LLAMAS". These aim to verify the generalization ability of models from a single scene to a complex scene.

**Tusimple** [1] is a small-scale dataset containing urban driving scene images with the resolution of $1280 \times 720$. The images are collected from high-speed roads. It provides 3626 training and 2782 testing images.

**CULane** [25] is a large-scale dataset for lane detection with the resolution of $1680 \times 590$. It collects more than 55 h videos and extract 133235 frames. The dataset is split into 88880 training, 9675 validation and 34680 testing images. The test set contains normal class and 8 challenging classes.

**LLAMAS** [4] is a large-scale and high-quality dataset for lane detection. It contains 100042 labeled images with the resolution of $1276 \times 717$. Since the label of testset is not public, we use the validation set for evaluation.

### 4.2 Experimental Setup

We conducted training on the Tusimple dataset and evaluated the performance of the model on the CULane and LLAMAS datasets to assess its generalization ability across a range of scenarios, from simple scenes to complex ones. In our comparison with other normalization methods, we re-implemented IBN-Net [24], IW [19], GIW [5], and ISW [6] on our baseline models. Metrics were consistently measured using the same evaluation standards for all approaches. To ensure fair comparisons, models were selected and trained on the last epoch.

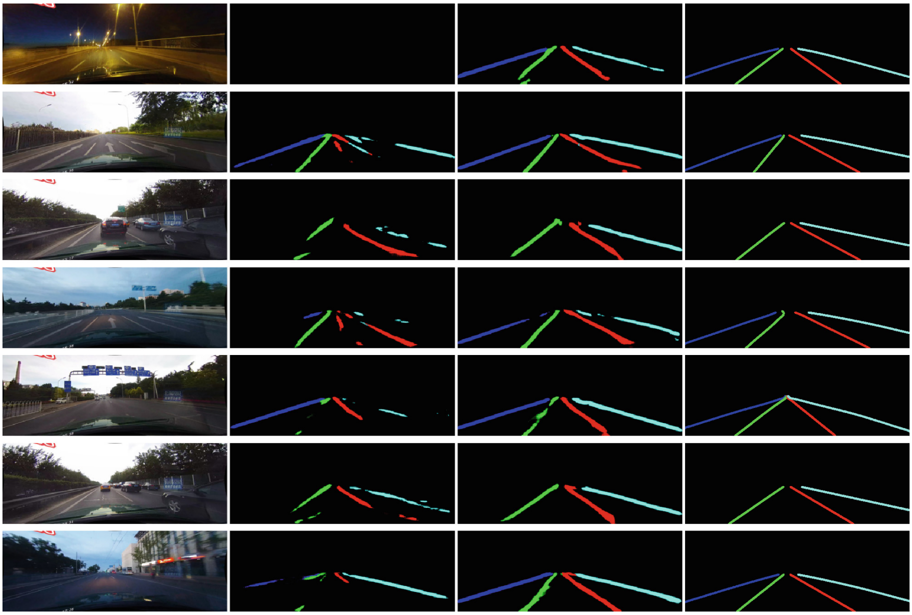

**Fig. 4.** Visualization results on CULane. From left to right are (a) Input image, (b) Baseline, (c) Ours (FSW) and (d) Groundtruth.

### 4.3 Implementation Details

We adopt ERFNet [29] pre-trained on ImageNet [31] as the network backbone. We use SGD optimizer with a momentum of 0.9 and weight decay of 1e-4. The initial learning rate is set to 1e-2 and is decreased using the polynomial policy with a power of 0.9. We train a model for 12 epochs with a batch size of 16. Before training, all datasets are preprocessed to contain a maximum of four categories (ranging from two lanes on each side), and all images are resized to $976 \times 208$.

We implement our method on PyTorch and use a single NVIDIA RTX 3090 GPU for our experiments. Following lane detection works, we use F1-score as the evaluation metric.

### 4.4 Comparison with DG Methods

We construct a simple to complex adaptation scenario to verify the effectiveness of FSW. We compare our results with other representative normalization methods: IBN-Net [24], IW [19], GIW [5] and ISW [6]. Table 1 shows the generalization performance test on CULane dataset. Note that all the methods use ERFNet [29] as the network backbone. We select the model from the last epoch of training to test for each method. FSW outperforms other methods on all challenging classes of CULane except the Crossroad. In addition, for styles that do

**Table 1.** Quantitative comparison on "TuSimple to CULane". *: For cross, it only shows FP.

| Methods | Normal | Crowded | Dazzle | Shadow | No line | Arrow | Curve | Cross* | Night | F1@50 |
|---|---|---|---|---|---|---|---|---|---|---|
| Baseline [29] | 32.26 | 13.12 | 8.09 | 6.29 | 7.37 | 20.82 | 20.76 | 9656 | 4.70 | 16.82 |
| IBN [24] | 36.07 | 13.73 | 7.59 | 7.07 | 6.82 | 23.66 | 25.18 | 8597 | 8.54 | 18.48 |
| IW [26] | 38.07 | 14.70 | 10.95 | 6.39 | 7.82 | 26.61 | 26.31 | 8249 | 8.06 | 19.62 |
| GIW [5] | 39.25 | 16.05 | 12.48 | 8.94 | 9.11 | 27.91 | 28.09 | **8180** | 10.88 | 21.11 |
| ISW [6] | 40.10 | 17.59 | 10.62 | 8.32 | 8.43 | 28.93 | 28.49 | 9075 | 12.37 | 21.69 |
| Ours | **41.84** | **19.65** | **13.20** | **10.00** | **10.09** | **30.30** | **29.35** | 8802 | **15.27** | **23.69** |

**Table 2.** Quantitative comparison over multiple domains.

| Methods | →CULane | →LLAMAS |
|---|---|---|
| Baseline [24] | 16.83 | 66.17 |
| IBN [24] | 18.48 | **69.95** |
| IW [26] | 19.62 | 65.19 |
| GIW [5] | 21.11 | 65.66 |
| ISW [6] | 21.69 | 65.17 |
| Ours | **23.69** | 68.99 |

not exist in the source domain, such as dazzle and night, our method has significant improvements compared with other methods. To prove the generalization ability over multiple domains, we use the same model to verify the performance on CULane and LLAMAS datasets. Given in Table 2, our approach also achieves significant results, only slightly lower than IBN on the LLAMAS dataset, but we are much higher than it on the more complex CULane dataset. The visualization results are shown in Fig. 4.

**Table 3.** Ablation analysis of FSE and SISW.

| Methods | FSE | SISW | F1@50 |
|---|---|---|---|
| Baseline [24] | | | 16.83 |
| +ISW | | | 21.69 |
| +ISW | ✓ | | 22.98 |
| +ISW | ✓ | ✓ | **23.69** |

### 4.5 Ablation Studies

In order to further analyze the effectiveness of FSE and SISW, we extend experiments on the domain generalization scene from Tusimple to CULane. We evaluate the effectiveness through method stacking. As shown in Table 3, the baseline model, trained only with FSE module, achieves an F1-score of 22.98% with +6.15% improvement. This shows that style extensions with more semantic information will contribute to selecting elements that are sensitive to domain shift more accurately. Further, we adopt SISW module to achieve an F1-score of 23.69%. This shows the distribution of global covariance based on standardization is better for representing the sensitivity of features to domain shift.

## 5 Conclusions

This paper proposes a novel Fourier-based standardized instance selective whitening framework, which contains two modules: Fourier-based Style Extension (FSE) and Standardized Instance Selective Whitening (SISW). FSE focuses on preserving semantic information through the Fourier spectral phase while extending the amplitude to diversify the style. SISW selectively removes the domain-specific information to enhance common feature learning. The collaboration between the two modules, SISW and FSE, proves to be complementary. SISW encourages networks to learn style-independent essential features while FSE contributes to improving the discrimination of style-sensitive feature elements. Extensive experiments confirm its effectiveness in generalizing from simpler to more complex scenes. However, challenges remain in extending styles not present in the source domain, such as shadow and night scenes, this leads to a considerable performance gap between the source and target domains. Addressing these challenges will be a primary focus of our future work.

## References


1. Tusimple. Tusimple benchmark (2017)
2. Araslanov, N., Roth, S.: Self-supervised augmentation consistency for adapting semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 15384–15394 (2021)
3. Balaji, Y., Sankaranarayanan, S., Chellappa, R.: Metareg: towards domain generalization using meta-regularization. Advances in neural information processing systems 31 (2018)
4. Behrendt, K., Soussan, R.: Unsupervised labeled lane markers using maps. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (2019)
5. Cho, W., Choi, S., Park, D.K., Shin, I., Choo, J.: Image-to-image translation via group-wise deep whitening-and-coloring transformation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10639–10647 (2019)

6. Choi, S., Jung, S., Yun, H., Kim, J.T., Kim, S., Choo, J.: Robustnet: Improving domain generalization in urban-scene segmentation via instance selective whitening. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11580–11590 (2021)
7. 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)
8. Ganin, Y., et al.: Domain-adversarial training of neural networks. J. Mach. Learn. Res. **17**(1), 2096–2030 (2016)
9. Gatys, L., Ecker, A.S., Bethge, M.: Texture synthesis using convolutional neural networks. Advances in neural information processing systems 28 (2015)
10. Gatys, L.A., Ecker, A.S., Bethge, M.: Image style transfer using convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2414–2423 (2016)
11. Ghifary, M., Kleijn, W.B., Zhang, M., Balduzzi, D.: Domain generalization for object recognition with multi-task autoencoders. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2551–2559 (2015)
12. Gong, R., Li, W., Chen, Y., Gool, L.V.: Dlow: domain flow for adaptation and generalization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2477–2486 (2019)
13. Hansen, B.C., Hess, R.F.: Structural sparseness and spatial phase alignment in natural scenes. JOSA A **24**(7), 1873–1885 (2007)
14. Hou, Y., Ma, Z., Liu, C., Loy, C.C.: Learning lightweight lane detection CNNs by self attention distillation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1013–1021 (2019)
15. 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)
16. Lee, S., Seong, H., Lee, S., Kim, E.: Wildnet: Learning domain generalized semantic segmentation from the wild. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9936–9946 (2022)
17. Li, D., Yang, Y., Song, Y.Z., Hospedales, T.: Learning to generalize: Meta-learning for domain generalization. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32 (2018)
18. Li, H., Pan, S.J., Wang, S., Kot, A.C.: Domain generalization with adversarial feature learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5400–5409 (2018)
19. Li, Y., Fang, C., Yang, J., Wang, Z., Lu, X., Yang, M.H.: Universal style transfer via feature transforms. Advances in neural information processing systems 30 (2017)
20. Motiian, S., Piccirilli, M., Adjeroh, D.A., Doretto, G.: Unified deep supervised domain adaptation and generalization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 5715–5725 (2017)
21. Nussbaumer, H.J., Nussbaumer, H.J.: The fast Fourier transform. Springer (1982)
22. Oppenheim, A., Lim, J., Kopec, G., Pohlig, S.: Phase in speech and pictures. In: ICASSP'79. IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 4, pp. 632–637. IEEE (1979)
23. Oppenheim, A.V., Lim, J.S.: The importance of phase in signals. Proc. IEEE **69**(5), 529–541 (1981)

24. 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)
25. Pan, X., Shi, J., Luo, P., Wang, X., Tang, X.: Spatial as deep: spatial CNN for traffic scene understanding. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32 (2018)
26. Pan, X., Zhan, X., Shi, J., Tang, X., Luo, P.: Switchable whitening for deep representation learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1863–1871 (2019)
27. Piotrowski, L.N., Campbell, F.W.: A demonstration of the visual importance and flexibility of spatial-frequency amplitude and phase. Perception **11**(3), 337–346 (1982)
28. Rahman, M.M., Fookes, C., Baktashmotlagh, M., Sridharan, S.: Correlation-aware adversarial domain adaptation and generalization. Pattern Recogn. **100**, 107124 (2020)
29. Romera, E., Alvarez, J.M., Bergasa, L.M., Arroyo, R.: Erfnet: efficient residual factorized convnet for real-time semantic segmentation. IEEE Trans. Intell. Transp. Syst. **19**(1), 263–272 (2017)
30. Roy, S., Siarohin, A., Sangineto, E., Bulo, S.R., Sebe, N., Ricci, E.: Unsupervised domain adaptation using feature-whitening and consensus loss. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9471–9480 (2019)
31. Russakovsky, O., et al.: Imagenet large scale visual recognition challenge. Int. J. Comput. Vision **115**, 211–252 (2015)
32. Siarohin, A., Sangineto, E., Sebe, N.: Whitening and coloring batch transform for gans. arXiv preprint arXiv:1806.00420 (2018)
33. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Instance normalization: The missing ingredient for fast stylization. arXiv preprint arXiv:1607.08022 (2016)
34. 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)
35. Wang, Z., Ren, W., Qiu, Q.: Lanenet: real-time lane detection networks for autonomous driving. arXiv preprint arXiv:1807.01726 (2018)
36. Wu, D., et al.: Yolop: you only look once for panoptic driving perception. Mach. Intell. Res. **19**(6), 550–562 (2022)
37. 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)
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)
39. Yue, X., Zhang, Y., Zhao, S., Sangiovanni-Vincentelli, A., Keutzer, K., Gong, B.: Domain randomization and pyramid consistency: simulation-to-real generalization without accessing target domain data. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2100–2110 (2019)
40. 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)

41. Zheng, T., et al.: Clrnet: cross layer refinement network for lane detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 898–907 (2022)
42. Zhou, K., Yang, Y., Hospedales, T., Xiang, T.: Learning to generate novel domains for domain generalization. In: Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVI 16, pp. 561–578. Springer (2020)