Revisiting Downsampling in Semantic Segmentation: Fighting Aliasing with Dynamic Gaussian and Gabor Frequency Filters

Yu Bing Luo[1], Nian Shi[1], Jia Qin[1*], Zekai Ji[1], Pinle Qin[1*], Jianchao Zeng[1], Jianghui Cai[1]

[1]College of Computer Science and Technology, North University of China
{sz202107003@st., s202407036@st., jqin@, s202407027@st., qpl@, zjc@}nuc.edu.cn, jianghui@tyust.edu.cn

Abstract

Downsampling is essential in semantic segmentation for reducing computational cost and guiding the learning of class-discriminative features. Existing models typically rely on strided convolutions or patch splitting to obtain features with lower resolution. However, we observe that such operations often introduce edge jagging and texture degradation, the underlying cause is that aliasing of the high frequency induces phase distortion. We conducted a systematic analysis of phase distortion and identified two key properties: spatial non-uniformity (concentrated near boundaries) and directional sparsity (accumulated along a few dominant directions). These properties cause crucial high-frequency cues to be misrepresented or lost during sampling. To address this issue, we propose a frequency aware filter consisting of two complementary modules: a dynamic Gaussian kernel (DGK) and a learnable Gabor-based frequency selector (LFS). To mitigate spatial non-uniformity, the DGK predicts edge normals from gradients, applies strong low-pass filtering along the normal direction, and leaves the tangential direction virtually untouched, thereby suppressing phase distortion while preserving contour continuity. To handle directional sparsity, the Learnable Gabor Selector (LFS) then performs directional band-pass filtering to attenuate residual aliasing peaks and adaptively boost informative texture. We further introduce phase-error energy (PE) to quantify distortion severity. Visualization and quantitative results demonstrate that frequency-aware filter offers a plug-and-play remedy for aliasing, yielding sharper boundaries and consistent gains across datasets.

1 Introduction

Semantic segmentation plays a vital role in real-world vision applications, such as autonomous driving (Yu et al. 2025; Chen et al. 2024c; Jiang et al. 2023), medical image segmentation(Wu et al. 2025; Specktor-Fadida et al. 2025; Ma et al. 2024), and development of intelligent robotics (Milioto and Stachniss 2019; Di Palo and Johns 2024; Bajcsy et al. 2024). These advanced visual tasks rely heavily on high-frequency structural details such as edges and textures(Zhou and Wang 2024; Zhou et al. 2024; Zhang, Panahi, and Gao 2023).

Existing models typically adopt multi-stage downsampling structures(Xie et al. 2021; Cheng et al. 2022; Guo et al. 2022; Chen, Gu, and Fu 2024; Xiong et al. 2024), such as patch splitting and strided convolutions, to progressively reduce feature resolution and balance model complexity with performance. However, according to the Nyquist–Shannon sampling theorem(Shannon 2006; Nyquist 1928), signals with frequencies exceeding half of the sampling rate suffer from distortion, leading to frequency-domain aliasing. For instance, in convolutional networks with a stride-2 downsampling, the sampling rate is 1/2, which means that frequencies above 1/4 are distorted. These high-frequency components are folded into the lower frequency spectrum, introducing phase distortions in the frequency domain.

In semantic segmentation, such phase distortions often manifest as blurred class boundaries and loss of fine texture details. Although prior methods have employed handcrafted low-pass filters(Zou et al. 2023; Zhang 2019; Grabinski et al. 2022; Chen et al. 2023; Wang et al. 2023) to suppress high frequencies and alleviate aliasing, they often rely on empirically defined kernels or fixed frequency bands. This limits the ability of the model to distinguish meaningful directional high-frequency components , which are vital for distinguishing the foreground from the background (Chen et al. 2024a; Zhao et al. 2025; Wang et al. 2024; Karras et al. 2021; Qin et al. 2021).

To thoroughly understand how aliasing affects semantic segmentation accuracy, we performed a feature-level analysis of phase distortions induced by aliasing. We observed that phase distortions exhibit spatial regularities and affect model predictions(as illustrated in Fig. 1): (1) Spatial non-uniformity: Phase distortions caused by aliasing tend to concentrate around object boundaries. Such boundary pixel displacements may result in discontinuities in thin structures or complete disappearance of small objects (Fig. 1(a), top). (2) Directional sparsity: Aliasing often accumulates along one or two dominant directions. The more concentrated the directionality, the more severe the phase distortion, which may lead to the model may misinterpret horizontally aliased textures as true edges and incorrectly split the target (Fig. 1(a), bottom). To better visualize these two properties, Fig. 1(b) presents a phase-error heatmap, where highlighted regions predominantly align with object edges. Fig. 1(c) shows the angular distribution of phase offsets in polar coordinates, where energy spikes are concentrated along the horizontal and vertical axes. For detailed analysis, see Appendix A.

*Corresponding author.

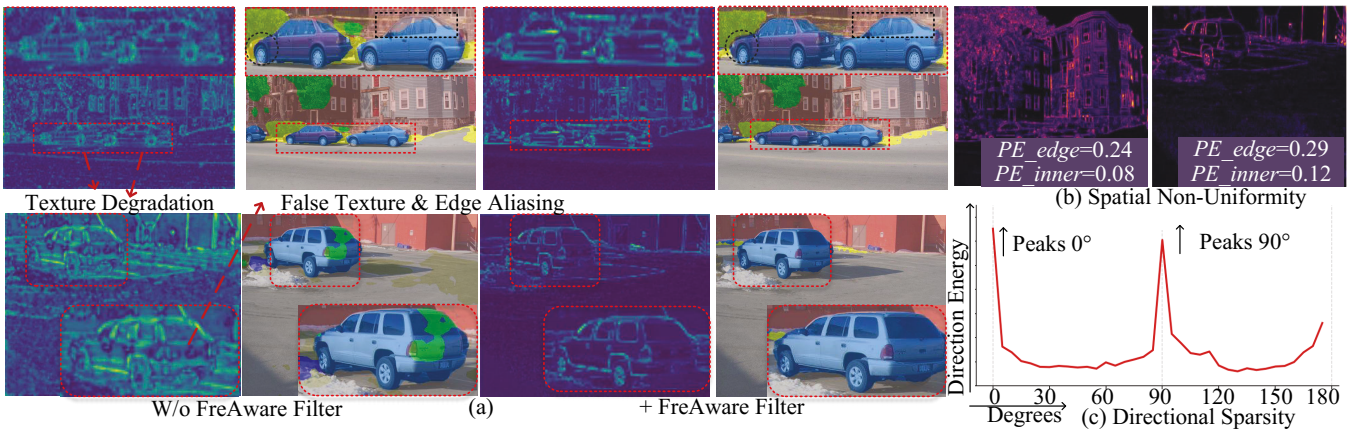


Figure 1: (a) Illustration of boundary texture loss caused by spatial non-uniformity and false interior textures arising from directional sparsity. (b) Visualisation of phase error after downsampling,the phase error at object boundaries is substantially greater than that within object interiors, confirming spatial non-uniformity. (c) Radial energy distribution obtained by projecting the heatmap over angles, sharp peaks appear only around 0° and 90°,validating the presence of directional sparsity.

Based on the above observations, we propose a frequency-aware filter (Freq Aware Filter), which consists of two components: a dynamic Gaussian kernel (DGK) and a learnable Gabor-based frequency selector (LFS) .

Owing to the spatial non-uniformity of aliasing, DGK predicts locally adaptive Gaussian kernels to flexibly suppress regional high-frequency interference while preserving critical edge information. It computes image gradients to identify local frequency distributions, maps them to blur radii, and introduces a rotation matrix to align kernel orientation. Anisotropic Gaussian filters are then generated per pixel, performing directional smoothing orthogonal to object edges to enhance intra-object consistency and maintain contour integrity.

However, Owing to directional sparsity, residual aliasing concentrated along specific directions can still cause phase distortion, resulting in jagged edges and false textures.To address this, we designed LFS to perform parametric Gabor filtering in the frequency domain.By explicitly encoding the scale, central frequency, and orientation into the convolution weights, LFS adaptively suppresses residual directional aliasing and performs residual-based high-pass reconstruction to selectively preserve discriminative textures. To quantitatively assess phase distortion, we introduced a new metric phase error energy (PE), which is defined as the difference between the sampled phase and the ideal phase in the frequency domain.

Experimental results demonstrate that the Freq Aware Filter effectively suppresses aliased high-frequency components while preserving informative textures and enhancing object boundaries. Quantitative analysis shows a significant reduction in PE. On the ADE20K dataset(Zhou et al. 2017), SegNext(Guo et al. 2022) achieved a 1.8-point improvement in mIoU. On Cityscapes(Cordts et al. 2016), both Mask2Former(Cheng et al. 2022) and SegNext saw a gain of 0.8 mIoU. On COCO-Stuff(Caesar, Uijlings, and Ferrari 2018), SegNext improved by 0.7 mIoU, with only a 0.26M increase in parameters.

Our contributions can be summarized as follows:

- Through systematic visualization and frequency-domain analysis, we are the first to reveal two key patterns of aliasing-induced phase distortion: spatial non-uniformity and directional sparsity. Based on this, we design a Frequency Aware Filter including DGK and LFS that dynamically and precisely suppresses aliasing and corrects phase distortion. This module can be seamlessly integrated into existing segmentation frameworks.
- We propose a new measurement to quantitatively analyze phase distortion,PE,which directly measures the magnitude of phase shifts caused by aliasing. This provides a unified scale for future comparison and analysis of anti-aliasing methods, laying a solid foundation for developing more effective anti-aliasing techniques.
- Quantitative and qualitative results demonstrate that the Freq Aware Filter effectively eliminates aliasing and significantly enhances the performance of state-of-the-art semantic segmentation frameworks.

2 Related Work

2.1 Alias in Deep Learning

In recent years, increasing attention has been paid to the aliasing effect. In convolutional neural network (CNN)-based models, early studies such as Blur Pool (Zhang 2019) employed fixed blur kernels to suppress high-frequency components. More recently, Zou et al. (Zou et al. 2023) proposed dynamically predicts convolution kernels along the channel dimension to alleviate aliasing.Hossain et al. (Hossain et al. 2023) introduced deep adaptive activation functions to reduce aliasing. In transformer-based architectures, aliasing is often more severe owing to the aggressive patch partitioning strategy. To address this, Qian et al. (Qian et al. 2021)incorporated low-pass filtering into vision transformers to reduce aliasing during the token embedding process.

Chen (Chen, Gu, and Fu 2024) conducted a analysis of aliasing , proposing a local frequency-based strategy to dynamically suppress high-frequency noise. While the aforementioned approaches have demonstrated effectiveness in mitigating aliasing, they remain limited in selectively preserving meaningful textures and capturing directional frequency components.

2.2 Frequency domain learning

Frequency-domain analysis has been widely employed in digital image processing. In recent years, such techniques have been successfully integrated into deep learning models, proving effective in model optimization (Yin et al. 2019), feature enhancement (Chen et al. 2025, 2024b; Huang et al. 2022; Fuoli, Van Gool, and Timofte 2021), and feature fusion (Chen et al. 2024a). FADC (Chen et al. 2024b) shows how to dynamically generate dilated convolutions in the frequency domain to expand the receptive field and improve prediction accuracy. FDConv (Chen et al. 2025) divided features into disjoint frequency groups in the Fourier domain, enabling the construction of diverse frequency-specific weights with minimal parameters, thereby enhancing performance in dense prediction tasks. Compared to existing methods, our approach offers superior frequency control granularity, enhanced spatial adaptivity, and improved boundary representation quality, while maintaining seamless compatibility with mainstream backbone networks.

2.3 Semantic Segmentation

Semantic segmentation requires precise classification of edge textures. SegFormer (Xie et al. 2021) integrates a lightweight Transformer encoder, significantly improving segmentation efficiency and establishing itself as a representative architecture for efficient segmentation. Mask2Former (Cheng et al. 2022) emphasizes global context modeling through mask-based attention to enhance segmentation performance. SegNeXt (Guo et al. 2022) represents a modern redesign of CNN-based segmentation frameworks. Built upon the ConvNeXt backbone, it introduces efficient optimizations tailored for semantic segmentation. The proposed Freq Aware Filter can be seamlessly integrated into state-of-the-art architectures, delivering performance gains with only minimal additional parameters.

3 Method

In this section, we present Freq Aware Filter together with the proposed PE. Freq Aware Filter comprises two core modules: (i) DGK, which flexibly suppresses aliasing-induced high frequencies to address spatial non-uniformity (see Fig. 2, top); and (ii) LFS, which removes residual directional aliasing while selectively preserving and enhancing useful textures to cope with directional sparsity. Section 3.1 details DGK, Section 3.2 describes LFS, and Section 3.3 introduces the PE metric and the concept of phase distortion.

3.1 Dynamic Guassian Blur Kernel

Owing to spatial non-uniformity, phase distortion exhibits localization patterns similar to high-frequency components in the image. DGK leverages this correspondence by explicitly using spatial gradients to generate a set of location-aware anisotropic Gaussian kernels for each spatial position. These kernels automatically adjust their bandwidth based on local semantic and texture information: larger bandwidth values are applied within object interiors for smoothing, while smaller bandwith values are used near boundaries to preserve sharpness and reduce blurring.

As illustrated in the upper part of Fig. 2, given features $F \in \mathbb{R}^{C\times H\times W}$, we first compute the Sobel gradients$\{g_x, g_y\}$ for each channel, along with the corresponding gradient magnitude m. A larger value of m indicates the presence of richer high-frequency structures (typically corresponding to edges). the gradients are defined as follows:

$$\begin{aligned} g_x(i,j) &= \frac{1}{c}\sum_c \left|\varphi_x\, F_c(i,j)\right|, \\ g_y(i,j) &= \frac{1}{c}\sum_c \left|\varphi_y\, F_c(i,j)\right|, \end{aligned} \tag{1}$$

$$m(i,j) \;=\; \sqrt{g_x^2 + g_y^2}\,. \tag{2}$$

Here, C denotes the number of channels, φ_x and φ_y are the Sobel kernels, and $g_x(i,j)$, $g_y(i,j)$ represent the average gradients in the x and y directions at spatial location (i,j). Then gradient magnitude $m(i,j) \in [0,1]$ is then mapped to a bandwidth range $[\sigma_{\min}, \sigma_{\max}]$.The bandwidth $\sigma(i,j)$ is defined as

$$\sigma(i,j) = \sigma_{\min} + (\sigma_{\max} - \sigma_{\min})\exp\big(-m(i,j)\big). \tag{3}$$

To suppress aliased high-frequency components without blurring valid contours, we derive the edge-normal direction from $g_x(i,j)$ and $g_y(i,j)$, and introduce a rotation matrix R_θ so that the kernel is orthogonal to the local edge applying low pass filtering only along the direction perpendicular to the edge. The DGK kernel direction is shown on the right side of Figure 4(c). The orientation angle $\theta(i,j)$ is defined as

$$\theta(i,j) = \operatorname{atan2}\big(g_y(i,j),\, g_x(i,j)\big), \tag{4}$$

and the corresponding rotation matrix is

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \tag{5}$$

where, θ is the dominant frequency orientation. From a frequency-domain perspective, DGK constructs a spatially-variant low-pass filter at each pixel, where the cutoff frequency is governed by $\sigma(i,j)$ and the orientation is controlled by $\theta(i,j)$. The kernel at each location is then defined over local coordinates $\mathbf{P} = (x,y)^{\mathrm{T}}$ with $(x,y) \in [-k/2, k/2]$ as follows:

$$G(\sigma,\theta) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{\|R_\theta \mathbf{P}\|^2}{2\sigma^2}\right), \tag{6}$$

The final output is

$$G_t = F \circledast G(\sigma,\theta), \tag{7}$$

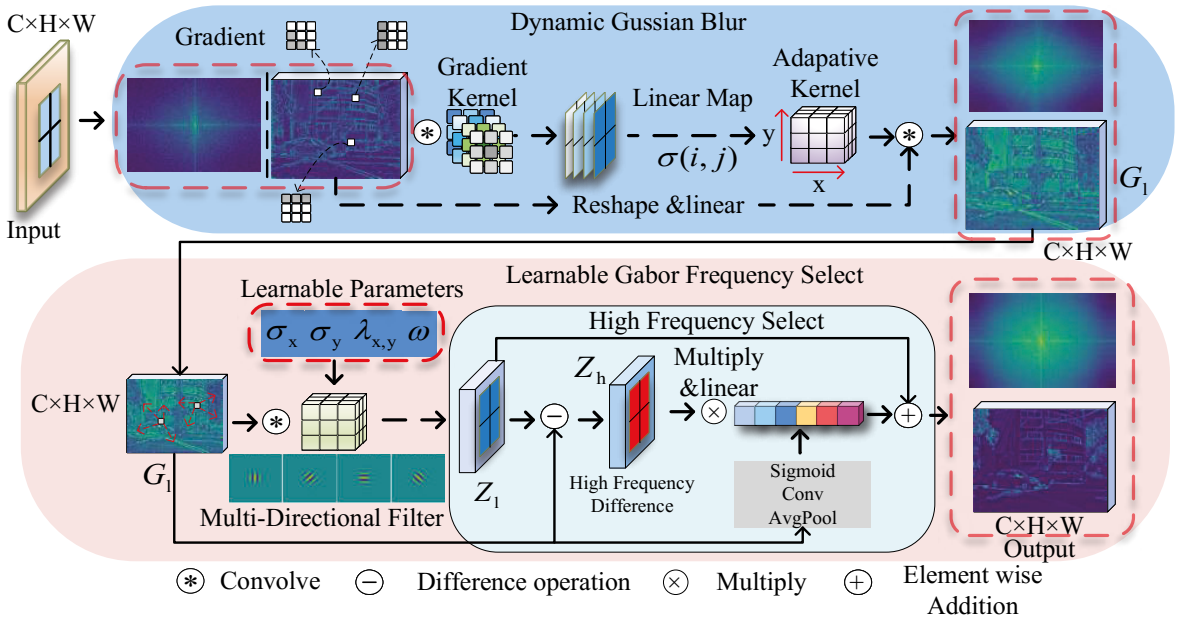


Figure 2: Architecture of the frequency aware filter.

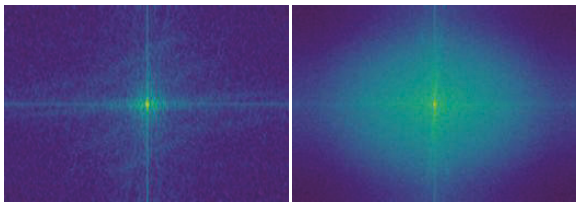


Figure 3: frequency visualisation of the DGK module.

where, $\circledast$ denotes convolution, and k is the size of the Gaussian kernel. Thus, DGK combines gradient-aware bandwidth control with edge-aligned kernel rotation to unify anti-aliasing and boundary preservation within a single lightweight operator.

As shown in Fig. 3, frequency spectra are presented. Before applying DGK, prominent green stripes and cross-shaped artifacts are visible, and high-frequency energy appears as scattered dots outside the Nyquist limit, indicating the presence of aliasing. After DGK processing, the peripheral high-frequency regions are significantly suppressed, with energy concentrated around the central low-frequency cross, demonstrating that DGK effectively attenuates high frequencies beyond the Nyquist range while preserving the dominant directional frequencies associated with object boundaries.

3.2 Learnable Gabor-based Frequency Selector

Although DGK effectively suppresses large-scale, spatially varying high-frequency noise, the directional sparsity of aliasing causes residual high-frequency energy to persist along specific orientations, thereby retaining phase distortion. These residuals manifest as sharp peaks in the frequency domain and spurious textures in the spatial domain. To suppress residual aliasing without losing edge detail, we introduce the Learnable Frequency Selector (LFS). As Fig.2 (bottom) shows, LFS combines a Gabor-inspired directional band-pass kernel with high-frequency residual gating, letting the network dynamically learn task-specific orientation and frequency responses.

Unlike anisotropic Gaussian filters that only adjust bandwidth, Gabor filters act as narrow band-pass filters in the frequency domain (see the left panel of Fig. 4(c)), where the central frequency $\lambda_{x,y}$ and orientation ω enable precise control over the preserved frequency components. The parameters σ_x and σ_y specify the filter's bandwidth in their respective directions. The learnable Gabor filter is defined as follows: For rotated local coordinates (x_ω, y_ω), the Gabor kernel is defined as

$$G_{\sigma_x,\sigma_y,\lambda_{x,y},\omega}(x,y) = \exp\Big(-\big(\tfrac{x_\omega^2}{2\sigma_x^2} + \tfrac{y_\omega^2}{2\sigma_y^2}\big)\Big) \times \cos\big(2\pi\lambda_{x,y}x_\omega\big), \tag{8}$$

Given the input feature map F_t, the low-frequency suppression (LFS) process is formulated as

$$Z_l = F_t \circledast G_{\sigma_x,\sigma_y,\lambda_{x,y},\omega}, \qquad Z_h = F_t - Z_l, \tag{9}$$

where $\circledast$ denotes convolution.

To adaptively select informative textures, we introduce a channel-attention mechanism:

$$S_l = \mathrm{Sigmoid}\Big(\mathrm{Conv}_{3\times3}\big(\mathrm{Avg}(F_t)\big)\Big), \tag{10}$$

and the final output is obtained by

$$O_l = Z_l + Z_h \otimes S_l. \tag{11}$$

where $\otimes$ indicates element-wise multiplication.

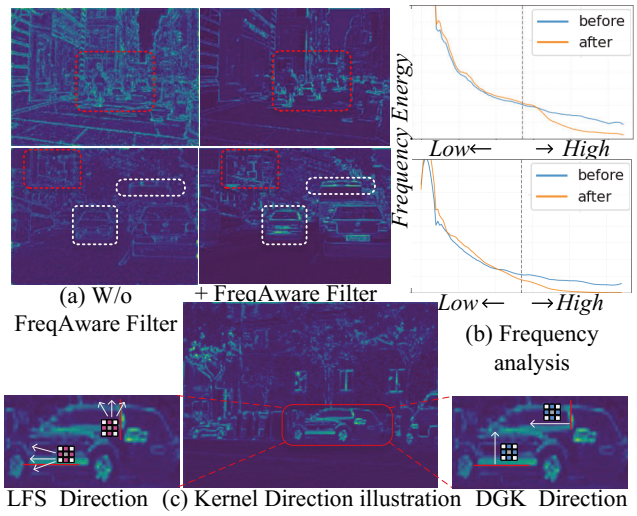


Figure 4: Visual effects of the frequency aware filter. (a) shows enhanced object boundaries. (b) illustrates the frequency magnitude before and after applying the frequency aware filter, where a notable reduction in high-frequency energy is observed. (c) shows the DGK and LFS kernel orientations are primarily aligned perpendicular and parallel to the object boundaries, respectively

As shown in Fig. 4(a), edges in the white box are broken before filtering. After applying Freq Aware Filter, contours become smoother. In the red box, noisy textures are smoothed In (b), the frequency curve shows a global reduction in high frequencies, indicating that DGK suppresses aliasing effectively. Meanwhile, edges remain enhanced in (a), suggesting that LFS restores useful edge frequencies. In (c), DGK filters perpendicular to edges for soft smoothing, while LFS aligns with edges for fine adjustment—together achieving anti-aliasing and boundary preservation.

3.3 Phase Error Energy

Aliasing refers to the folding of frequency components above the Nyquist limit into the lower frequency band after sampling, resulting in distortions in both magnitude and phase. Phase distortion,leads to pixel-level displacement that manifests in the spatial domain as boundary disappearance, misalignment, or false textures. This may also explain why hard pixels in segmentation are always located at the edges. To quantitatively evaluate the phase distortion caused by aliasing, we define the *PE* E_ϕ based on the Euclidean distance in the complex plane. This metric performs a mean-squared analysis over the phase differences at all significant frequency points (u, v), as shown in Equation (12):

$$E_\phi = \frac{1}{|\Omega|} \sum_{(u,v)\in\Omega} \left| F_{\text{alias}} - F_{\text{ideal}} \right|^2. \tag{12}$$

where F_{ideal} denotes the ideal frequency spectrum without alias, F_{alias} represents the actual spectrum affected by aliasing, and Ω denotes the set of significant frequency points. At each spatial frequency coordinate (u, v), we assume that the ideal and aliased signals share the same magnitude $A(u, v)$. F_{ideal} and F_{alias} are calculated as follows:

$$\begin{aligned} F_{\text{ideal}}(u,v) &= A\, e^{j\phi_{\text{ideal}}(u,v)}, \\ \Delta\phi &= \phi_{\text{alias}} - \phi_{\text{ideal}}, \\ F_{\text{alias}}(u,v) &= A\, e^{j\left(\phi_{\text{ideal}}(u,v)+\Delta\phi(u,v)\right)}, \end{aligned} \tag{13}$$

where $\Delta\phi$ denotes the phase error introduced by aliasing. Since phase distortion is the primary factor affecting segmentation in this context, we focus solely on the phase component in our analysis; the amplitude $A(u, v)$ is therefore treated as a constant when computing E_ϕ. When $\Delta\phi = 0$, $E_\phi = 0$, indicating the complete absence of phase distortion. Because E_ϕ compares image phases in the frequency domain, it provides a purely phase-based measure of pixel-location and structural misalignments, precisely revealing both the spatial distribution and directional characteristics of phase shifts. In experiments, E_ϕ shows a strong negative correlation with mIoU, confirming the impact of aliasing-induced phase distortion on segmentation performance.

4 Experiments

4.1 Metrics and Datasets

Following prior works (Cheng et al. 2022; Guo et al. 2022), we use mIoU to evaluate segmentation accuracy, and BioU and BAcc (Cheng et al. 2021) to assess boundary quality. We report PE introduced in Section 3 to measure phase correction effectiveness.We employed on widely used semantic segmentation benchmarks datasets: Cityscapes (Cordts et al. 2016), ADE20K (Zhou et al. 2017), and COCO-Stuff (Caesar, Uijlings, and Ferrari 2018).

4.2 Implementation Details

We integrated our method into existing segmentation frameworks including SegNext(Guo et al. 2022) , Mask2Former(Cheng et al. 2022) , and UperNet (Xiao et al. 2018), using their original training configurations. More details are described in the Appendix.

4.3 Ablation Study

Effectiveness of DGK and LFS: As shown in Table 1, using DGK alone improved mIoU by 1.3 and reduced PE by 0.02. LFS significantly improves boundary accuracy, lowering PE by 0.09 and increasing mIoU by 1.7. When combined, DGK and LFS achieve a 2.0 gain in mIoU and a 0.12 reduction in PE.More detail experiments are described in the Appendix.

SegNext-T	DGK	LFS	Params(M)	mIoU↑	PE↓
Base	×	×	4.26	41.1	0.31
Freq-Aware Filter	✓	×	4.33	41.9	0.24
	×	✓	4.45	42.3	0.22
	✓	✓	**4.52**	**42.9**	**0.19**

Table 1: Ablation study of DGK and LFS on the ADE20K validation set.

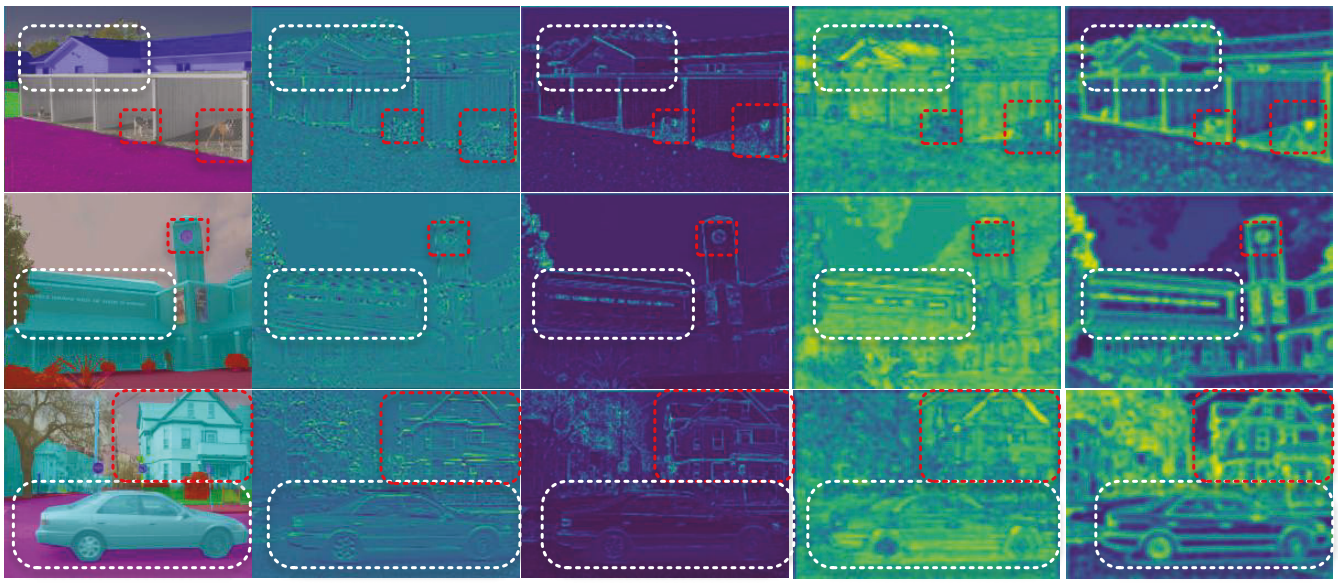


Figure 5: Visualization of the effects of the Frequency-Aware Filter. Red boxes highlight regions where fine textures disappear due to aliasing. White boxes indicate the emergence of spurious textures.

SegNext-T	DGK	LFS	mIoU↑	PE↓
Base	×	×	41.1	0.31
Freq-Aware Filter	K=3	K=3	42.3	0.27
	K=5	K=3	**42.9**	**0.19**
	K=7	K=5	42.7	0.23
	K=5	K=5	42.6	0.23

Table 2: Ablation study on kernel sizes of DGK and LFS on the ADE20K validation set.

SegNext-T	mIoU↑	PE↓
Base	41.1	0.31
Freq-AwareFilter$_1$	40.2	0.23
Freq-AwareFilter$_2$(Default)	**42.9**	**0.19**

Table 3: Ablation study on DGK/LFS order on the ADE20K.

Feature Map Visualization: As shown in Fig. 5, textures of small objects within red boxes vanish progressively because of aliasing, becoming nearly invisible at 4× downsampling . White boxes highlight spurious textures.After applying Freq Aware Filter, object boundaries become sharper and interiors cleaner. To further verify phase correction, we visualize phase spectra in Fig. 7. On the left, reveal strong aliasing patterns such as cross-shaped stripes and checkerboard noise. On the right, high-frequency noise is suppressed while important textures and edges remain well preserved. In particular, artifacts around tree trunks and object contours are removed after LFS, resulting in smoother, more continuous boundaries.

Effect of Kernel Size: As shown in Table 2, we analyse the impact of DGK and LFS kernel sizes on mIoU. Results show that the best performance is achieved with a kernel size of (5,3), which we adopt in our final configuration.

Effect of Module Order: As shown in Table 3, we compare two versions of the Freq Aware Filter. Freq Aware Filter$_1$ applies LFS before DGK, while Filter$_2$ follows the default order (DGK LFS). Reversing the order leads to a 0.9 drop in mIoU, although PE slightly improves by 0.8. This is because enhancing features without first low-pass filtering introduces a conflict between noise suppression and detail enhancement.

4.4 Comparison Study

Comparison with Anti-Aliasing Methods: As shown in Table 4, we compare our FreqAware Filter with leading anti-aliasing methods, including Blur (Zhang 2019), AdaBlur (Zou et al. 2023), FLC (Grabinski et al. 2022), and DAF (Chen, Gu, and Fu 2024). our method achieves significant improvements in BioU and BAcc, leading to the highest mIoU among all methods.

Methods	Params(M)	mIoU↑	BIoU↑	BAcc↑
UperNet[ECCV'18]	31.16	78.1	61.8	74.4
UperNet+Blur[ICML'19]	31.16	78.8	62.3	75.0
UperNet+AdaBlur[IJCV'23]	32.32	78.9	62.3	75.1
UperNet+FLC[ECCV'22]	31.16	78.6	62.7	74.9
UperNet+DAF[ICLR'24]	31.16	79.3	62.6	75.7
UperNet+Freq Aware Filter	31.42	**79.8**	**63.1**	**76.4**

Table 4: Comparison with state-of-the-art sampling methods on the Cityscapes validation set

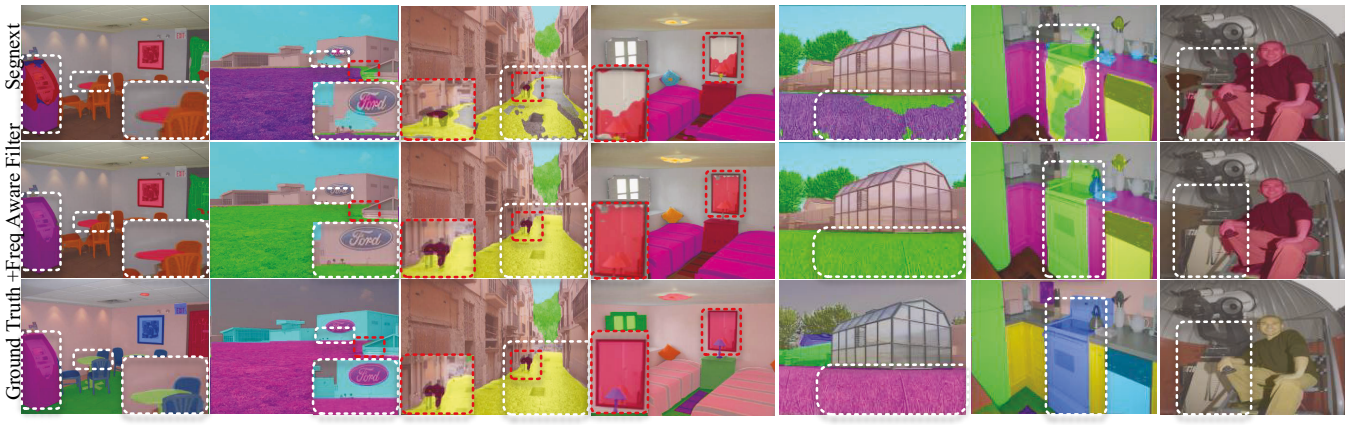
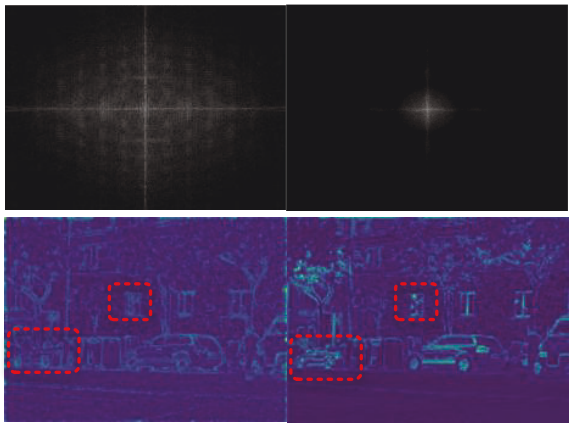


Figure 6: Visual results on the ADE20K dataset. The first row shows segmentation output from SegNext, the second row shows results after applying our method, and the third row displays the ground truth.



W/o Freq Aware Filter + Freq Aware Filter

Figure 7: Visualization of phase distortion. The first row shows the phase error spectra, and the second row displays the corresponding feature maps. Red boxes highlight regions where details are lost and subsequently recovered.

Comparison with State-of-the-Art: As shown in Table 5, integrating Freq Aware Filter into Mask2Former (Cheng et al. 2022) improves mIoU by 0.8 on the Cityscapes dataset, outperforming competing methods. Notably, despite using a ResNet-50 backbone, our method surpasses several ResNet-101-based models in accuracy.

Methods	Backbone	mIoU↑
AlignSeg[TPAMI'21]	ResNet-50	78.5
IFA[ECCV'22]	ResNet-50	78.0
SFNet[ICCV'23]	ResNet-50	79.2
FaPN[ICCV'21]	ResNet-50	80.0
Mask2former[CVPR'22]	ResNet-50	79.4
Mask2former[CVPR'22]	ResNet-101	80.1
Mask2former+FreqAwareFilter	ResNet-50	**80.2**

Table 5: Comparison with state-of-the-art methods on the Cityscapes validation set.

Multi-Dataset Evaluation: As shown in Table 6, our method improves mIoU by 1.8 on ADE20K , 0.8 on Cityscapes , and 0.7 on COCO-Stuff .

Methods	ADE20K mIoU↑	Cityscapes mIoU↑	COCO-Stuff mIoU↑
SegNext-T	41.1	79.8	38.7
+Freq Aware Filter	**42.9**	**80.6**	**40.4**

Table 6: Comparison across ADE20K, Cityscapes, and COCO-Stuff datasets.

Qualitative Results: Fig. 6 shows that our method significantly improves segmentation quality by mitigating phase distortion issues such as incomplete object masks, false texture-induced splits, and blurry boundaries.

5 Conclusion

In this work, we aim to address phase distortion introduced by downsampling in semantic segmentation. Through feature-level analysis, we identify two key properties of phase distortion: spatial non-uniformity and directional sparsity. We introduce the Phase Error Energy metric to quantitatively measure phase distortion. These insights guide the design of our Frequency-Aware Filter, which consists of a Dynamic Gaussian Kernel and a Learnable Gabor-based Frequency Selector. These components jointly suppress irrelevant high-frequency noise, smooth intra-class textures, and selectively preserve directional high-frequency boundaries. This effectively mitigates issues such as incomplete segmentation and missing object structures caused by aliasing. Both qualitative and quantitative results demonstrate the effectiveness of Freq Aware Filter in semantic segmentation. The insights gained and the proposed metric offer new perspectives for future research in this area and may inspire broader applications beyond segmentation.

Acknowledgments

We thank the anonymous reviewers for their valuable feedback. This work was partly supported by the Natural Science Foundation of China (62302466), Natural Science Basic Research Program of Shanxi Province (202303021212188).

References

Bajcsy, A.; Loquercio, A.; Kumar, A.; and Malik, J. 2024. Learning vision-based pursuit-evasion robot policies. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 9197–9204. IEEE.

Caesar, H.; Uijlings, J.; and Ferrari, V. 2018. Coco-stuff: Thing and stuff classes in context. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1209–1218.

Chen, L.; Fu, Y.; Gu, L.; Yan, C.; Harada, T.; and Huang, G. 2024a. Frequency-aware feature fusion for dense image prediction. *IEEE transactions on pattern analysis and machine intelligence*.

Chen, L.; Fu, Y.; Wei, K.; Zheng, D.; and Heide, F. 2023. Instance segmentation in the dark. *International Journal of Computer Vision*, 131(8): 2198–2218.

Chen, L.; Gu, L.; and Fu, Y. 2024. When semantic segmentation meets frequency aliasing. *arXiv preprint arXiv:2403.09065*.

Chen, L.; Gu, L.; Li, L.; Yan, C.; and Fu, Y. 2025. Frequency Dynamic Convolution for Dense Image Prediction. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 30178–30188.

Chen, L.; Gu, L.; Zheng, D.; and Fu, Y. 2024b. Frequency-adaptive dilated convolution for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3414–3425.

Chen, L.; Wu, P.; Chitta, K.; Jaeger, B.; Geiger, A.; and Li, H. 2024c. End-to-end autonomous driving: Challenges and frontiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Cheng, B.; Girshick, R.; Dollár, P.; Berg, A. C.; and Kirillov, A. 2021. Boundary IoU: Improving object-centric image segmentation evaluation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 15334–15342.

Cheng, B.; Misra, I.; Schwing, A. G.; Kirillov, A.; and Girdhar, R. 2022. Masked-attention mask transformer for universal image segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1290–1299.

Cordts, M.; Omran, M.; Ramos, S.; Rehfeld, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; and Schiele, B. 2016. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3213–3223.

Di Palo, N.; and Johns, E. 2024. Dinobot: Robot manipulation via retrieval and alignment with vision foundation models. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2798–2805. IEEE.

Fuoli, D.; Van Gool, L.; and Timofte, R. 2021. Fourier space losses for efficient perceptual image super-resolution. In *Proceedings of the IEEE/CVF international conference on computer vision*, 2360–2369.

Grabinski, J.; Jung, S.; Keuper, J.; and Keuper, M. 2022. Frequencylowcut pooling-plug and play against catastrophic overfitting. In *European Conference on Computer Vision*, 36–57. Springer.

Guo, M.-H.; Lu, C.-Z.; Hou, Q.; Liu, Z.; Cheng, M.-M.; and Hu, S.-M. 2022. Segnext: Rethinking convolutional attention design for semantic segmentation. *Advances in neural information processing systems*, 35: 1140–1156.

Hossain, M. T.; Teng, S. W.; Lu, G.; Rahman, M. A.; and Sohel, F. 2023. Anti-aliasing deep image classifiers using novel depth adaptive blurring and activation function. *Neurocomputing*, 536: 164–174.

Huang, J.; Liu, Y.; Zhao, F.; Yan, K.; Zhang, J.; Huang, Y.; Zhou, M.; and Xiong, Z. 2022. Deep fourier-based exposure correction network with spatial-frequency interaction. In *European Conference on Computer Vision*, 163–180. Springer.

Jiang, B.; Chen, S.; Xu, Q.; Liao, B.; Chen, J.; Zhou, H.; Zhang, Q.; Liu, W.; Huang, C.; and Wang, X. 2023. Vad: Vectorized scene representation for efficient autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 8340–8350.

Karras, T.; Aittala, M.; Laine, S.; Härkönen, E.; Hellsten, J.; Lehtinen, J.; and Aila, T. 2021. Alias-free generative adversarial networks. *Advances in neural information processing systems*, 34: 852–863.

Ma, J.; He, Y.; Li, F.; Han, L.; You, C.; and Wang, B. 2024. Segment anything in medical images. *Nature Communications*, 15(1): 654.

Milioto, A.; and Stachniss, C. 2019. Bonnet: An open-source training and deployment framework for semantic segmentation in robotics using cnns. In *2019 international conference on robotics and automation (ICRA)*, 7094–7100. IEEE.

Nyquist, H. 1928. Abridgment of certain topics in telegraph transmission theory. *Journal of the AIEE*, 47(3): 214–217.

Qian, S.; Shao, H.; Zhu, Y.; Li, M.; and Jia, J. 2021. Blending anti-aliasing into vision transformer. *Advances in Neural Information Processing Systems*, 34: 5416–5429.

Qin, Z.; Zhang, P.; Wu, F.; and Li, X. 2021. Fcanet: Frequency channel attention networks. In *Proceedings of the IEEE/CVF international conference on computer vision*, 783–792.

Shannon, C. E. 2006. Communication in the presence of noise. *Proceedings of the IRE*, 37(1): 10–21.

Specktor-Fadida, B.; Ben-Sira, L.; Ben-Bashat, D.; and Joskowicz, L. 2025. SegQC: a segmentation network-based framework for multi-metric segmentation quality control and segmentation error detection in volumetric medical images. *Medical Image Analysis*, 103638.

Wang, L.; You, Z.-H.; Lu, W.; Chen, S.-B.; Tang, J.; and Luo, B. 2024. Attention-aware sobel graph convolutional

network for remote sensing image change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–12.

Wang, W.; Dai, J.; Chen, Z.; Huang, Z.; Li, Z.; Zhu, X.; Hu, X.; Lu, T.; Lu, L.; Li, H.; et al. 2023. Internimage: Exploring large-scale vision foundation models with deformable convolutions. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 14408–14419.

Wu, J.; Wang, Z.; Hong, M.; Ji, W.; Fu, H.; Xu, Y.; Xu, M.; and Jin, Y. 2025. Medical sam adapter: Adapting segment anything model for medical image segmentation. *Medical image analysis*, 102: 103547.

Xiao, T.; Liu, Y.; Zhou, B.; Jiang, Y.; and Sun, J. 2018. Unified perceptual parsing for scene understanding. In *Proceedings of the European conference on computer vision (ECCV)*, 418–434.

Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J. M.; and Luo, P. 2021. SegFormer: Simple and efficient design for semantic segmentation with transformers. *Advances in neural information processing systems*, 34: 12077–12090.

Xiong, Y.; Li, Z.; Chen, Y.; Wang, F.; Zhu, X.; Luo, J.; Wang, W.; Lu, T.; Li, H.; Qiao, Y.; et al. 2024. Efficient deformable convnets: Rethinking dynamic and sparse operator for vision applications. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5652–5661.

Yin, D.; Gontijo Lopes, R.; Shlens, J.; Cubuk, E. D.; and Gilmer, J. 2019. A fourier perspective on model robustness in computer vision. *Advances in Neural Information Processing Systems*, 32.

Yu, H.; Yang, W.; Zhong, J.; Yang, Z.; Fan, S.; Luo, P.; and Nie, Z. 2025. End-to-end autonomous driving through v2x cooperation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 9598–9606.

Zhang, F.; Panahi, A.; and Gao, G. 2023. FsaNet: Frequency self-attention for semantic segmentation. *IEEE Transactions on Image Processing*, 32: 4757–4772.

Zhang, R. 2019. Making convolutional networks shift-invariant again. In *International conference on machine learning*, 7324–7334. PMLR.

Zhao, W.; Zhang, R.; Wang, Q.; Cheng, G.; and Huang, K. 2025. BFANet: Revisiting 3D Semantic Segmentation with Boundary Feature Analysis. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 29395–29405.

Zhou, B.; Zhao, H.; Puig, X.; Fidler, S.; Barriuso, A.; and Torralba, A. 2017. Scene parsing through ade20k dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 633–641.

Zhou, T.; and Wang, W. 2024. Cross-image pixel contrasting for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(8): 5398–5412.

Zhou, Y.; Li, L.; Wang, C.; Song, L.; and Yang, G. 2024. Gobletnet: Wavelet-based high-frequency fusion network for semantic segmentation of electron microscopy images. *IEEE Transactions on Medical Imaging*.

Zou, X.; Xiao, F.; Yu, Z.; Li, Y.; and Lee, Y. J. 2023. Delving deeper into anti-aliasing in convnets. *International Journal of Computer Vision*, 131(1): 67–81.