北京理工大學
BEIJING INSTITUTE OF TECHNOLOGY

Frequency-aware Feature Fusion for Dense Image Prediction

Code Paper Github Google Scholar

Linwei Chen 陈林蔚 Ph.D

charleschen2013218@gmail.com

Supervisor: Prof. Ying Fu（付莹）

Team and collaborators

北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY

Research Team

Team Leader: Prof. Jun Zhang（张军）

Academician of the Chinese Academy of Engineering, Distinguished Professor of the Ministry of Education, National-level Leading Talent, and Deputy Secretary of the Party Committee of the Council of the Chinese Institute of Electronics.

Supervisor: Prof. Ying Fu（付莹）

Professor and Ph.D. advisor, National-level high-caliber young talent and a recipient of the Shi Qingyun Award for Women Scientists from the China Society of Image and Graphics.

Collaborators

- YOU Shaodi
  Associate Professor, Universit…
- Kaixuan Wei
  KAUST
- Lin Gu
  Research Scientist, RIKEN AI…
- Felix Heide
  Princeton University | Torc Ro…
- Chenggang Yan
  Hangzhou Dianzi University
- Tatsuya Harada
  The University of Tokyo
- Gao Huang （黄高）
  Associate Professor, Tsinghua…

- Felix Heide - Princeton University
- Tatsuya Harada - The University of Tokyo
- Lin Gu - The University of Tokyo / RIKEN AIP
- Shaodi You - University of Amsterdam
- Gao Huang - Tsinghua University
- Chenggang Yan - Hangzhou Dianzi University
- Kaixuan Wei - King Abdullah University of Science and Technology

Academic profile

北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY

Google Scholar

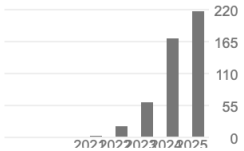
Linwei Chen（陈林蔚） 已关注

Beijing Institute of Technology
在 bit.edu.cn 的电子邮件经过验证 - 首页
Computer Vision Artificial Intelligence

标题	引用次数	年份
LE-GAN: Unsupervised low-light image enhancement network using attention module and identity invariant loss Y Fu, Y Hong, L Chen, S You Knowledge-Based Systems 240, 108010	116	2022
Instance Segmentation in the Dark L Chen, Y Fu, K Wei, D Zheng, F Heide IJCV 2023, CCF-A, https://github.com/Linwei-Chen/LIS 131 (8), 2198–2218	83	2023
Crafting Object Detection in Very Low Light. Y Hong, K Wei, L Chen, Y Fu BMVC 1 (2), 1-15	66	2021
Frequency-aware feature fusion for dense image prediction L Chen, Y Fu, L Gu, C Yan, T Harada, G Huang TPAMI 2024, CCF-A, https://github.com/Linwei-Chen/FreqFusion 46 (12), 10763 …	62	2024
Frequency-adaptive dilated convolution for semantic segmentation L Chen, L Gu, D Zheng, Y Fu CVPR 2024 Highlighted, Top 2.8%, CCF-A, 3414-3425	50	2024
Efficient hybrid supervision for instance segmentation in aerial images L Chen, Y Fu, S You, H Liu Remote Sensing 13 (2), 252	24	2021
A large-scale climate-aware satellite image dataset for domain adaptive land-cover semantic segmentation S Liu, L Chen, L Zhang, J Hu, Y Fu ISPRS Journal of Photogrammetry and Remote Sensing 205, 98-114	15	2023
Hybrid supervised instance segmentation by learning label noise suppression L Chen, Y Fu, S You, H Liu Neurocomputing 496, 131-146	15	2022
When Semantic Segmentation Meets Frequency Aliasing L Chen, L Gu, Y Fu ICLR 2024, https://github.com/Linwei-Chen/Seg-Aliasing	14	2024
Level-aware consistent multilevel map translation from satellite imagery Y Fu, Z Fang, L Chen, T Song, D Lin IEEE Transactions on Geoscience and Remote Sensing 61, 1-14	10	2022
Consistency-aware map generation at multiple zoom levels using aerial image L Chen, Z Fang, Y Fu IEEE Journal of Selected Topics in Applied Earth Observations and Remote …	9	2022
Frequency Dynamic Convolution for Dense Image Prediction L Chen, L Gu, L Li, C Yan, Y Fu CVPR 2025, final score=554 (5=full), CCF-A, https://github.com/Linwei-Chen …	3	2025
Spatial Frequency Modulation for Semantic Segmentation L Chen, Y Fu, L Gu, D Zheng, J Dai IEEE Transactions on Pattern Analysis and Machine Intelligence 1 (1), 1-18		2025
Frequency-Dynamic Attention Modulation for Dense Prediction L Chen, L Gu, Y Fu International Conference on Computer Vision (ICCV), 1-13		2025

引用次数

	总计	2020 年至今
引用	467	467
h 指数	10	10
i10 指数	10	10



开放获取的出版物数量 查看全部

1 篇文章 9 篇文章

无法查看的文章 可查看的文章

根据资助方的强制性开放获取政策

合著作者 修改

- Ying Fu, Beijing Institute of Tech…
- Lin Gu, Research Scientist, RI…
- YOU Shaodi, Associate Professor, U…
- Kaixuan Wei, KAUST
- Chenggang Yan, Hangzhou Dianzi Univ…
- Felix Heide, Princeton University | T…
- Tatsuya Harada, The University of Tokyo
- Gao Huang （黄高）, Associate Professor, T…

- Published 10+ papers, including 8 first-author papers in top-tier journals and CCF-A conferences like TPAMIx2, IJCV, ICLR, CVPR Highlight (Top 2.8%), CVPR, ICCV and ISPRS.
- 5 pending national patents, with 2 already granted.
- CVPR 2024 PBDL Low-Light Detection and Segmentation Challenge: 2nd Place.
- PLA Rocket Force Intelligent Fire Eye National Competition: Excellence Award.
- Best Reviewer at BMVC 2024.

Highlight: TPAMI FreqFusion

北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY

Our GitHub repository has already received over 400 stars

Journals & Magazines > IEEE Transactions on Pattern ... > Volume: 46 Issue: 12

Frequency–Aware Feature Fusion for Dense Image Prediction

Publisher: IEEE Cite This PDF

Linwei Chen ; Ying Fu ; Lin Gu ; Chenggang Yan ; Tatsuya Harada ; Gao Huang **All Authors**

19 Cites in Papers

3870 Full Text Views

Abstract

Document Sections

I. Introduction

II. Related Work

III. Feature Similarity Analysis Metrics

IV. Frequency–Aware Feature Fusion

V. Experimental Results

Show Full Outline

Authors

Figures

References

Citations

Keywords

Metrics

Footnotes

Abstract:
Dense image prediction tasks demand features with strong category information and precise spatial boundary details at high resolution. To achieve this, modern hierarchical models often utilize feature fusion, directly adding upsampled coarse features from deep layers and high–resolution features from lower levels. In this paper, we observe rapid variations in fused feature values within objects, resulting in intra–category inconsistency due to disturbed high–frequency features. Additionally, blurred boundaries in fused features lack accurate high frequency, leading to boundary displacement. Building upon these observations, we propose Frequency–Aware Feature Fusion (FreqFusion), integrating an Adaptive Low–Pass Filter (ALPF) generator, an offset generator, and an Adaptive High–Pass Filter (AHPF) generator. The ALPF generator predicts spatially–variant low–pass filters to attenuate high–frequency components within objects, reducing intra–class inconsistency during upsampling. The offset generator refines large inconsistent features and thin boundaries by replacing inconsistent features with more consistent ones through resampling, while the AHPF generator enhances high–frequency detailed boundary information lost during downsampling. Comprehensive visualization and quantitative analysis demonstrate that FreqFusion effectively improves feature consistency and sharpens object boundaries. Extensive experiments across various dense prediction tasks confirm its effectiveness.

Published in: IEEE Transactions on Pattern Analysis and Machine Intelligence (Volume: 46 , Issue: 12, December 2024)

Page(s): 10763 – 10780

DOI: 10.1109/TPAMI.2024.3449959

Date of Publication: 26 August 2024

Publisher: IEEE

ISSN Information:

PubMed ID: 39186415

Funding Agency:

FreqFusion Public

Unpin | Unwatch 2 | Fork 21 | Starred 420

main | 1 Branch | 0 Tags | Go to file | Add file | Code

Linwei-Chen ... fb5542f · 5 months ago 48 Commits

Mask2Former	...	10 months ago
README.assets	...	5 months ago
SegNeXt	...	10 months ago
feat_tool	...	5 months ago
mmdetection	...	10 months ago
FreqFusion.py	...	9 months ago
README.md	...	5 months ago

README

FreqFusion

TPAMI 2024：Frequency-aware Feature Fusion for Dense Image Prediction

The preliminary official implementation of our TPAMI 2024 paper "Frequency-aware Feature Fusion for Dense Image Prediction" (IEEE arxiv) , which is also available at https://github.com/ying-fu/FreqFusion.

Interested readers are also referred to an insightful Note about this work in Zhihu (TODO).

About

TPAMI：Frequency-aware Feature Fusion for Dense Image Prediction

Readme

Activity

420 stars

2 watching

21 forks

Releases

No releases published

Create a new release

Packages

No packages published

Publish your first package

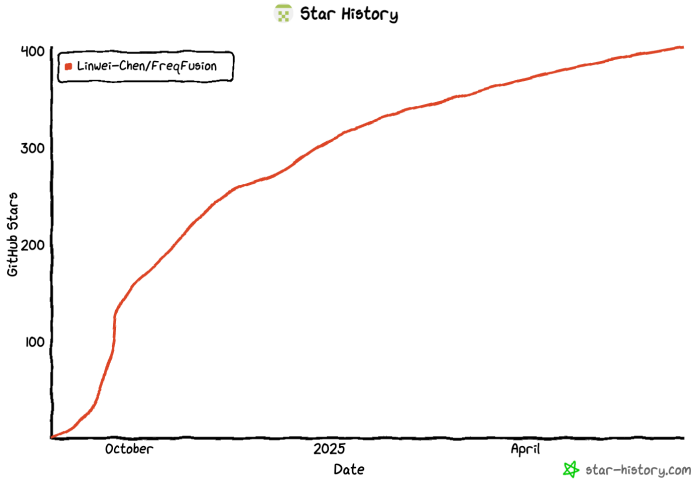
Languages

Jupyter Notebook 79.1%

Python 20.6% Cuda 0.2%

Shell 0.1% C++ 0.0%

Dockerfile 0.0%



Highlight: TPAMI FreqFusion

北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY

- **Citation count reached the Top ~1.3%, Google Scholar: Cited 63**
- **Influencing subsequent research in the field.**

Browse Journals & Magazines > IEEE Transactions on Pattern Analysis and Machine Intelligence

IEEE Transactions on Pattern Analysis and Machine Intelligence

20.8	0.09432	6.861	28.4
Impact Factor	Eigenfactor	Article Influence Score	CiteScore Powered by Scopus

Home | Popular | Early Access | Current Issue | All Issues | About Journal

Issue 12 • Dec.-2024 | Get Entire Issue Now Purchase or Sign in | Back to navigation

Search within results | Items Per Page | Export | Email Selected Results

Showing 26–50 of 256

Author | Affiliation | Supplemental Items

Select All on Page | Sort By Most Cited [By Papers]

Fast-BEV: A Fast and Strong Bird's-Eye View Perception Baseline
Yangguang Li; Bin Huang; Zeren Chen; Yufeng Cui; Feng Liang; Mingzhu Shen; Fenggang Liu; Enze Xie; Lu Sheng; Wanli Ouyang; Jing Shao
Publication Year: 2024 , Page(s): 8665 – 8679
Cited by: Papers (19)
Abstract HTML

Frequency-Aware Feature Fusion for Dense Image Prediction
Linwei Chen; Ying Fu; Lin Gu; Chenggang Yan; Tatsuya Harada; Gao Huang
Publication Year: 2024 , Page(s): 10763 – 10780
Cited by: Papers (19)
Abstract HTML

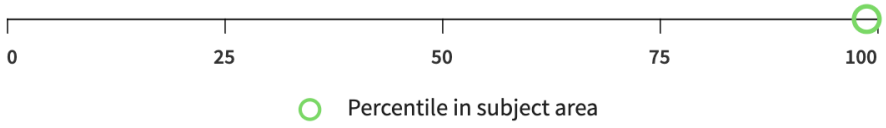
Secrets of Event-Based Optical Flow, Depth and Ego-Motion Estimation by Contrast Maximization
Shintaro Shiba; Yannick Klose; Yoshimitsu Aoki; Guillermo Gallego
Publication Year: 2024 , Page(s): 7742 – 7759
Cited by: Papers (19)
Abstract HTML

In the category

Percentile in subject area

98.76

The percentage of documents of the same type, from the same category and published in the same year that have a lower citation count than this paper.



Category Normalized Citation Impact

7.40 Above average

Category Normalized Citation Impact (CNCI) is the ratio of a document's actual times cited count to the expected count for a document of the same type, from the same category, and published in the same year. If the ratio is above 1, then the document's citation performance is above average.

Category expected citations

2.97

The expected number of citations calculated from other items of the same type (article, review, etc.) published in the same subject categories and year.

Highlight: TPAMI FreqFusion

北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY

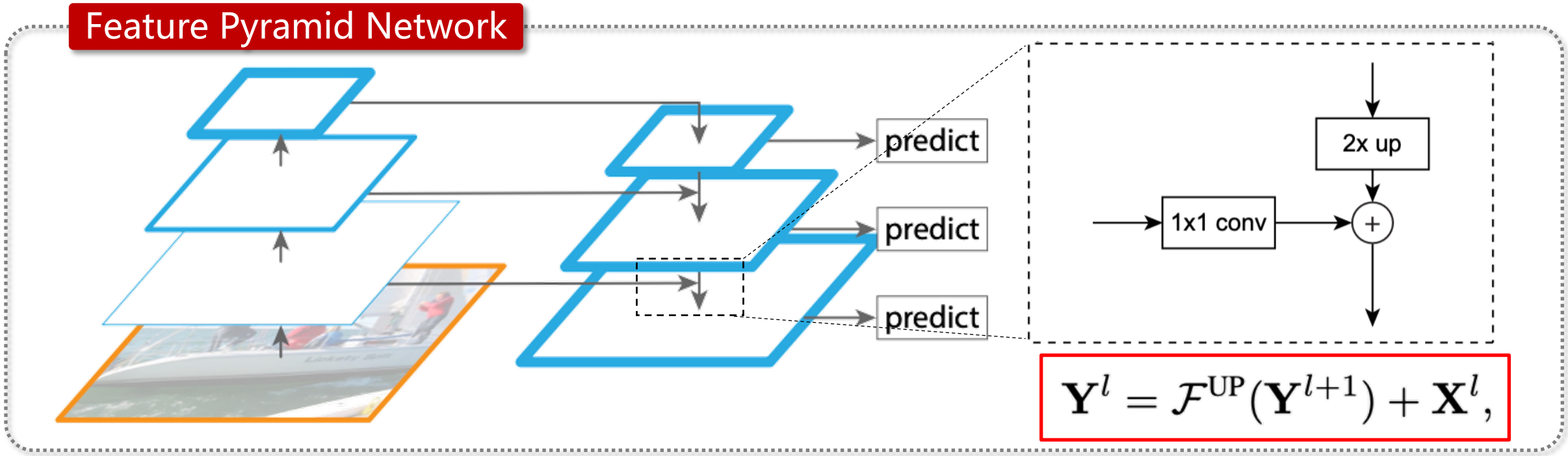
- **Discussions on social medias such as Zhihu, Bilibili, Xiaohongshu, indicating strong interest from the community.**

TPAMI 2024 FreqFusion

北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY

Research Background

- Feature fusion is fundamental and widely-used in modern vision models
- It integrates low-level details and high-level semantic, such as **FPN**



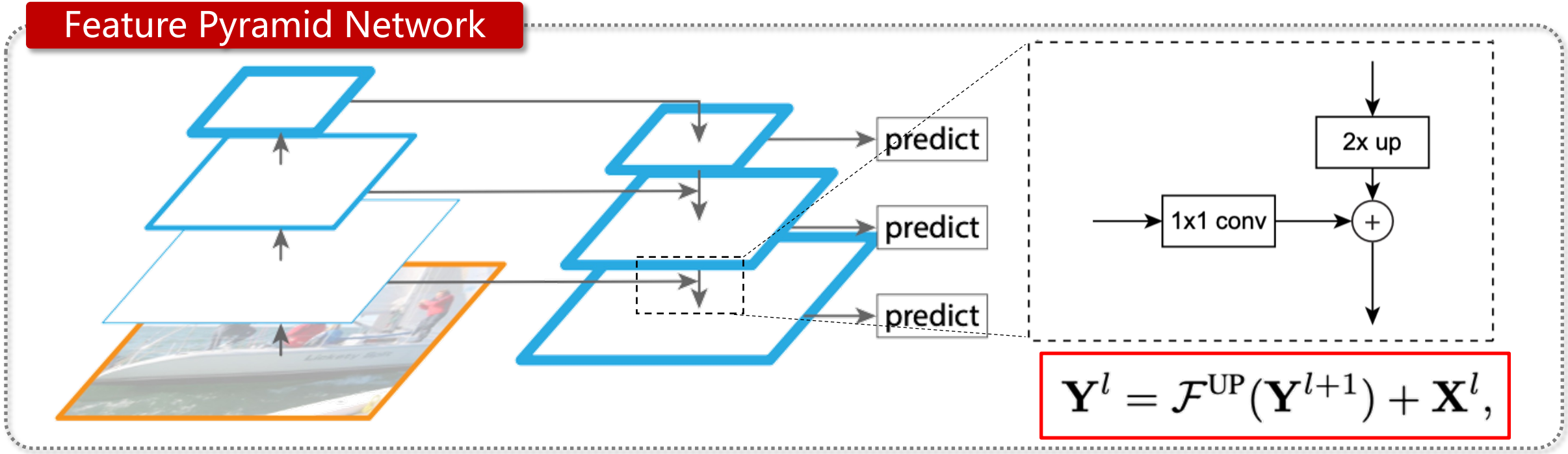
$$\mathbf{Y}^l = \mathcal{F}^{\mathrm{UP}}(\mathbf{Y}^{l+1}) + \mathbf{X}^l,$$

TPAMI 2024 FreqFusion

北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY

Research Background

- It involves two main steps: Upsampling and Element-wise Add/Concat
- The upsampling ensure the low&high level feature have the same size



$$\mathbf{Y}^l = \mathcal{F}^{\mathrm{UP}}(\mathbf{Y}^{l+1}) + \mathbf{X}^l,$$

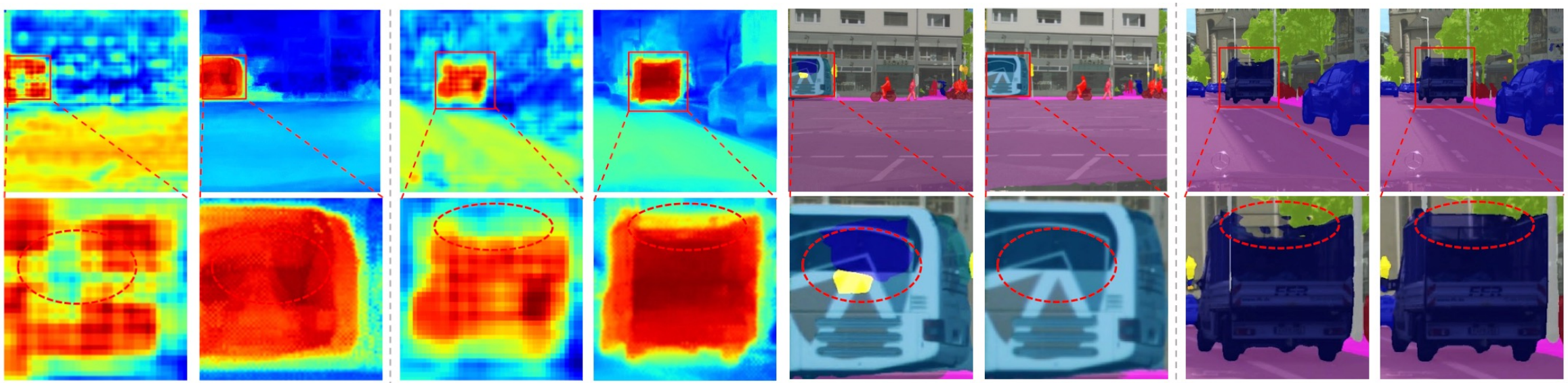
TPAMI 2024 FreqFusion

北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY

Scientific Question

- Details from low-level and semantic info from high-level are crucial
- Directly fusion leads to intra-class inconsistency&boundary displacement

Feature Visualization



Standard fusion FreqFusion Standard fusion FreqFusion Standard fusion FreqFusion Standard fusion FreqFusion

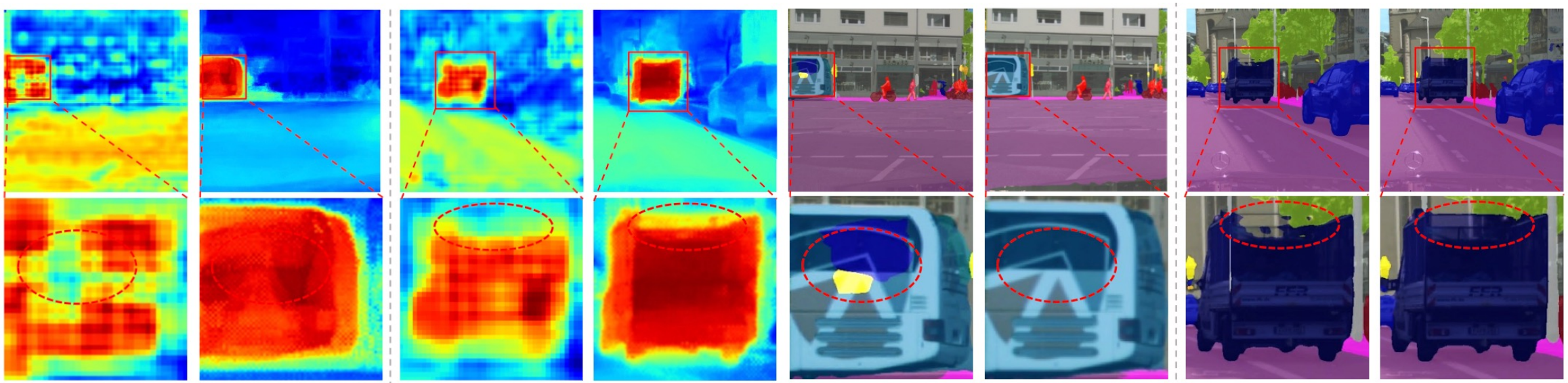
TPAMI 2024 FreqFusion

北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY

Scientific Question

- **Intra-class inconsistency** occurs when features belonging to the same object are not uniform. Inside the car, the features are noisy and contain unnecessary high-frequency components, leading to inconsistency.

Feature Visualization



Standard fusion FreqFusion Standard fusion FreqFusion Standard fusion FreqFusion Standard fusion FreqFusion

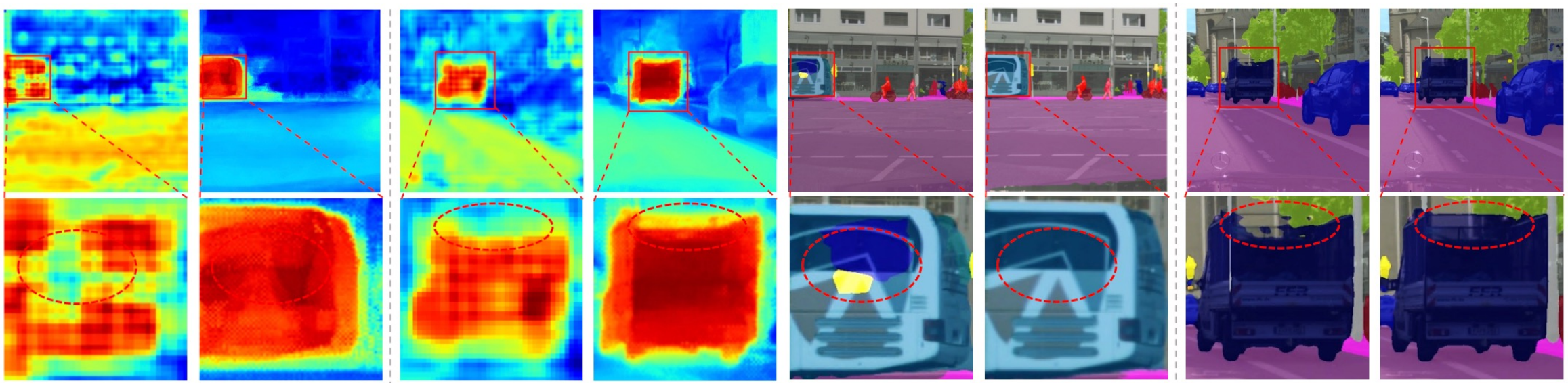
TPAMI 2024 FreqFusion

北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY

Scientific Question

- **Boundary displacement** occurs when object edges blur and are inaccurately localized, reflecting a lack of the high-frequency components necessary for precise prediction.

Feature Visualization



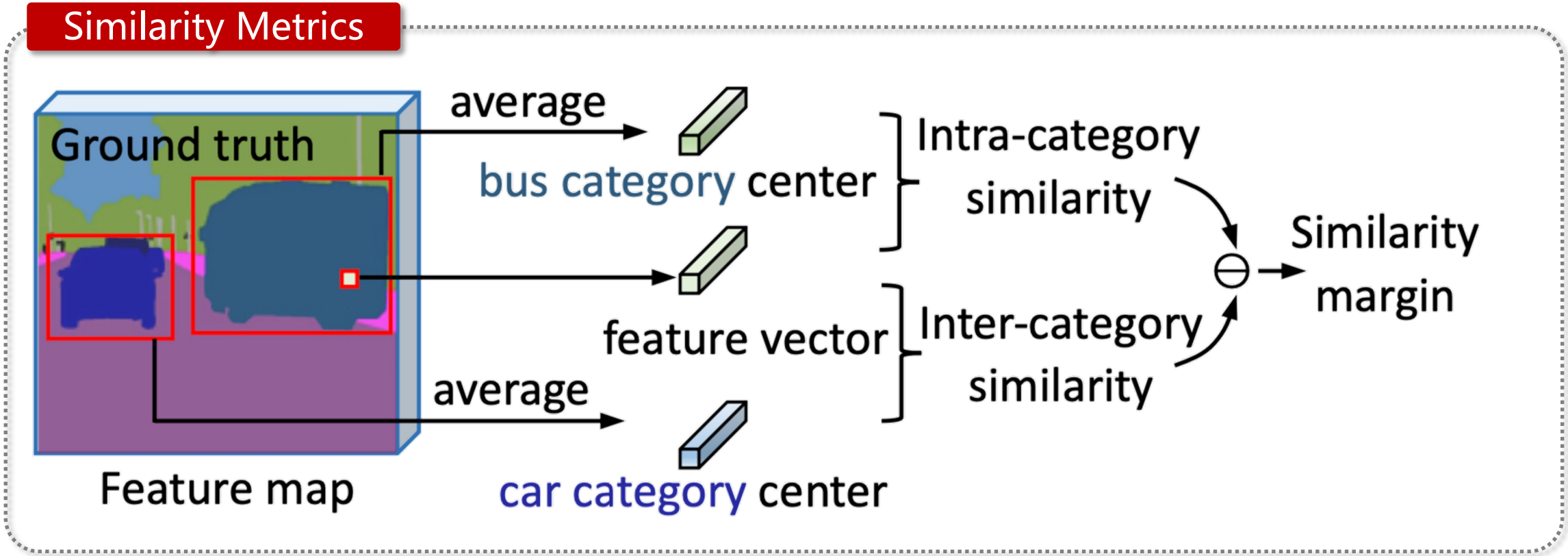
Standard fusion FreqFusion Standard fusion FreqFusion Standard fusion FreqFusion Standard fusion FreqFusion

TPAMI 2024 FreqFusion

北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY

How to quantify inconsistency & displacement?

- We develope a set of similarity metrics
- Including Intra-category/Inter-category Similarity, Similarity Margin

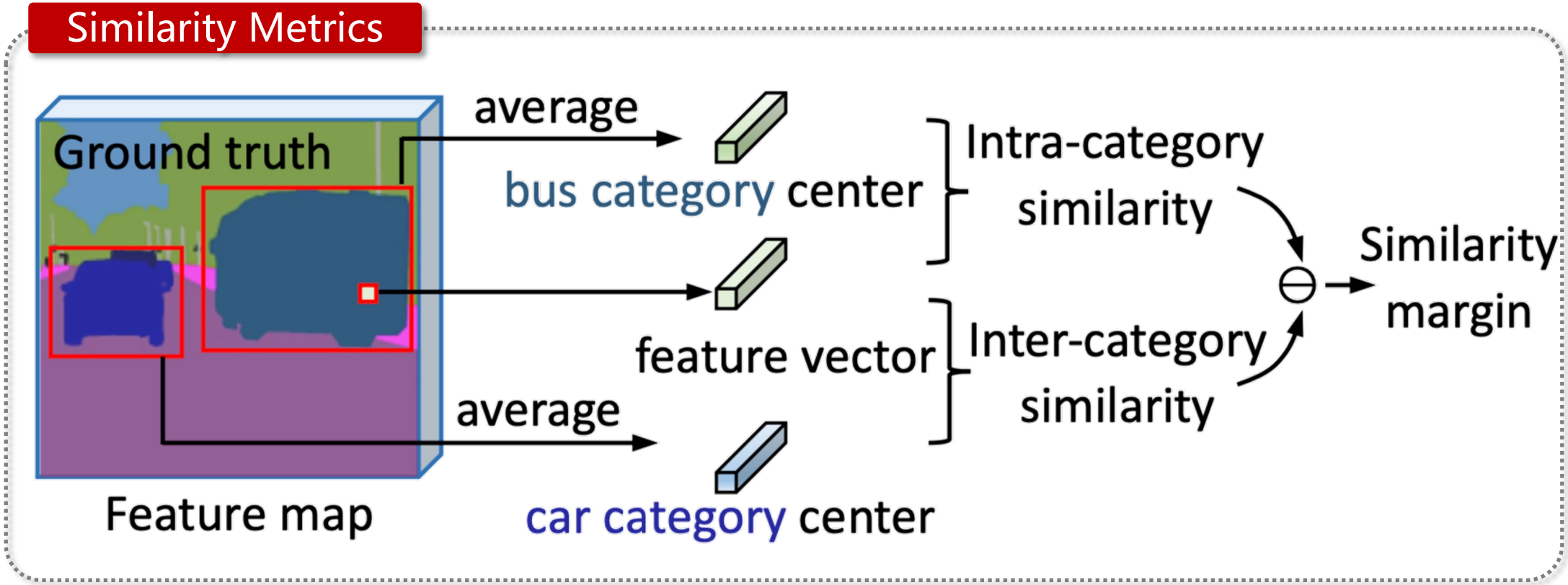


TPAMI 2024 FreqFusion

北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY

How to quantify inconsistency & displacement?

- IntraSim measures how similar a feature is to its own category
- Category center is the the average feature of its own class

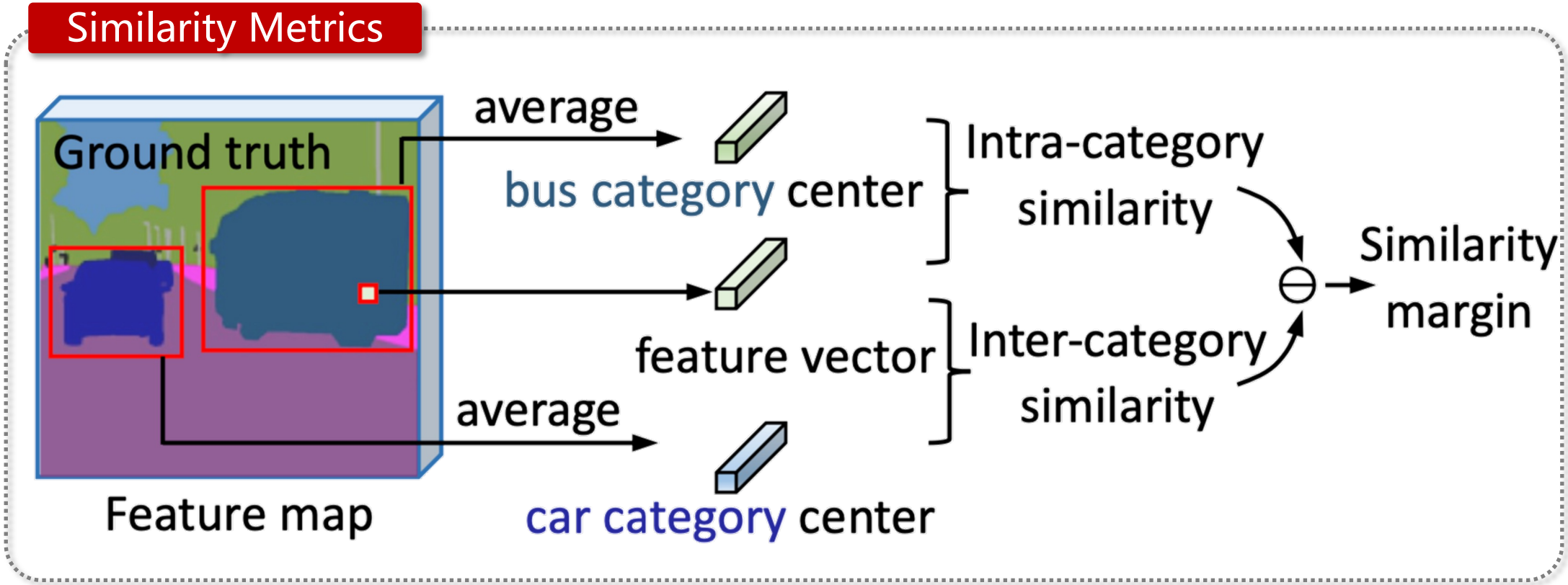


TPAMI 2024 FreqFusion

北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY

How to quantify inconsistency & displacement?

- InterSim measures how similar a feature is to different category
- The difference between these two gives us the Similarity Margin

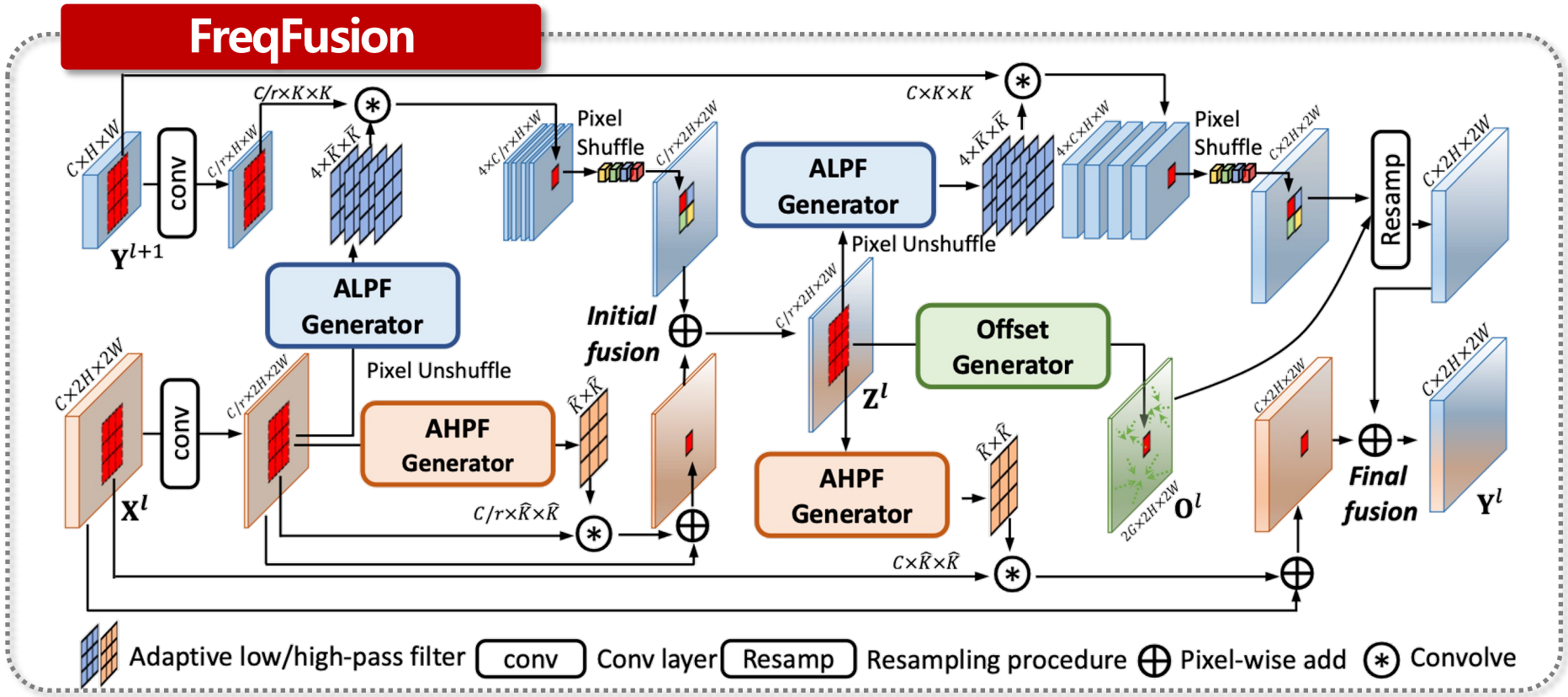


TPAMI 2024 FreqFusion

北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY

Proposed Method: FreqFusion

- To solve inconsistency and displacement, we propose FreqFusion
- It is a **light-weight**, **pretrain-free**, **pulg-in** module (*one line code*)



3 key components

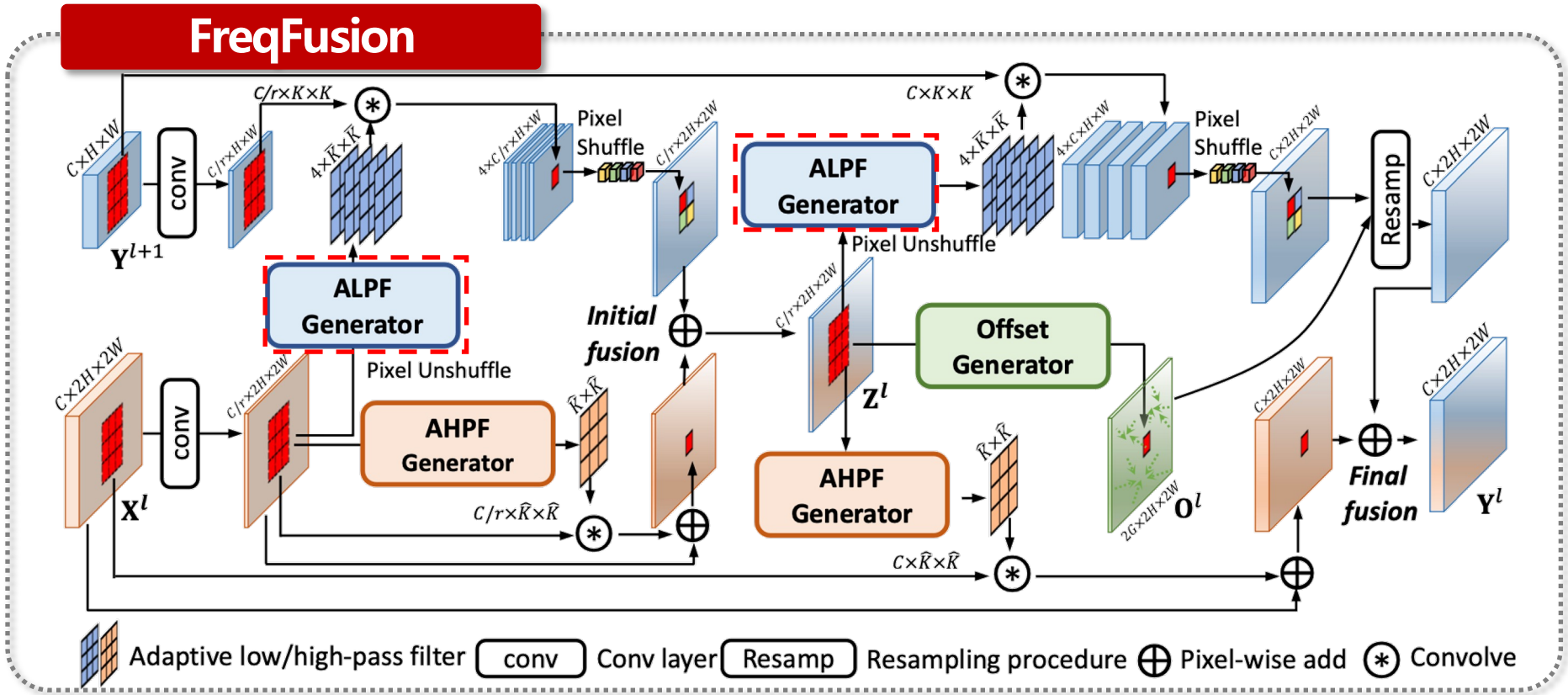
1. Adaptive Low-Pass Filter (ALPF) Generator
2. Offset Generator
3. Adaptive High-Pass Filter (AHPF) Generator

TPAMI 2024 FreqFusion

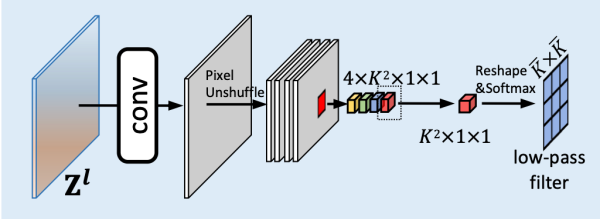
北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY

Adaptive Low-Pass Filter (ALPF) Generator

- We observed that intra-class inconsistency often appears as unnecessary high-frequency componets within an object's feature representation



Generation of ALPF:



$$\bar{\mathbf{V}}^l = \mathrm{Conv}_{3\times3}(\mathbf{Z}^l),$$

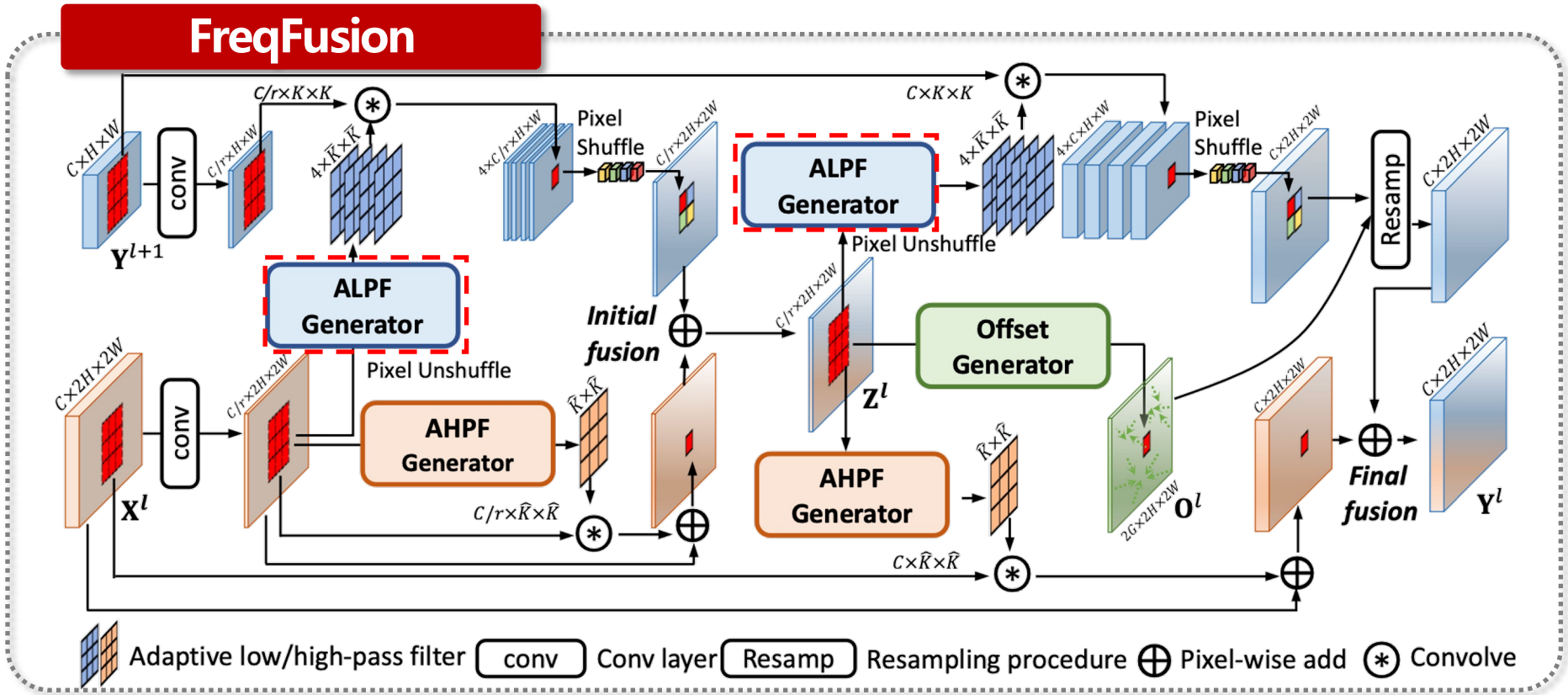
$$\bar{\mathbf{W}}^{l,p,q}_{i,j} = \mathrm{Softmax}(\bar{\mathbf{V}}^l_{i,j}) = \frac{\exp(\bar{\mathbf{V}}^{l,p,q}_{i,j})}{\sum_{p,q\in\Omega}\exp(\bar{\mathbf{V}}^{l,p,q}_{i,j})},$$

TPAMI 2024 FreqFusion

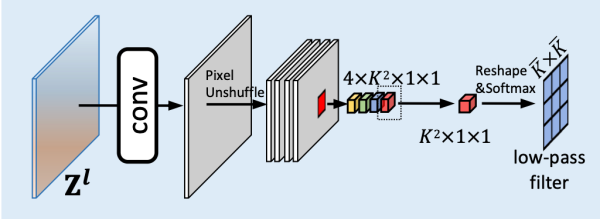
北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY

Adaptive Low-Pass Filter (ALPF) Generator

- ALPF Generator produces spatially-variant low-pass filters to attenuates the unnecessary high-frequency components, leading to much more uniform and consistent features within each object



Generation of ALPF:



$$\bar{\mathbf{V}}^l = \mathrm{Conv}_{3\times3}(\mathbf{Z}^l),$$

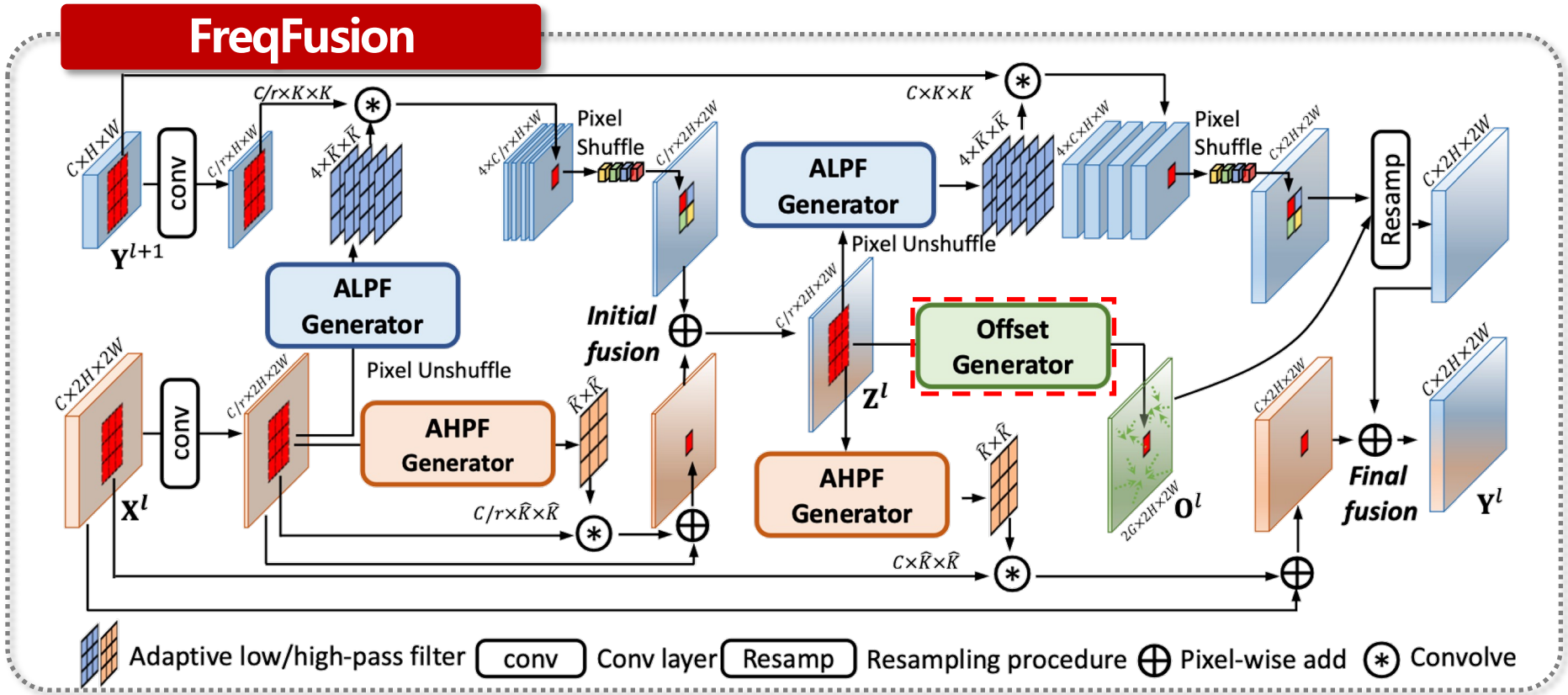
$$\bar{\mathbf{W}}^{l,p,q}_{i,j} = \mathrm{Softmax}(\bar{\mathbf{V}}^l_{i,j}) = \frac{\exp(\bar{\mathbf{V}}^{l,p,q}_{i,j})}{\sum_{p,q\in\Omega}\exp(\bar{\mathbf{V}}^{l,p,q}_{i,j})},$$

TPAMI 2024 FreqFusion

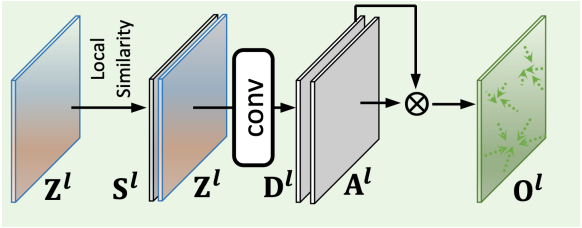
北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY

Offset Generator

- ALPF cannot fix both thin (prefer small filters) and large (prefer large filters) inconsistent regions. Regions with inconsistent features are often surrounded by regions with highly consistent features



Generation of offset:



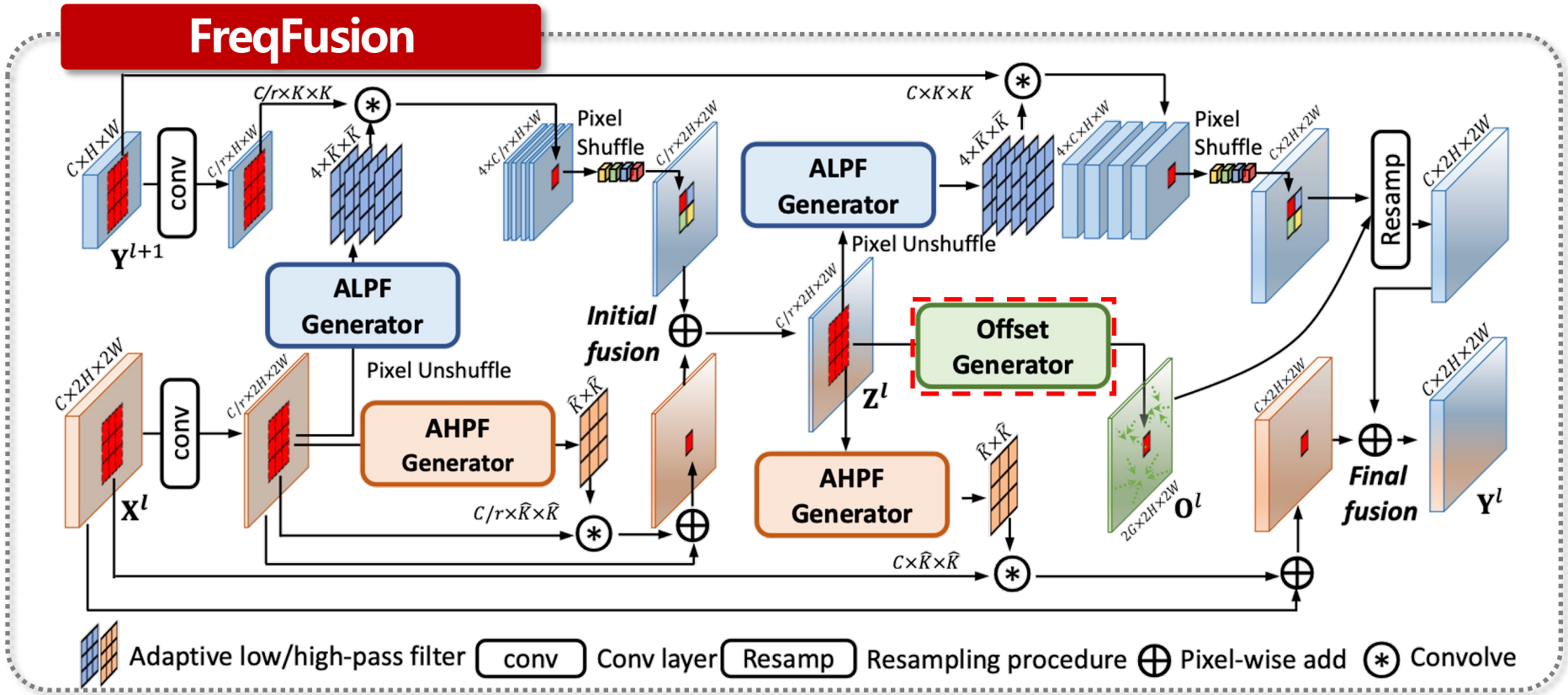
$$\mathbf{O}^l = \mathbf{D}^l \cdot \mathbf{A}^l,$$
$$\mathbf{D}^l = \mathrm{Conv3} \times 3(\mathrm{Concat}(\mathbf{Z}^l, \mathbf{S}^l)),$$
$$\mathbf{A}^l = \mathrm{Sigmoid}(\mathrm{Conv3} \times 3(\mathrm{Concat}(\mathbf{Z}^l, \mathbf{S}^l))),$$

$$\mathbf{S}^{l,p,q}_{i,j} = \frac{\sum_{c=1}^{C} \mathbf{Z}^l_{c,i,j} \cdot \mathbf{Z}^l_{c,i+p,j+q}}{\sqrt{\sum_{c=1}^{C} (\mathbf{Z}^l_{c,i,j})^2} \sqrt{\sum_{c=1}^{C} (\mathbf{Z}^l_{c,i+p,j+q})^2}},$$

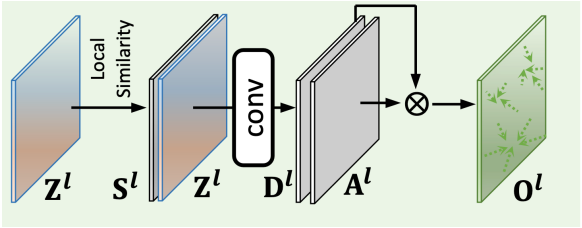
TPAMI 2024 FreqFusion

Proposed Method: FreqFusion

- The Offset Generator learns to resample features. It replaces inconsistent features with more consistent ones from the local neighborhood, which improves intra-class consistency for both large areas and thin boundaries.



Generation of offset:



$$\mathbf{O}^l = \mathbf{D}^l \cdot \mathbf{A}^l,$$
$$\mathbf{D}^l = \mathrm{Conv3} \times 3(\mathrm{Concat}(\mathbf{Z}^l, \mathbf{S}^l)),$$
$$\mathbf{A}^l = \mathrm{Sigmoid}(\mathrm{Conv3} \times 3(\mathrm{Concat}(\mathbf{Z}^l, \mathbf{S}^l))),$$

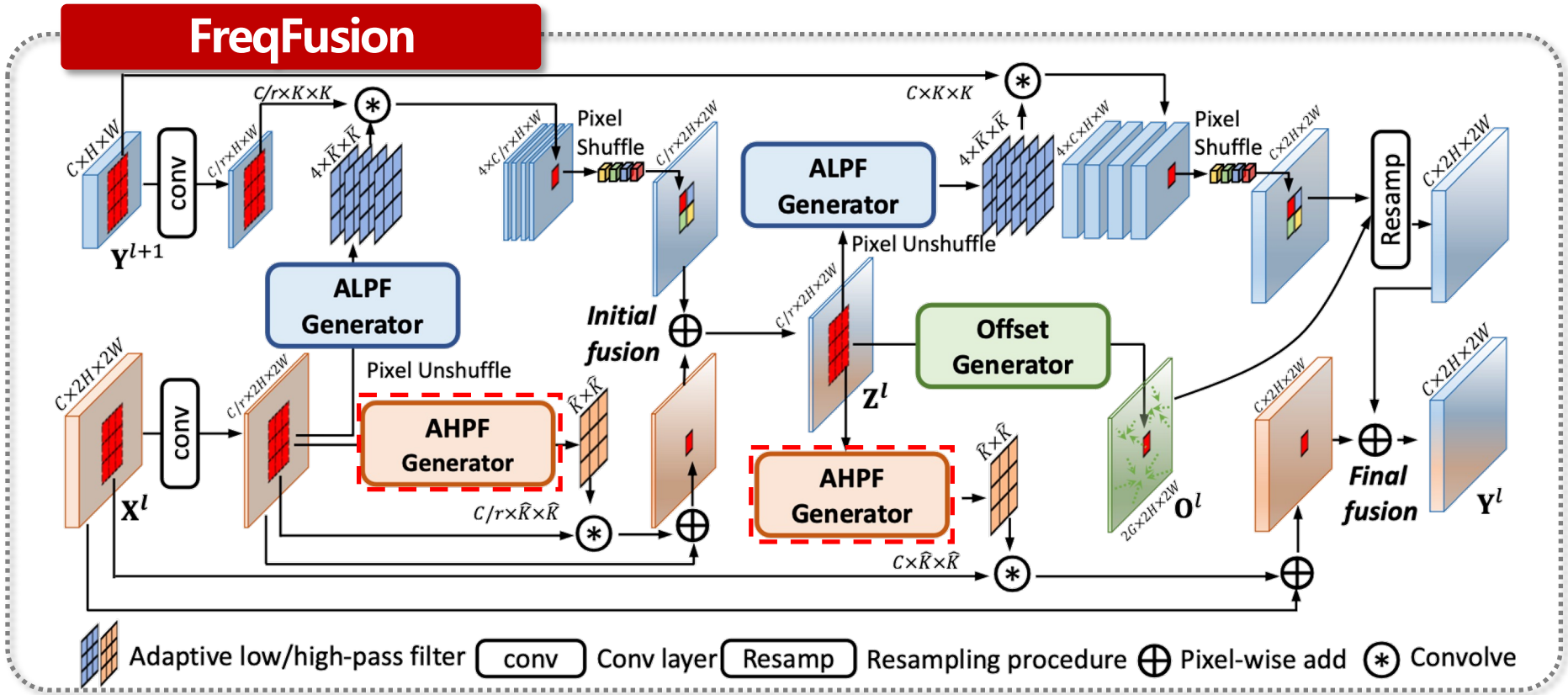
$$\mathbf{S}^{l,p,q}_{i,j} = \frac{\sum_{c=1}^{C} \mathbf{Z}^l_{c,i,j} \cdot \mathbf{Z}^l_{c,i+p,j+q}}{\sqrt{\sum_{c=1}^{C} (\mathbf{Z}^l_{c,i,j})^2} \sqrt{\sum_{c=1}^{C} (\mathbf{Z}^l_{c,i+p,j+q})^2}},$$

TPAMI 2024 FreqFusion

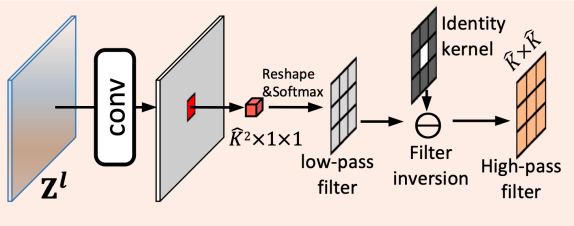
北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY

Adaptive High-Pass Filter (AHPF) Generator

- Boundary blur is caused by the loss of high-frequency information during the downsampling stages in the network backbone



Generation of AHPF:



$$\hat{\mathbf{V}}^l = \mathrm{Conv}_{3\times3}(\mathbf{Z}^l),$$

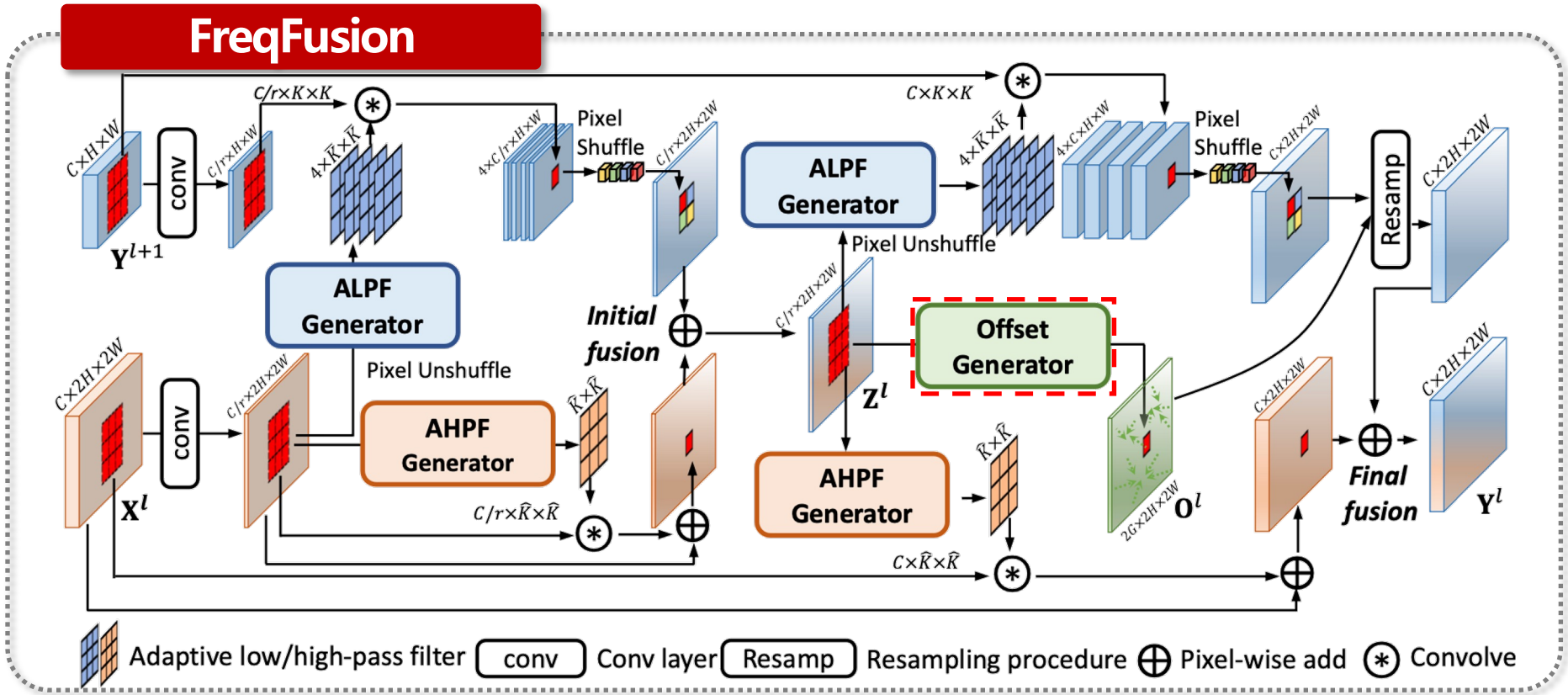
$$\hat{\mathbf{W}}^{l,p,q}_{i,j} = \mathbf{E} - \mathrm{Softmax}(\hat{\mathbf{V}}^l_{i,j}) = \mathbf{E}^{p,q} - \frac{\exp(\hat{\mathbf{V}}^{l,p,q}_{i,j})}{\sum_{p,q\in\Omega}\exp(\hat{\mathbf{V}}^{l,p,q}_{i,j})},$$

$$\tilde{\mathbf{X}}^l_{i,j} = \mathbf{X}^l_{i,j} + \sum_{p,q\in\Omega} \hat{\mathbf{W}}^{l,p,q}_{i,j} \cdot \mathbf{X}^l_{i,j}.$$

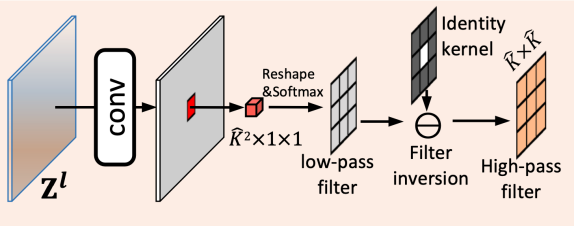
TPAMI 2024 FreqFusion

Proposed Method: FreqFusion

- The Adaptive High-Pass Filter (AHPF) Generator enhances the final fusion. It strengthens the detailed high-frequency boundary information, resulting in sharper and more accurate object boundaries.



Generation of AHPF:



$$\hat{\mathbf{V}}^l = \mathrm{Conv}_{3\times3}(\mathbf{Z}^l),$$

$$\hat{\mathbf{W}}^{l,p,q}_{i,j} = \mathbf{E} - \mathrm{Softmax}(\hat{\mathbf{V}}^l_{i,j}) = \mathbf{E}^{p,q} - \frac{\exp(\hat{\mathbf{V}}^{l,p,q}_{i,j})}{\sum_{p,q\in\Omega}\exp(\hat{\mathbf{V}}^{l,p,q}_{i,j})},$$

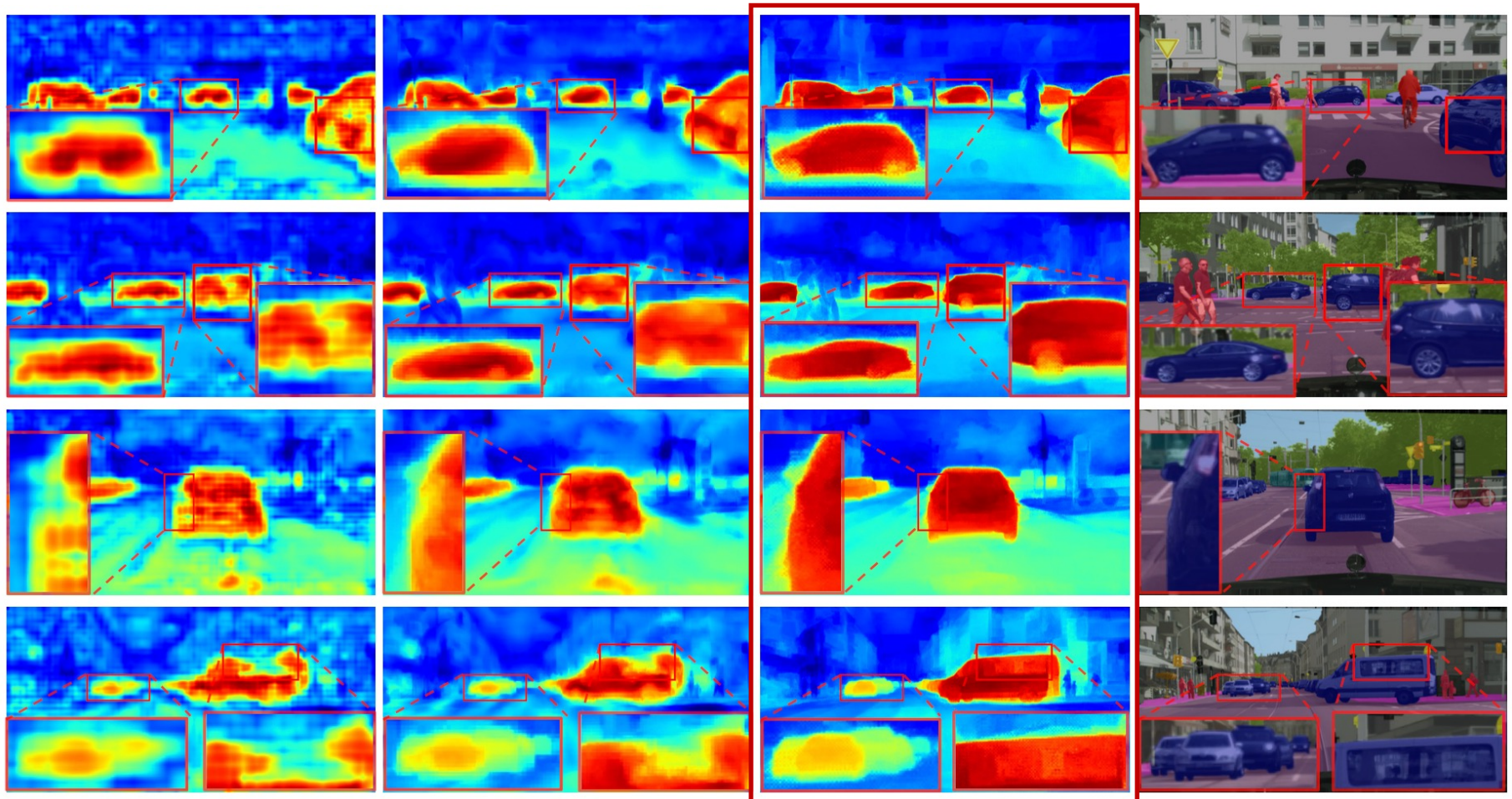
$$\tilde{\mathbf{X}}^l_{i,j} = \mathbf{X}^l_{i,j} + \sum_{p,q\in\Omega} \hat{\mathbf{W}}^{l,p,q}_{i,j} \cdot \mathbf{X}^l_{i,j}.$$

TPAMI 2024 FreqFusion

北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY

Feature Visualization

ALPF (b) and Offset generators (c) significantly improves feature consistency compared to standard fusion (a)



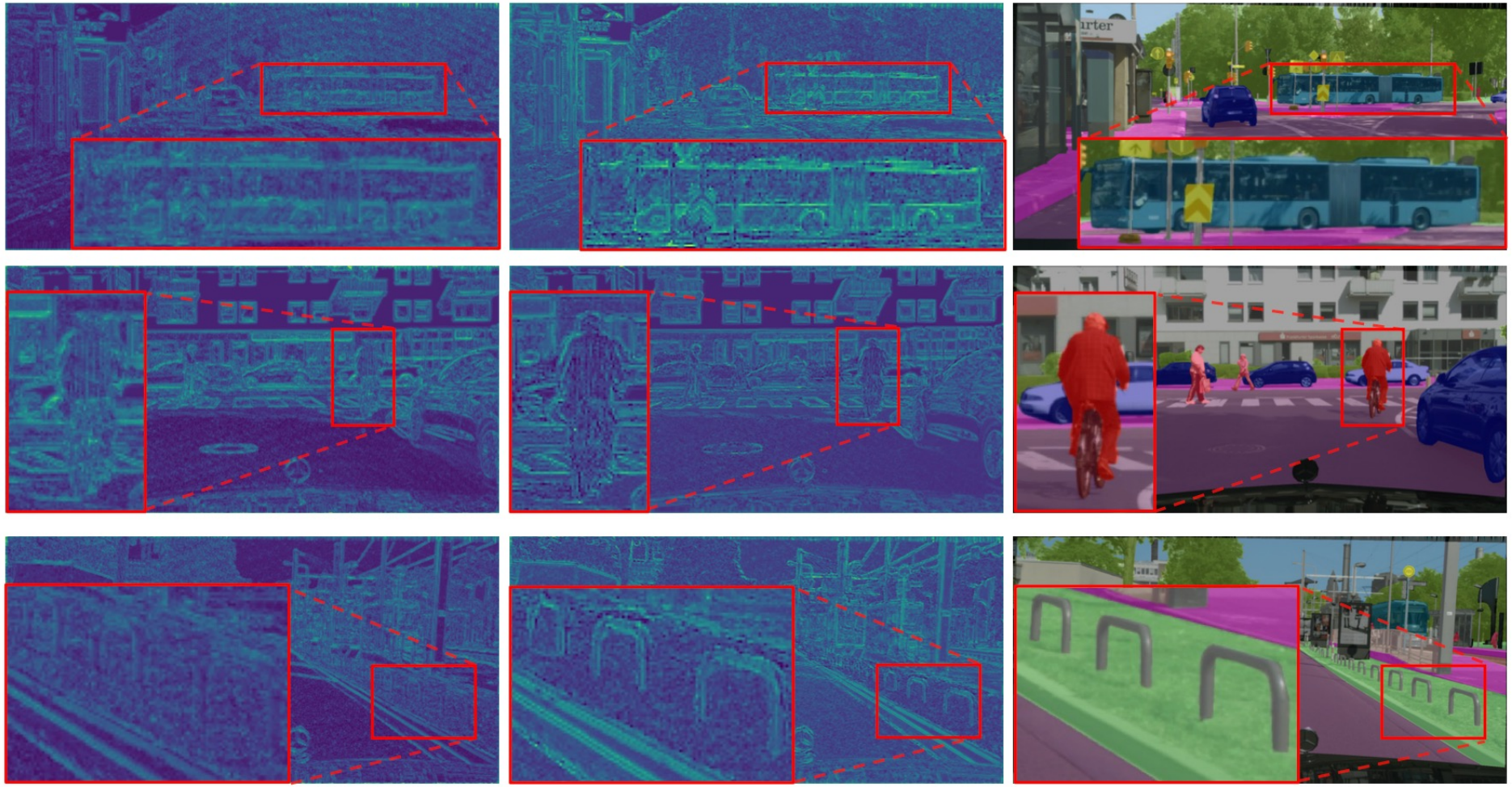
(a) Standard feature fusion (b) +ALPF generator (c) +ALPF generator +Offset generator (d) Ground truth

TPAMI 2024 FreqFusion

北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY

Feature Visualization

The AHPF component strengthens fine details in low-level features, making object contours clearer and more complete.



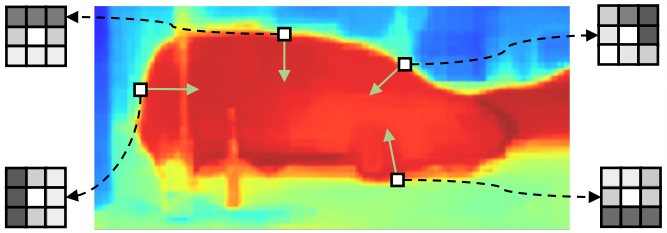
(a) Low-level feature (b) +AHPF generator (c) Ground truth

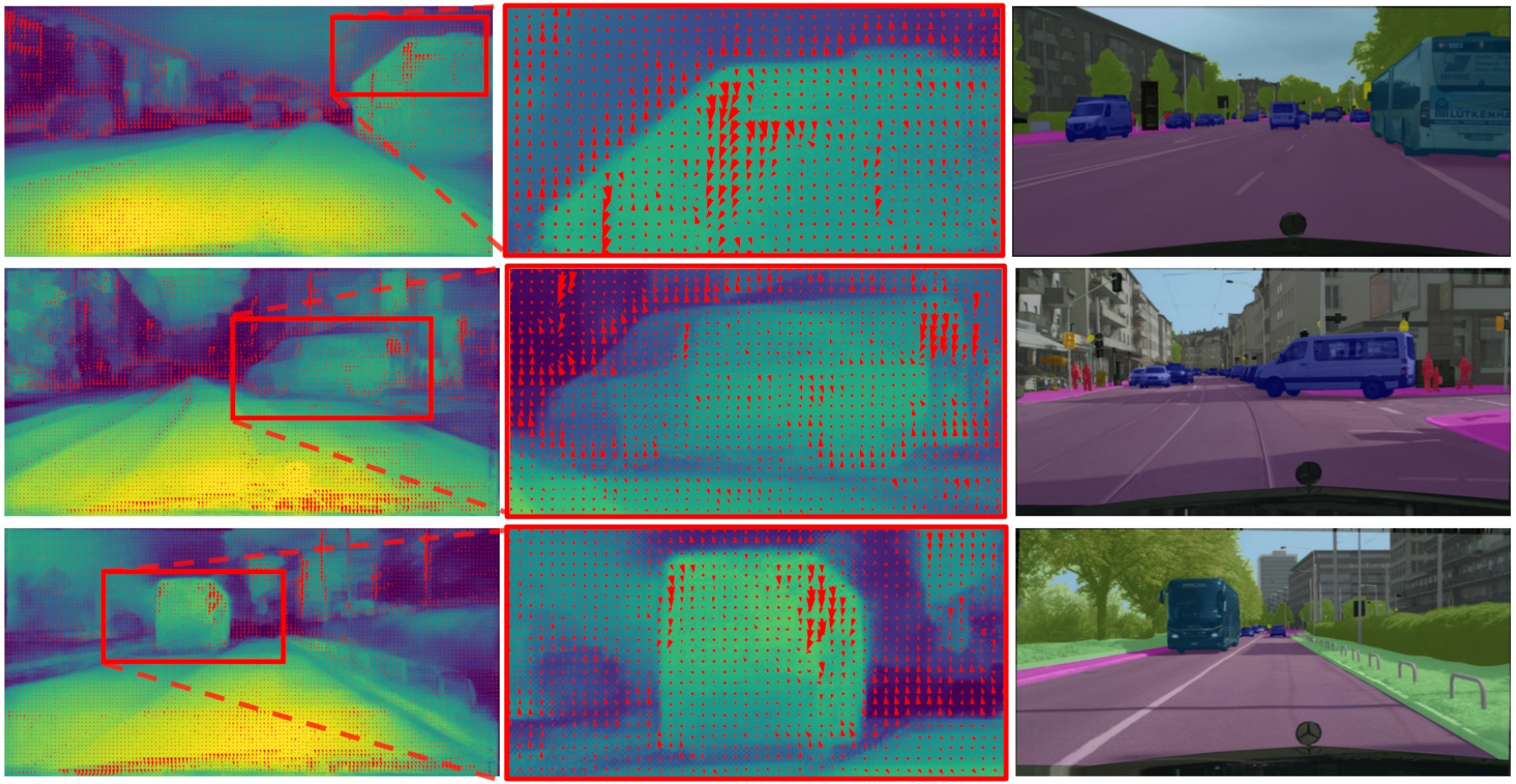
TPAMI 2024 FreqFusion

北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY

Feature Visualization

The Offset Generator learns to sample from more consistent feature regions to correct inconsistent areas nearby.





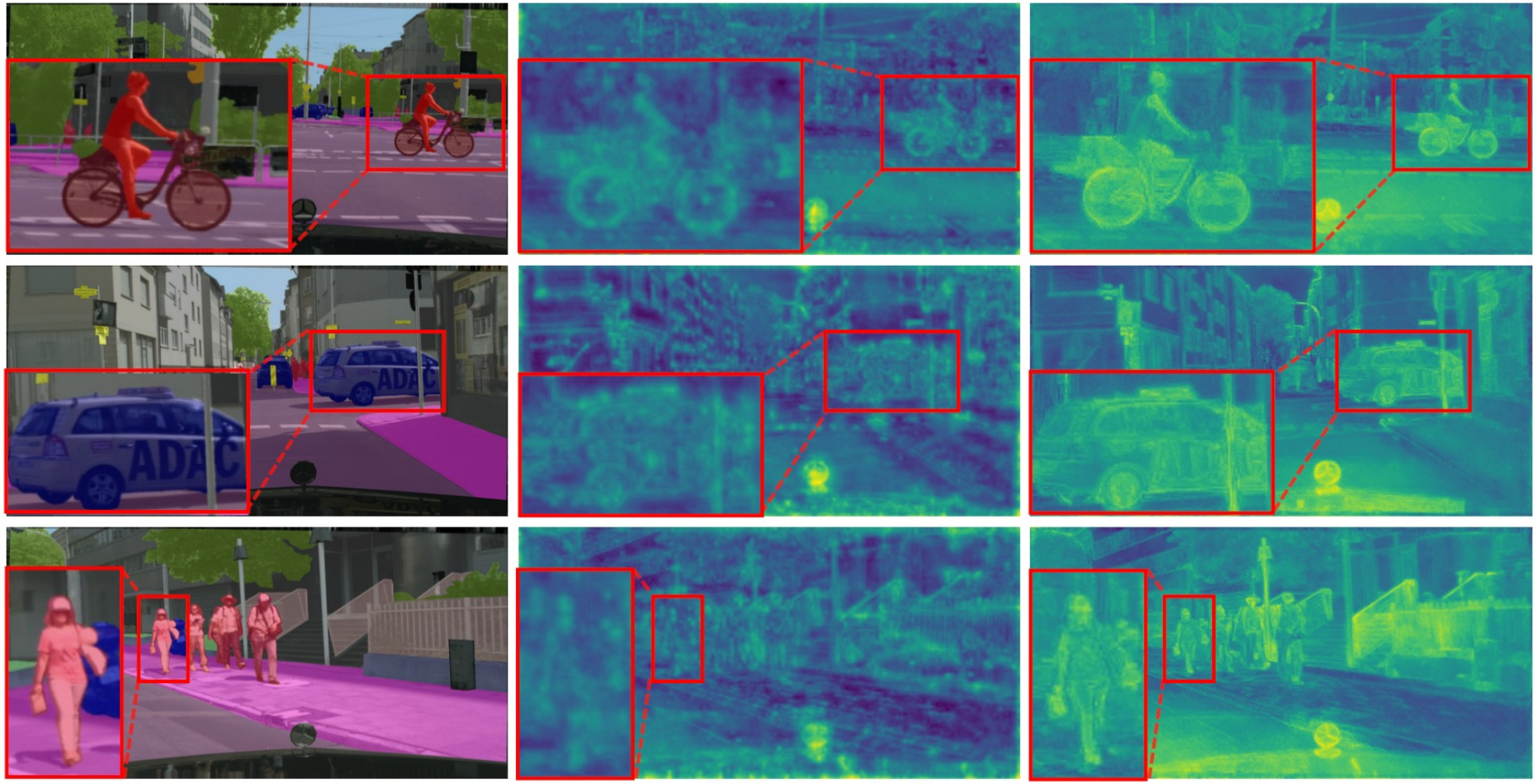
Resampling offsets | Zoom in | Ground truth

TPAMI 2024 FreqFusion

北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY

Feature Visualization

The features generated by FreqFusion are visibly superior to those from standard feature fusion, with less internal noise and sharper boundaries.



Ground truth | Standard feature fusion | FreqFusion (Ours)

TPAMI 2024 FreqFusion

北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY

Quantitative Feature Analysis

- The proposed FreqFusion significantly improves the overall feature quality, including IntraSim, SimMargin, SimAcc

Method	IntraSim↑	SimMargin↑	SimAcc↑	Boundary IntraSim↑	Boundary SimMargin↑	Boundary SimAcc↑
Standard feature fusion	0.697	0.255	0.907	0.584	0.202	0.632
FreqFusion	**0.799**	**0.297**	**0.941**	**0.694**	**0.239**	**0.728**
w/o ALPF Generator (Initial fusion)	0.792	0.288	0.929	0.687	0.230	0.725
w/o AHPF Generator (Initial fusion)	0.798	0.296	0.940	0.690	0.232	0.726
w/o ALPF Generator (Final fusion)	0.727	0.245	0.918	0.610	0.229	0.724
w/o Offset Generator (Final fusion)	0.760	0.295	0.925	0.648	0.235	0.720
w/o AHPF Generator (Final fusion)	0.796	0.295	0.938	0.688	0.228	0.718

TPAMI 2024 FreqFusion

北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY

Quantitative Results on Downstream Tasks

- The improvements in feature quality translate directly to better performance on various dense prediction tasks, including Semantic/ Instance/Panoptic Segmentation, Object Detection

Mask2Former[CVPR2022] [32]	Backbone	mIoU
Bilinear	Swin-B†	53.9
FreqFusion (Ours)	Swin-B†	**55.3 (+1.4)**
Bilinear	Swin-L†	56.1
FreqFusion (Ours)	Swin-L†	**56.8 (+0.7)**

Method	Params (M)	FLOPs (G)	mIoU	bIoU
Segformer-B1 [NeurIPS2021] [29]	13.74	15.91	41.7	27.8
+ Deconv[CVPR2010] [79]	+3.5	+34.4	40.7	25.9
+ PixelShuffle[CVPR2016] [41]	+14.2	+34.4	41.5	26.6
+ CARAFE[ICCV2019] [43]	+0.44	+1.45	42.8	29.8
+ IndexNet[TPAMI2020] [44]	+12.6	+30.65	41.5	28.3
+ A2U[CVPR2021] [45]	+0.12	+0.41	41.5	27.3
+ FADE[ECCV2022] [22]	+0.29	+2.65	43.1	31.7
+ SAPA-B[NeurIPS2022] [21]	+0.1	+1.0	43.2	31.0
+ Dysample-S+[ICCV2023] [50]	+0.01	+0.3	43.6	29.9
+ Dysample+[ICCV2023] [50]	+0.1	+0.4	43.3	29.2
+ FreqFusion (Ours)	+0.34	+2.35	**44.5**	**32.8**

Method	Backbone	Param.	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Nearest	R50	46.8M	37.5	58.2	40.8	21.3	41.1	48.9
Deconv	R50	+2.4M	37.3	57.8	40.3	21.3	41.1	48.0
PixelShuffle[CVPR2016] [41]	R50	+9.4M	37.5	58.5	40.4	21.5	41.5	48.3
CARAFE[ICCV2019] [43]	R50	+0.3M	38.6	59.9	42.2	23.3	42.2	49.7
IndexNet[TPAMI2020] [44]	R50	+8.4M	37.6	58.4	40.9	21.5	41.3	49.2
A2U[CVPR2021] [45]	R50	+38.9K	37.3	58.7	40.0	21.7	41.1	48.5
FADE[ECCV2022] [22]	R50	+0.2M	38.5	59.6	41.8	23.1	42.2	49.3
SAPA-B[NeurIPS2022] [21]	R50	+0.1M	37.8	59.2	40.6	22.4	41.4	49.1
DySample-S+[ICCV2023] [50]	R50	+8.2K	38.6	59.8	42.1	22.5	42.1	50.0
DySample+[ICCV2023] [50]	R50	+65.5K	38.7	60.0	42.2	22.5	42.4	50.2
FreqFusion (Ours)	R50	+0.3M	**39.4**	**60.9**	**42.7**	**23.0**	**43.3**	**50.9**
Nearest	R101	65.8M	39.4	60.1	43.1	22.4	43.7	51.1
DySample+[ICCV2023] [50]	R101	+65.5K	40.5	61.6	43.8	24.2	44.5	52.3
FreqFusion (Ours)	R101	+0.3M	**41.0**	**62.2**	**44.9**	**24.7**	**45.0**	**53.4**

Method	Backbone	Params	PQ	PQ^{th}	PQ^{st}	SQ	RQ
Nearest	R50	46.0M	40.2	47.8	28.9	77.8	49.3
Deconv	R50	+1.8M	39.6	47.0	28.4	77.1	48.5
PixelShuffle[CVPR2016] [41]	R50	+7.1M	40.0	47.4	28.8	77.1	49.1
CARAFE[ICCV2019] [43]	R50	+0.2M	40.8	47.7	30.4	78.2	50.0
IndexNet[TPAMI2020] [44]	R50	+6.3M	40.2	47.6	28.9	77.1	49.3
A2U[CVPR2021] [45]	R50	+29.2K	40.1	47.6	28.7	77.3	48.0
FADE[ECCV2022] [22]	R50	+0.1M	40.9	48.0	30.3	78.1	50.1
SAPA-B[NeurIPS2022] [21]	R50	+0.1M	40.6	47.7	29.8	78.0	49.6
DySample-S+[ICCV2023] [50]	R50	+6.2K	41.1	48.1	30.5	78.2	50.2
DySample+[ICCV2023] [50]	R50	+49.2K	41.5	48.5	30.8	78.3	50.7
FreqFusion	R50	+0.3M	**42.7**	**49.3**	**32.7**	**79.0**	**51.9**
Nearest	R101	65.0M	42.2	50.1	30.3	78.3	51.4
DySample+[ICCV2023] [50]	R101	+49.2K	43.0	50.2	32.1	78.6	52.4
FreqFusion	R101	+0.3M	**44.0**	**50.8**	**33.7**	**79.4**	**53.4**

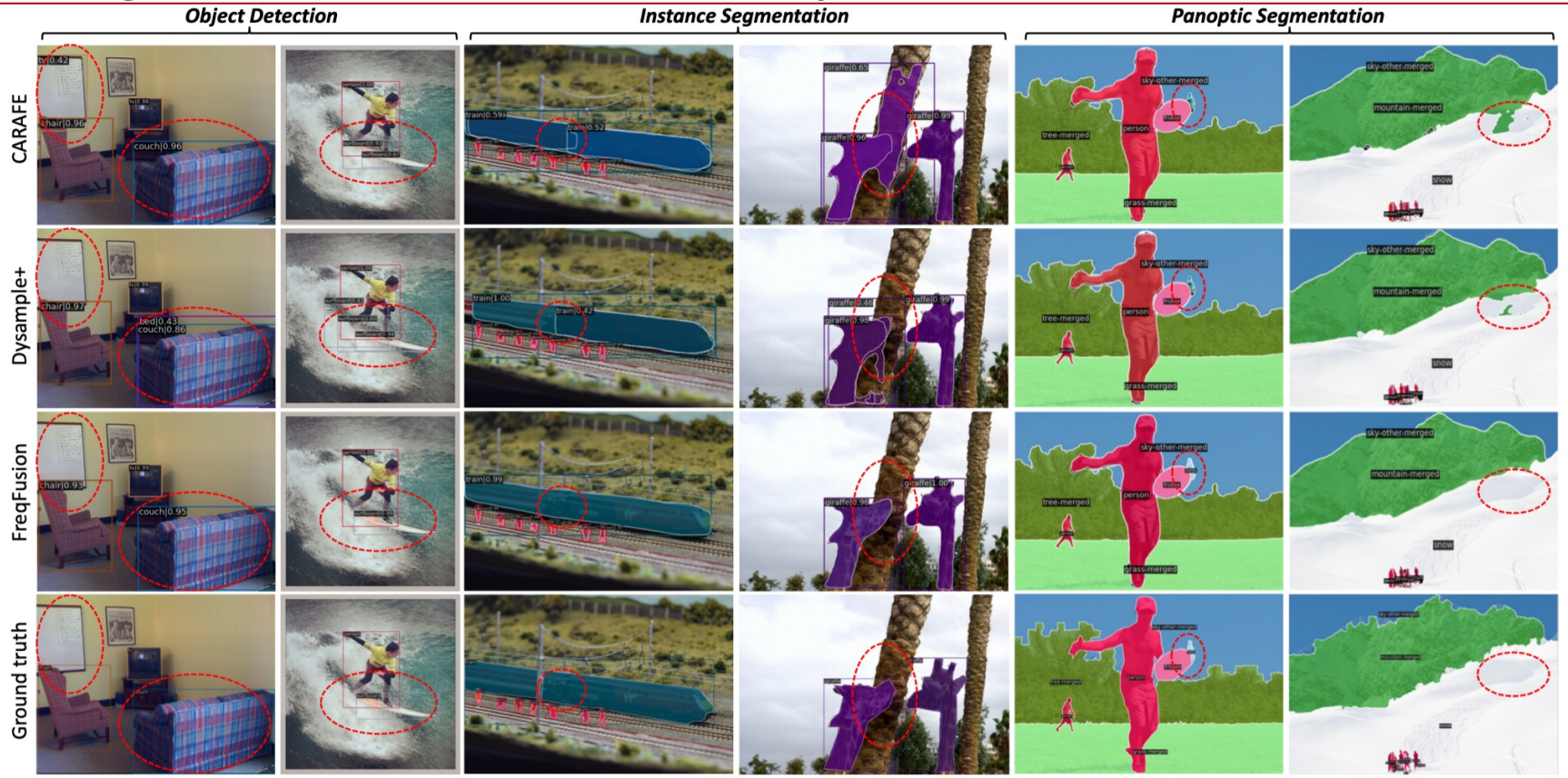
Method	Backbone	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Nearest	R50	38.3	58.7	42	21.9	41.8	50.2
Deconv	R50	37.9	58.5	41.0	22.0	41.6	49.0
PixelShuffle[CVPR2016] [41]	R50	38.5	59.4	41.9	22.0	42.3	49.8
CARAFE[ICCV2019] [43]	R50	39.2	60.0	43.0	23.0	42.8	50.8
IndexNet[TPAMI2020] [44]	R50	38.4	59.2	41.7	22.1	41.7	50.3
A2U[CVPR2021] [45]	R50	38.2	59.2	41.4	22.3	41.7	49.6
FADE[ECCV2022] [22]	R50	39.1	60.3	42.4	23.6	42.3	51.0
SAPA-B[NeurIPS2022] [21]	R50	38.7	59.7	42.2	23.1	41.8	49.9
DySample-S+[ICCV2023] [50]	R50	39.3	60.3	42.8	23.2	42.7	50.8
DySample+[ICCV2023] [50]	R50	39.6	60.4	43.5	23.4	42.9	**51.7**
FreqFusion	R50	**40.0**	**61.2**	**43.5**	**24.3**	**43.9**	51.5
Nearest	R101	40.0	60.4	43.7	22.8	43.7	52.0
DySample+[ICCV2023] [50]	R101	41.0	61.9	44.9	24.3	45.0	53.5
FreqFusion	R101	**41.6**	**62.4**	**45.7**	**25.0**	**45.6**	**54.2**
Instance Segmentation Mask Results							
Nearest	R50	34.7	55.8	37.2	16.1	37.3	50.8
Deconv	R50	34.5	55.5	36.8	16.4	37.0	49.5
PixelShuffle[CVPR2016] [41]	R50	34.8	56.0	37.3	16.3	37.5	50.4
CARAFE[ICCV2019] [43]	R50	35.4	56.7	37.6	16.9	38.1	51.3
IndexNet[TPAMI2020] [44]	R50	34.7	55.9	37.1	16.0	37.0	51.1
A2U[CVPR2021] [45]	R50	34.6	56.0	36.8	16.1	37.4	50.3
FADE[ECCV2022] [22]	R50	35.1	56.7	37.2	16.7	37.5	51.4
SAPA-B[NeurIPS2022] [21]	R50	35.1	56.5	37.4	16.7	37.6	50.6
DySample-S+[ICCV2023] [50]	R50	35.5	56.8	37.8	17.0	37.9	51.9
DySample+[ICCV2023] [50]	R50	35.7	57.3	38.2	17.3	38.2	51.8
FreqFusion	R50	**36.0**	**57.9**	**38.1**	**17.9**	**39.0**	**52.3**
Nearest	R101	36.0	57.6	38.5	16.5	39.3	52.2
DySample+[ICCV2023] [50]	R101	36.8	58.7	39.5	17.5	40.0	53.8
FreqFusion	R101	**37.4**	**59.6**	**39.9**	**18.4**	**40.6**	**54.3**

TPAMI 2024 FreqFusion

北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY

Qualitative Results on Downstream Tasks

- The visual results show that models enhanced with FreqFusion produce more detailed and accurate predictions. This is especially evident at object edges and for small, thin objects.

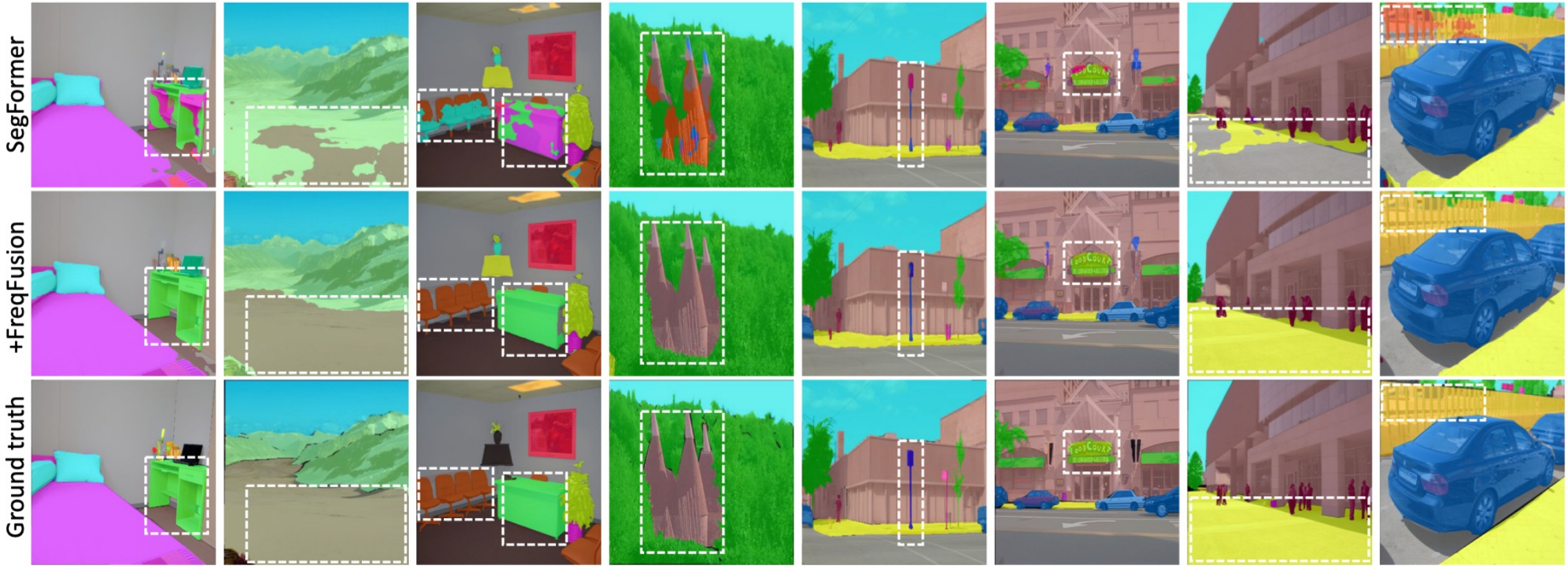


TPAMI 2024 FreqFusion

北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY

Qualitative Results on Downstream Tasks

- The predictions are cleaner, with fewer misclassified pixels and more precise segmentation masks



Conclusion & Future Work

- Conclusion
  - This paper addresses the key challenges of intra-class inconsistency and boundary displacement in dense image prediction tasks.
  - We proposed FreqFusion, a novel module that uses frequency-aware principles and feature similarity analysis to resolve these issues, leading to significant performance gains.
- Future Work
  - We plan to explore improving computational efficiency and real-time applicability for deployment in resource-constrained environments.
  - We will also investigate extending FreqFusion to handle different data modalities, such as video, to address feature fusion for dynamic scenes.

北京理工大學
BEIJING INSTITUTE OF TECHNOLOGY

Frequency-aware Feature Fusion for Dense Image Prediction

Code Paper Github Google Scholar

Linwei Chen 陈林蔚 Ph.D

charleschen2013218@gmail.com

Supervisor: Prof. Ying Fu（付莹）

Team member

北京理工大學 BEIJING INSTITUTE OF TECHNOLOGY

ISP Lab：https://ying-fu.github.io/

Prof. Ying Fu

Beijing Insitute of Technology

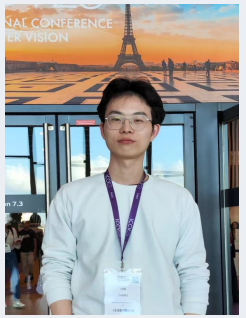
- National High-Level Talent Program in 2017.
- Best Paper Awards: ICML 2020, PRCV 2019.

Yunhao Zou-Ph.D

- Denoising, HDR,

Linwei Chen-Ph.D

- Segmentation, Detection



Fan Zhang-Ph.D

- Image enhancement

Yingping Liang

- 3D Vision

Yingping Liang

- Super-resolution

Lab Website