



浙江大學
ZHEJIANG UNIVERSITY

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

The CVPR logo for Vancouver, Canada, featuring a stylized blue skyline of the city's buildings.

Recurrent Homography estimation using Homography-guided image **W**arping and **F**ocus transformer (RHWF)

Si-Yuan Cao, Runmin Zhang, Lun Luo, Beinan Yu, Zehua Sheng,
Junwei Li, Hui-Liang Shen

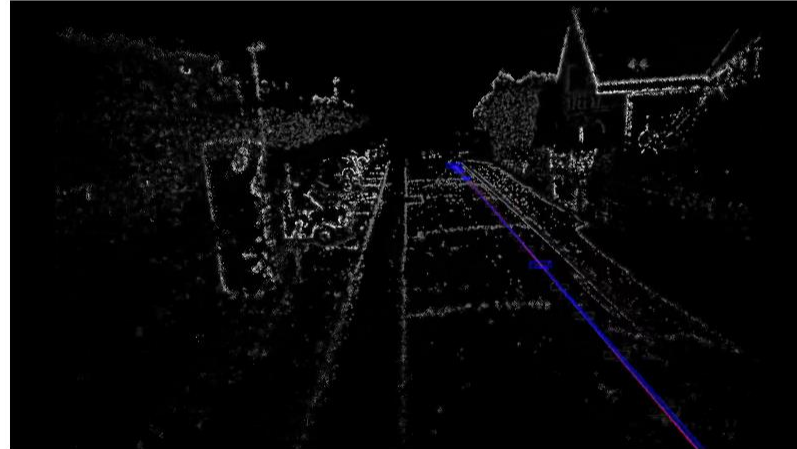
Poster: WED-AM-153

Background

Video Stabilization



SLAM



Planar Object Tracking



Motivation

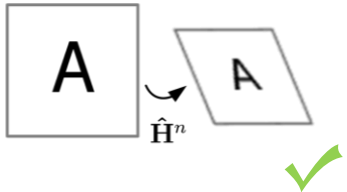
Warping strategies



- No warping
- Only translation-equivariance



- Pre-warping
- Redundant in computation



- Homography guided warping
- Improve feature consistency during recurrence

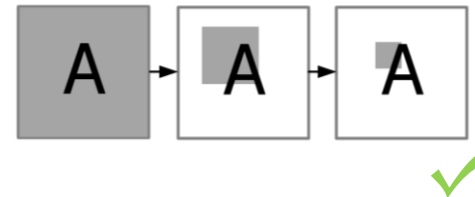
Attention strategies



- Pure global attention
- Time-consuming

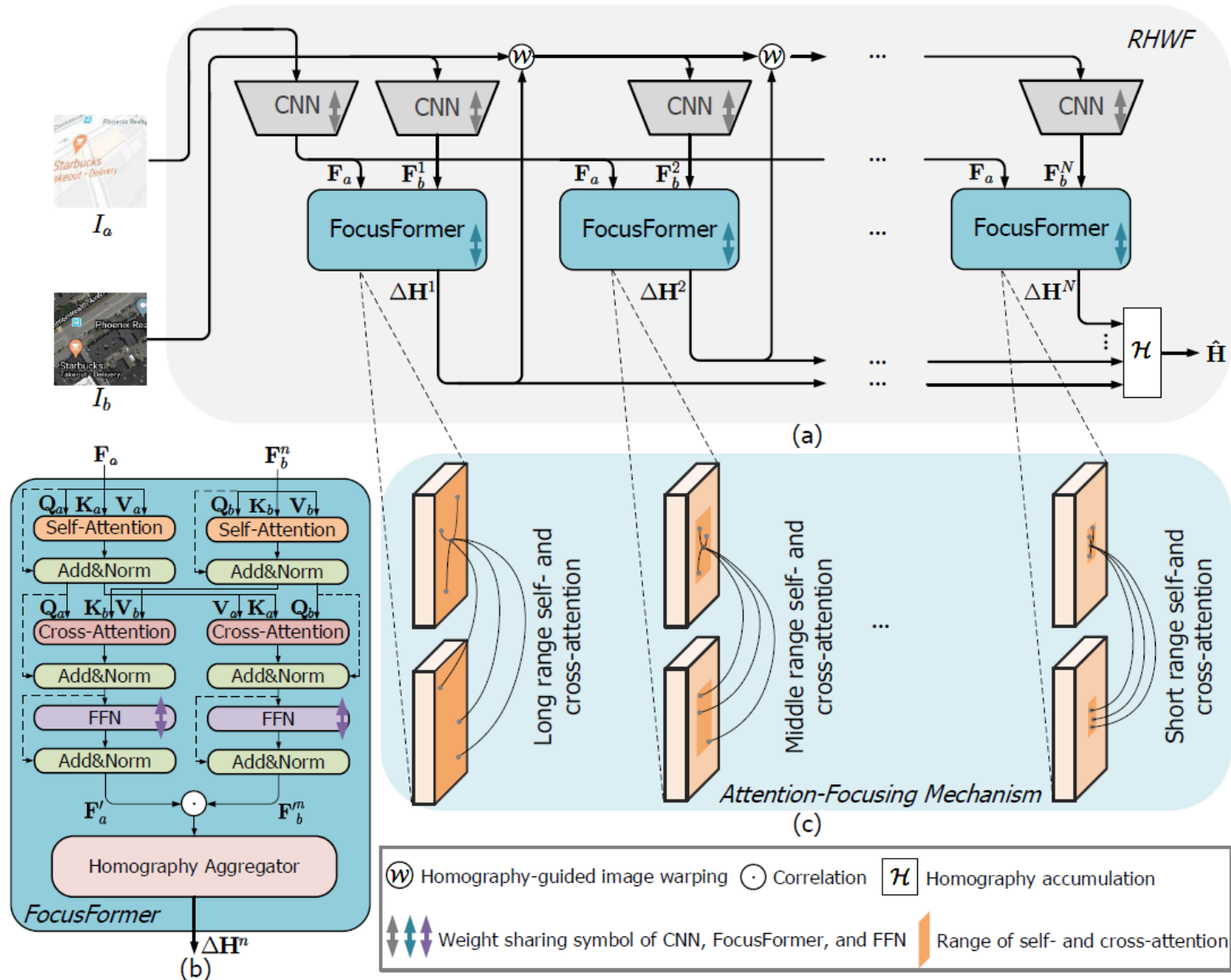


- Pure local attention
- Small receptive field

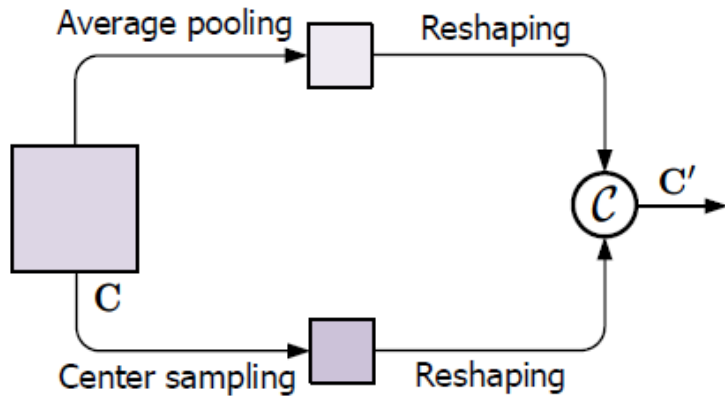


- Focus attention
- Gradually clarify the attention targets

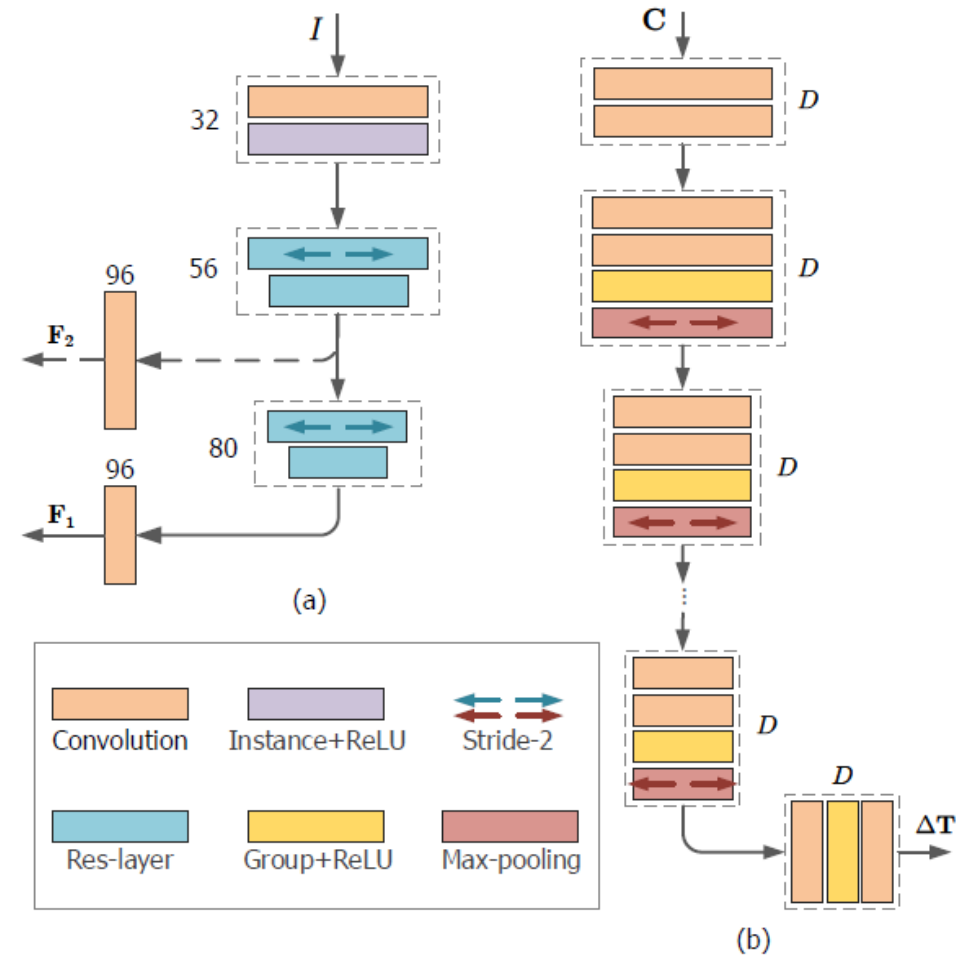
RHWF: Overview



Network details



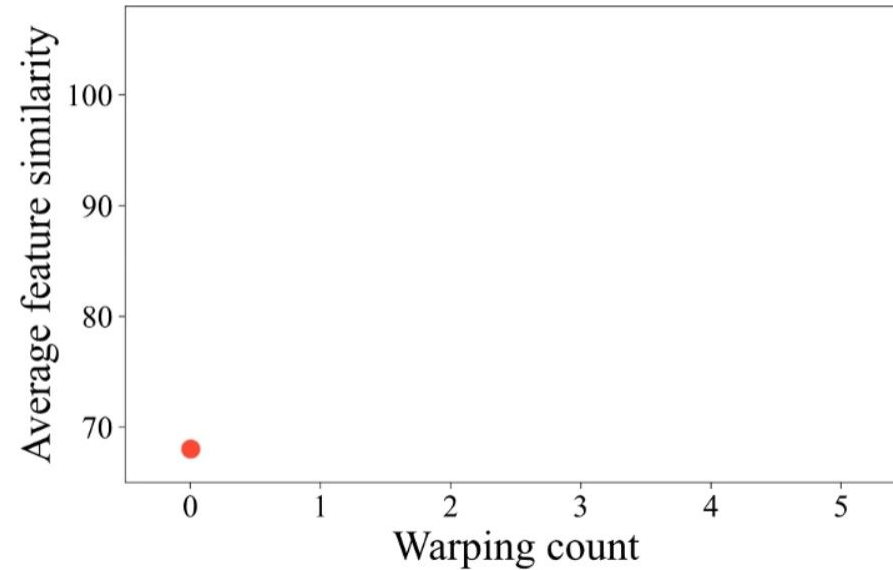
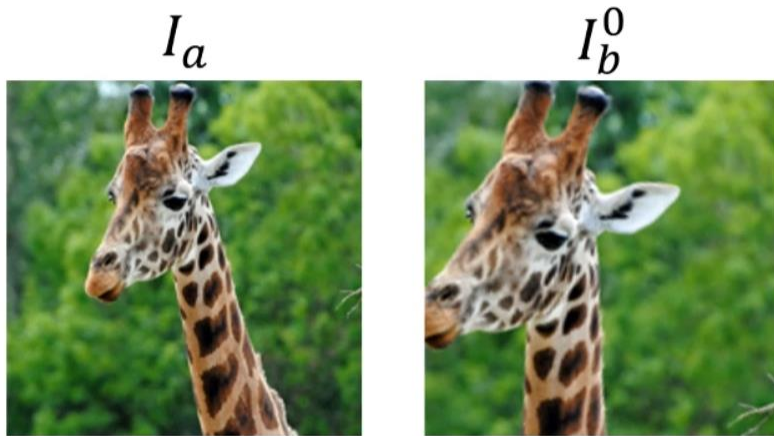
Correlation pooling operation



(a) CNN backbone

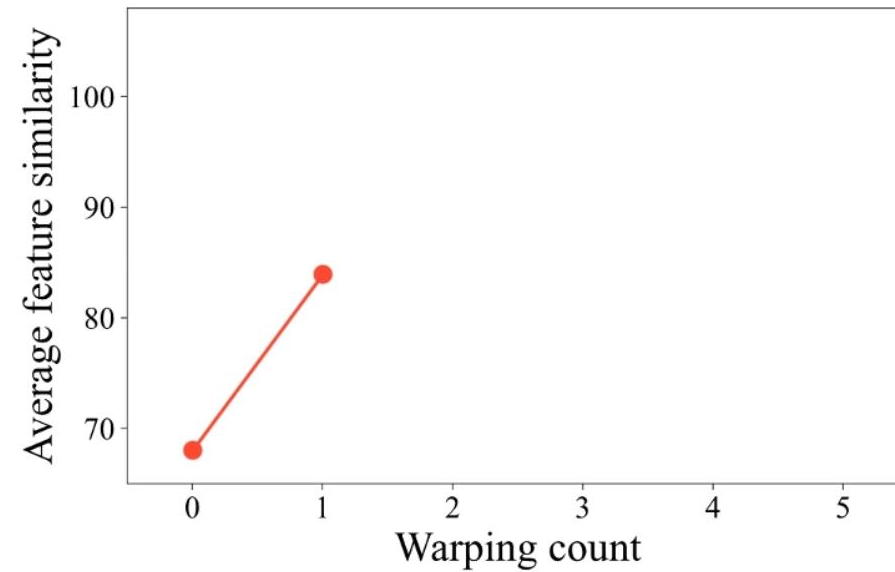
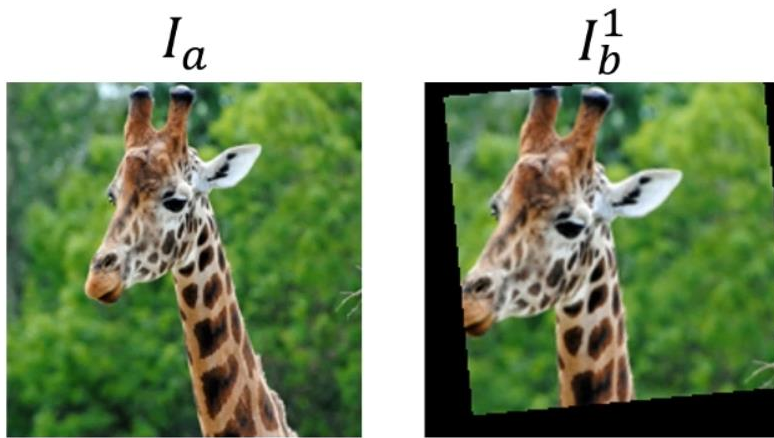
(b) Homography aggregator

Homography-guided Image Warping



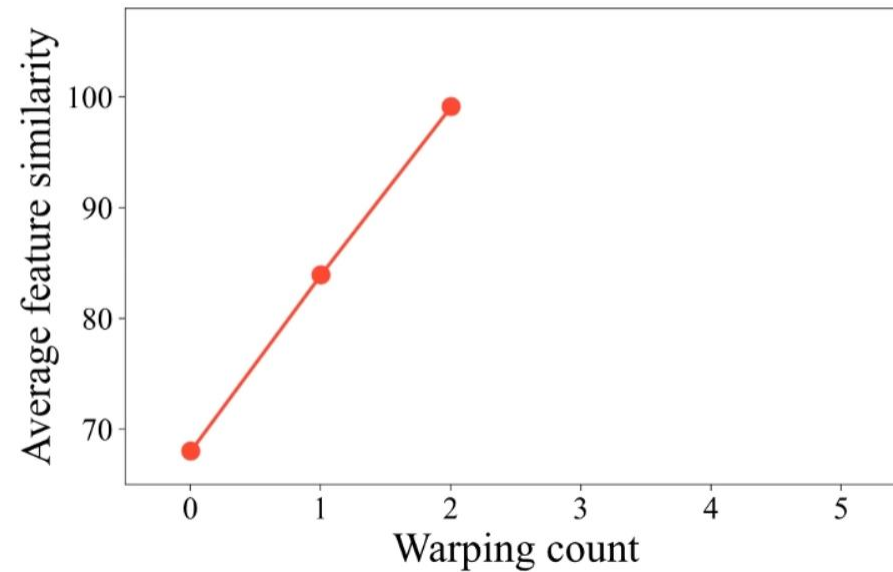
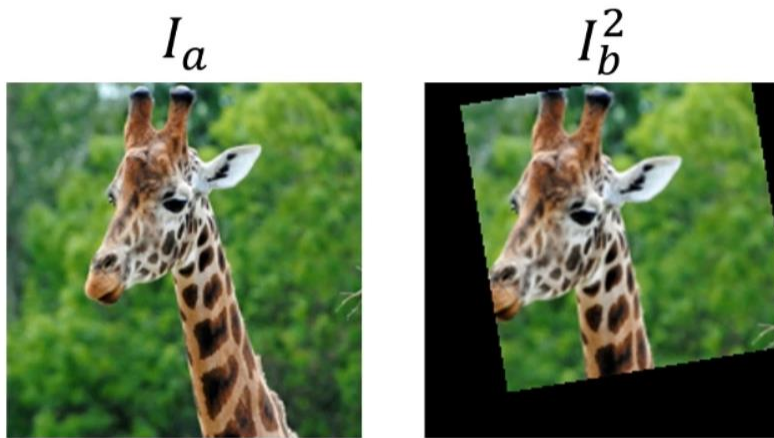
Gradually improve the feature consistency

Homography-guided Image Warping



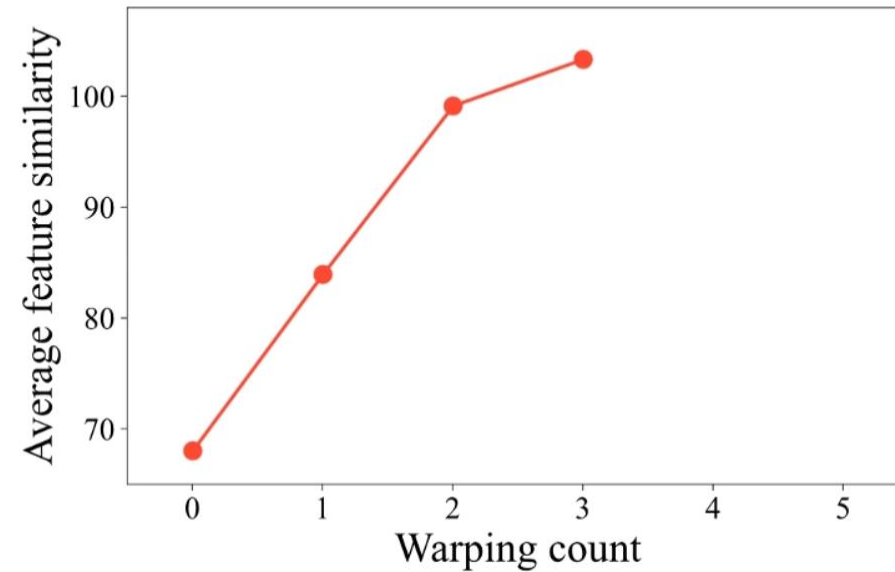
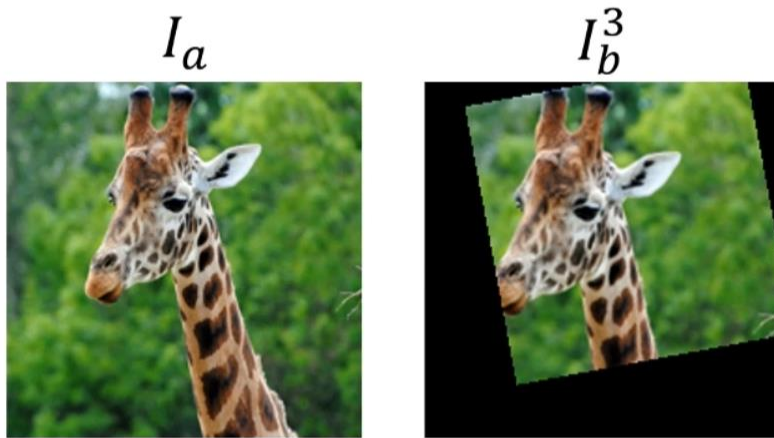
Gradually improve the feature consistency

Homography-guided Image Warping



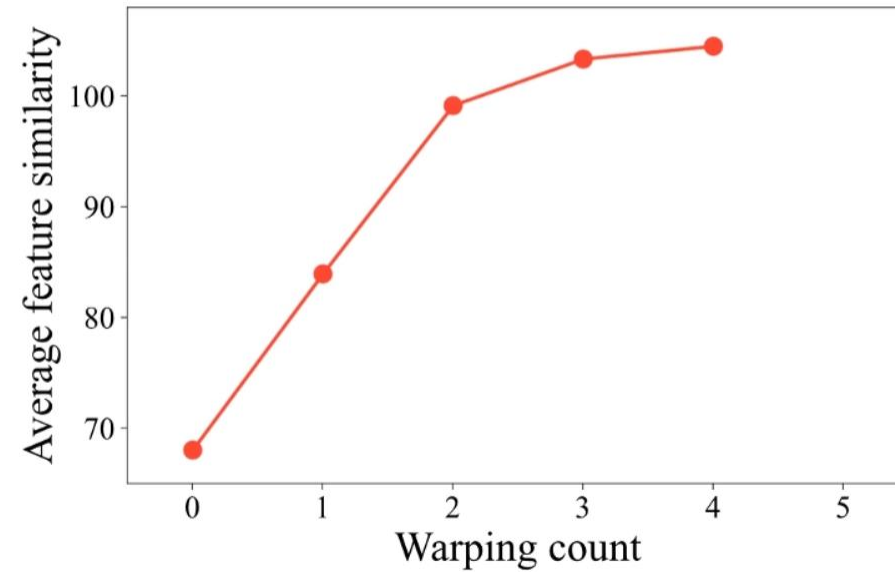
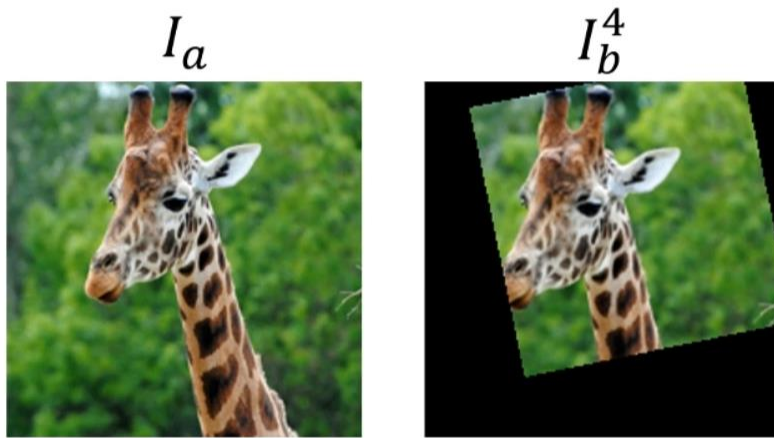
Gradually improve the feature consistency

Homography-guided Image Warping



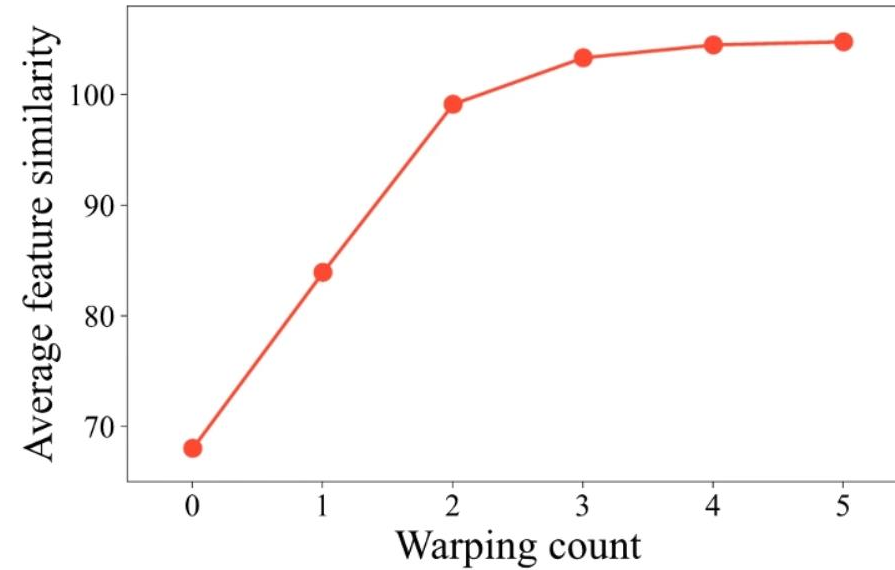
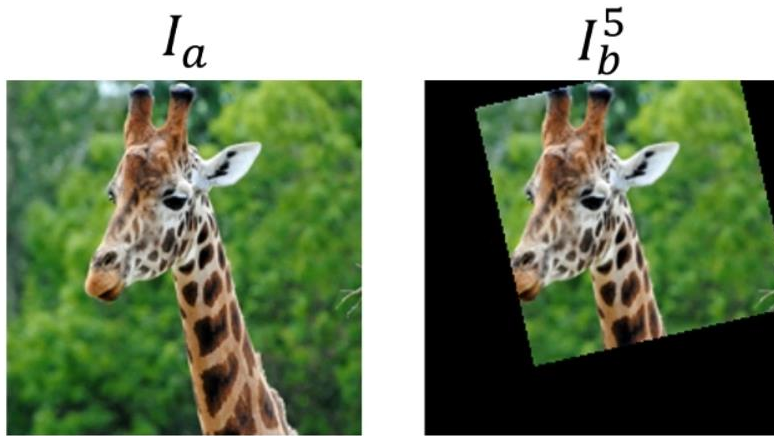
Gradually improve the feature consistency

Homography-guided Image Warping



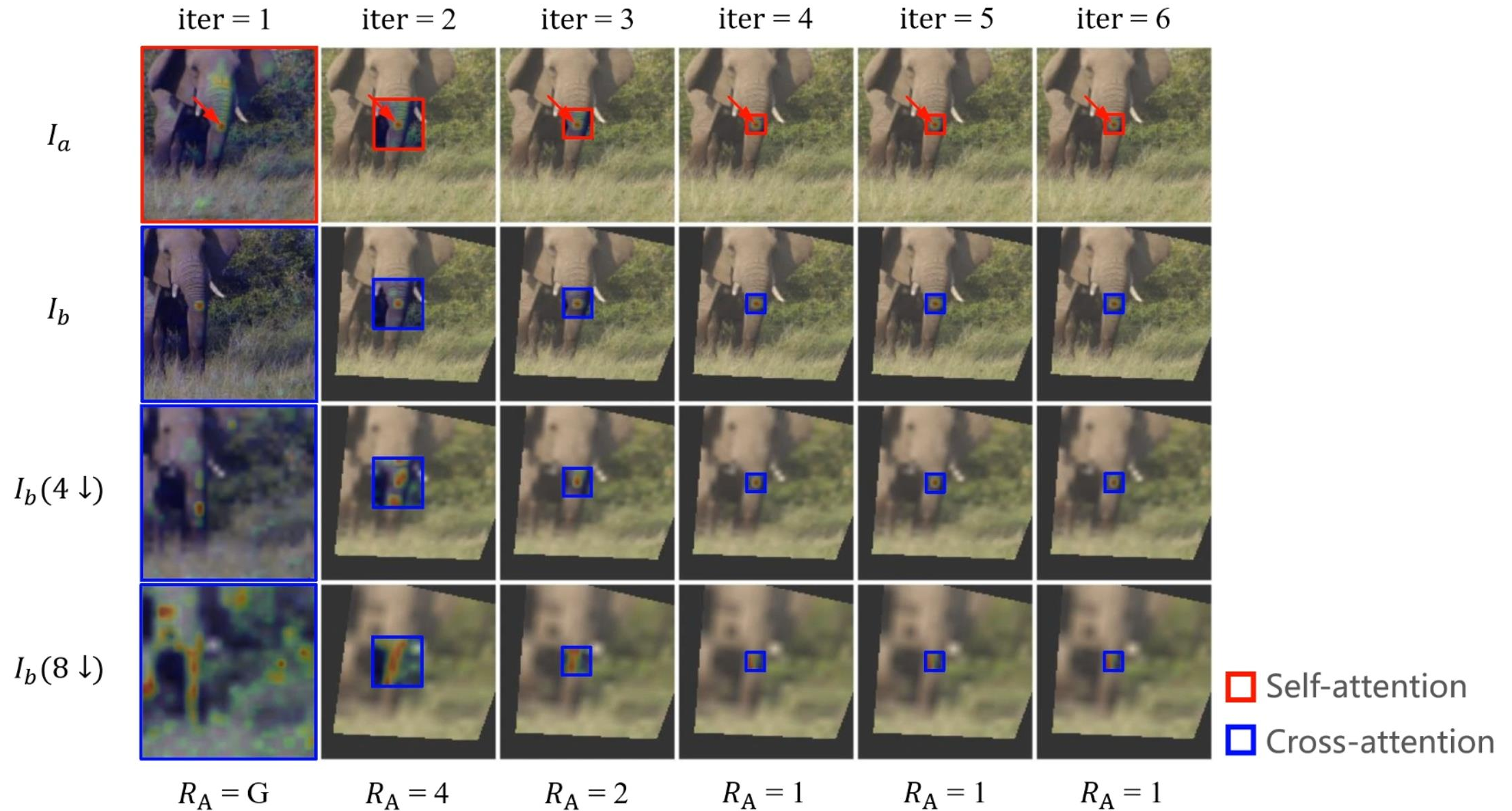
Gradually improve the feature consistency

Homography-guided Image Warping

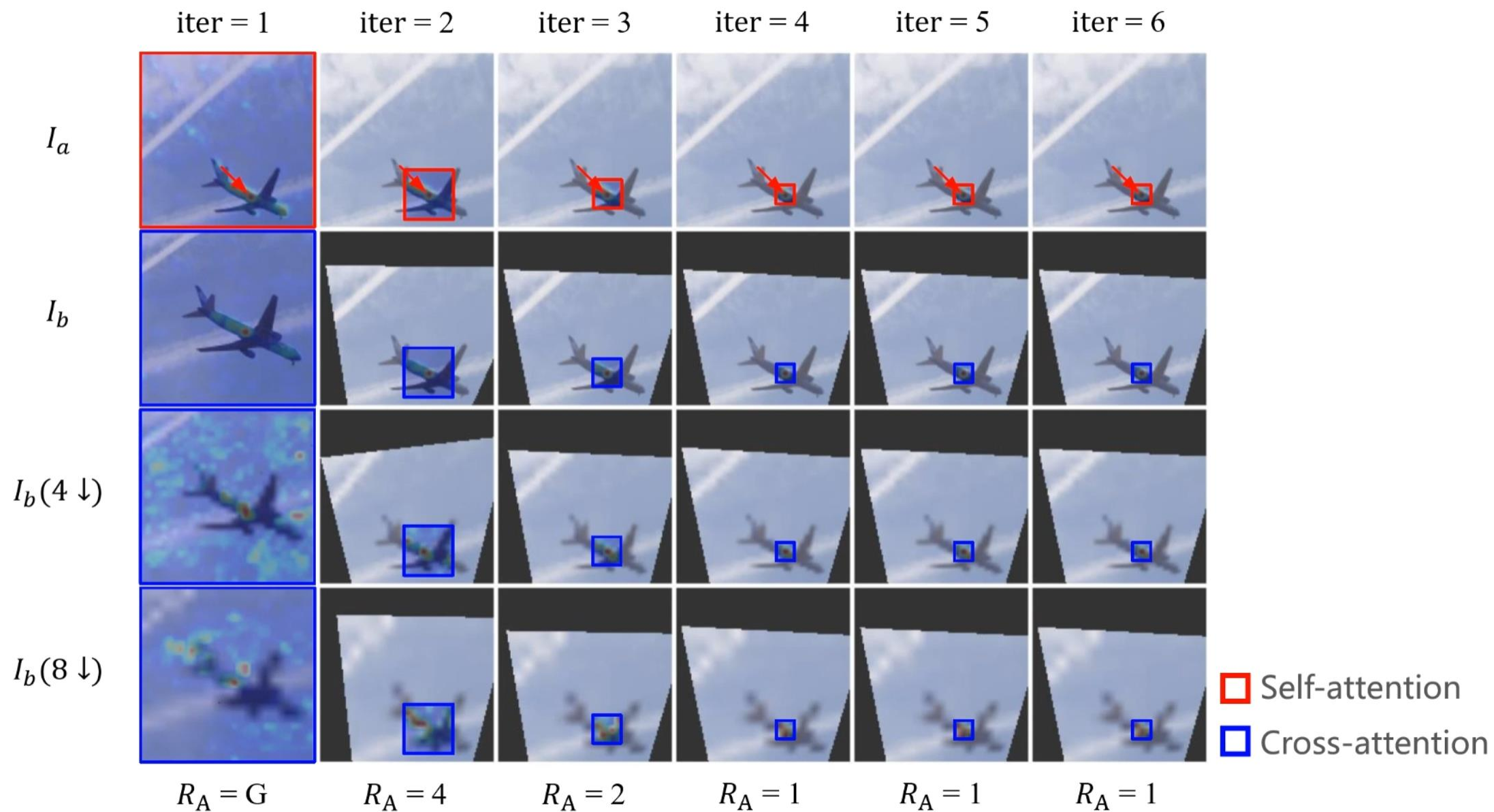


Gradually improve the feature consistency

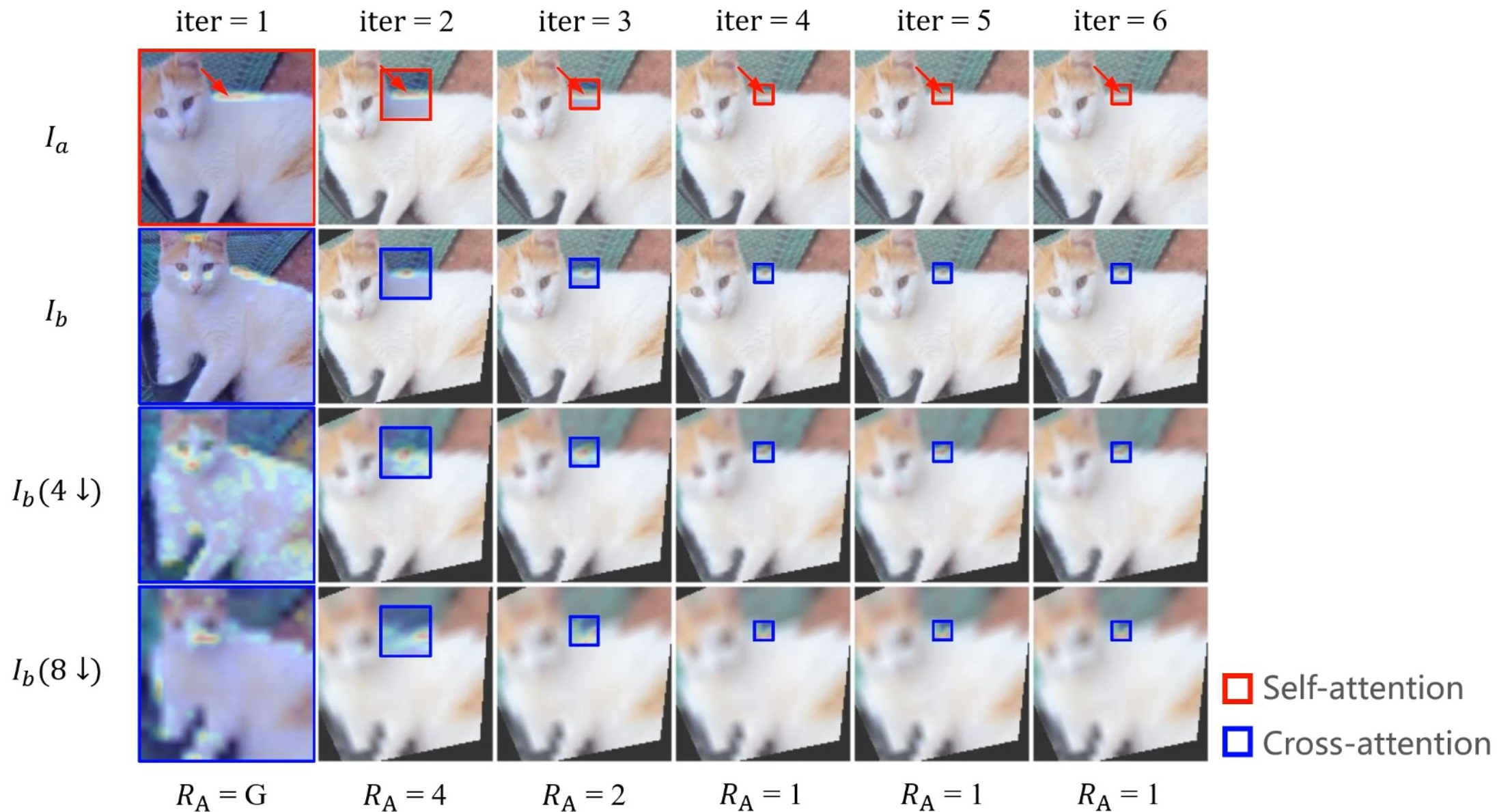
Attention-focusing mechanism



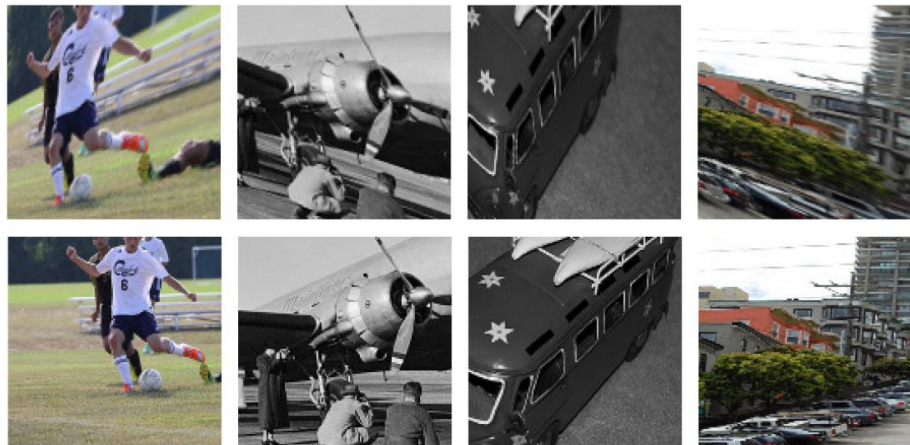
Attention-focusing mechanism



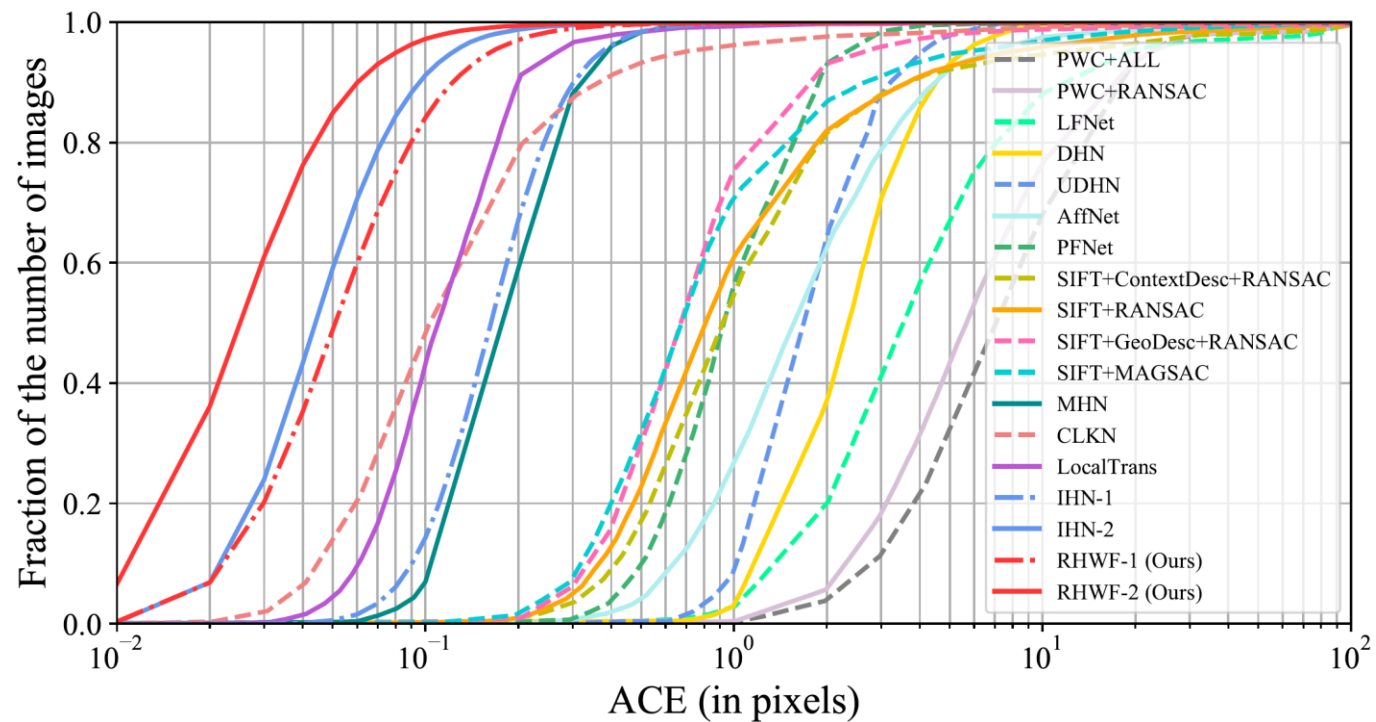
Attention-focusing mechanism



Performance: MSCOCO



MSCOCO image pairs



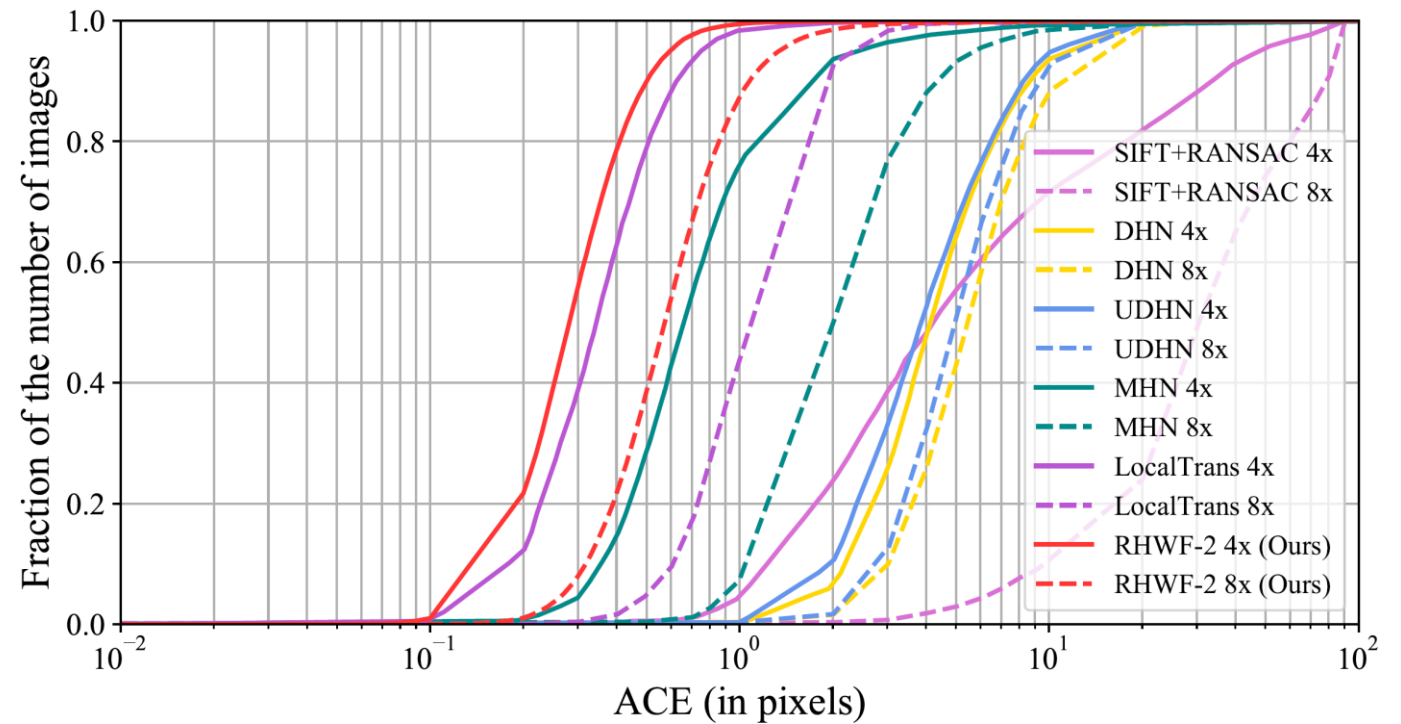
Performance: Cross-Resolution MSCOCO



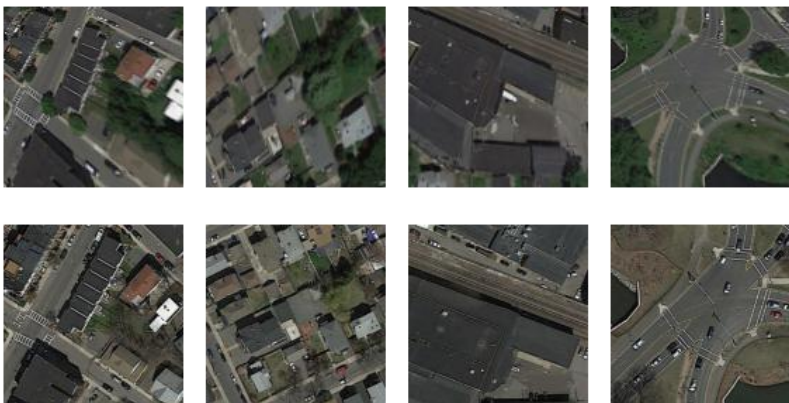
4x cross-resolution MSCOCO image pairs



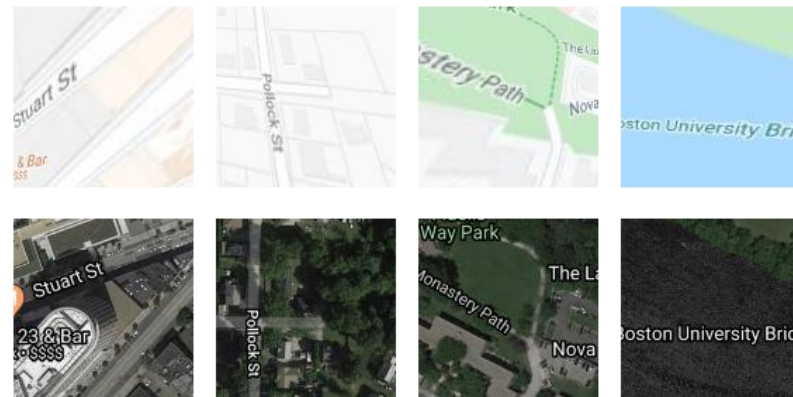
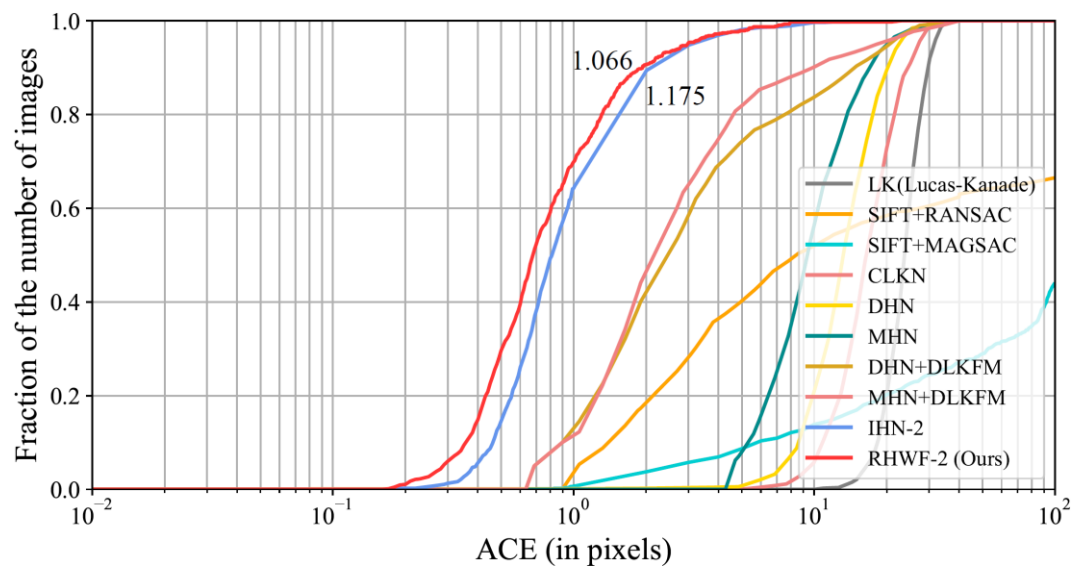
8x cross-resolution MSCOCO image pairs



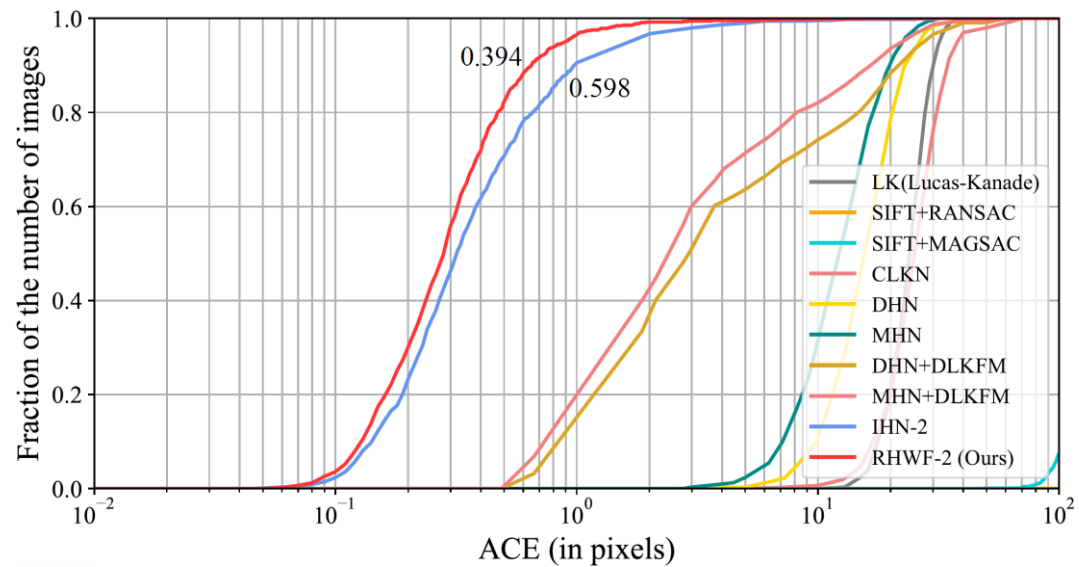
Performance: Cross-Modal Datasets



Google Earth image pairs



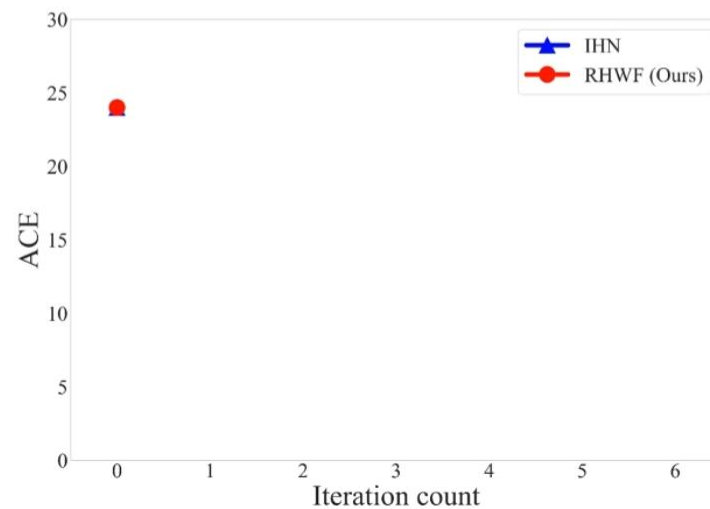
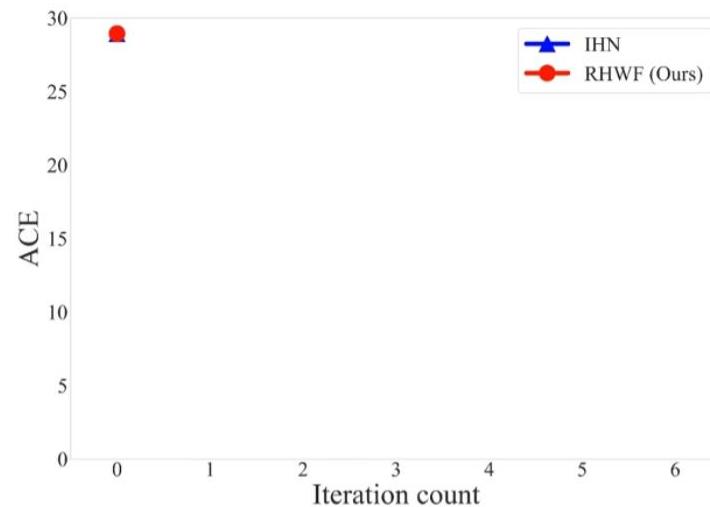
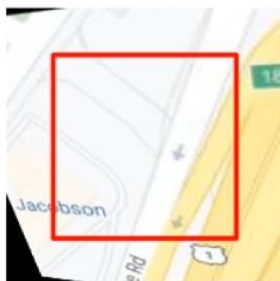
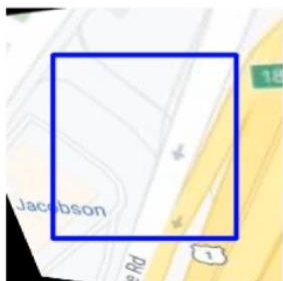
Google Map image pairs



Performance: Cross-Modal Datasets

IHN^[1]

RHWF (Ours)



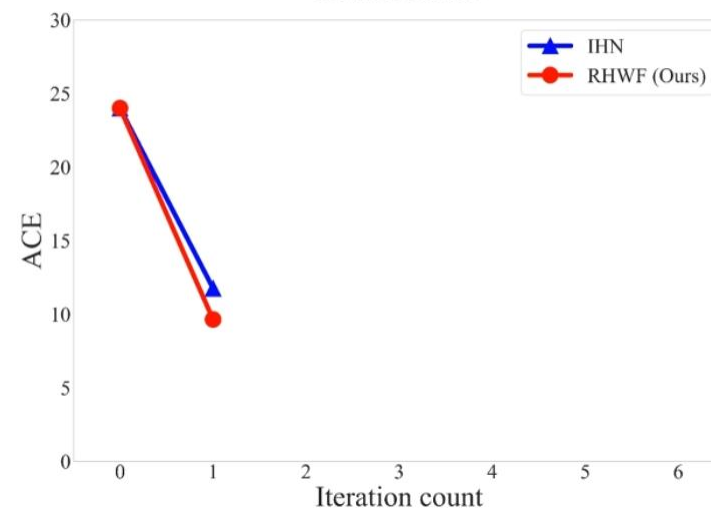
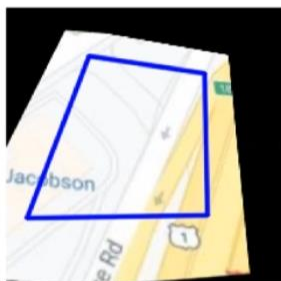
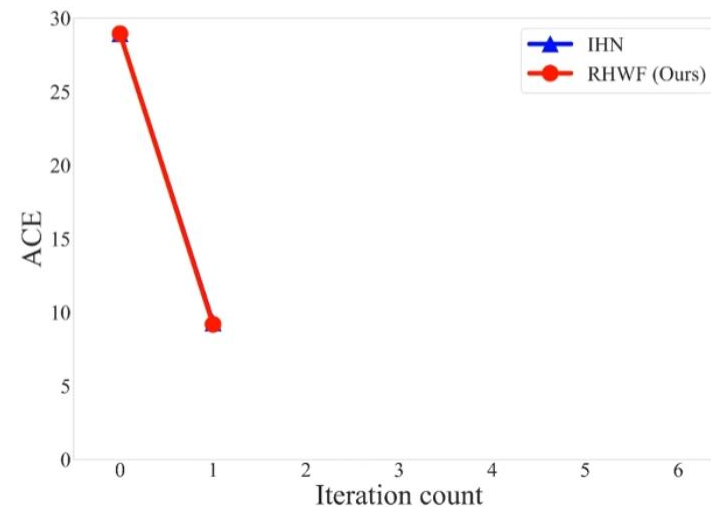
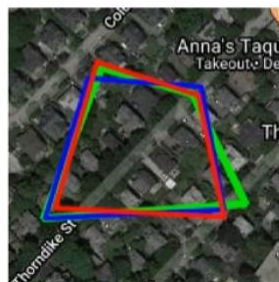
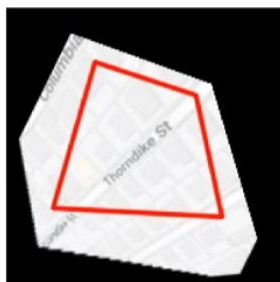
[1] Si-Yuan Cao, Jianxin Hu, Zehua Sheng, and Hui-Liang Shen. Iterative deep homography estimation. CVPR, 2022.

Performance: Cross-Modal Datasets

IHN^[1]



RHWF (Ours)



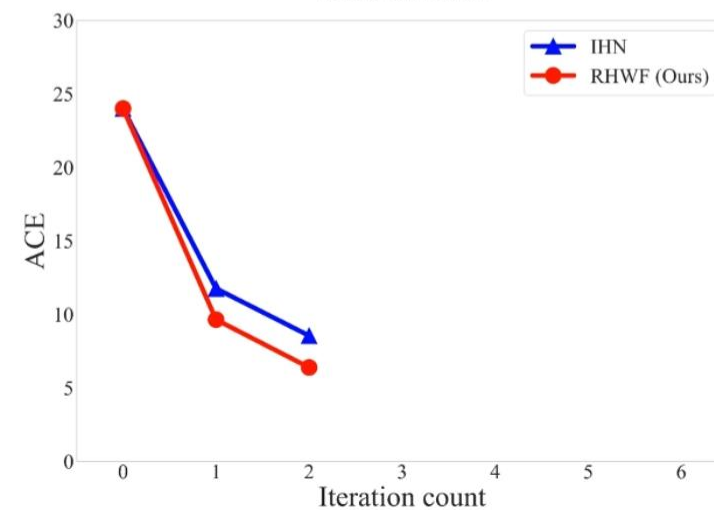
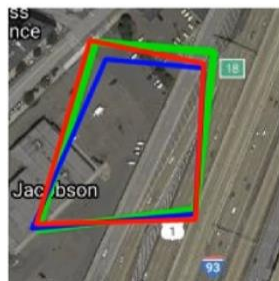
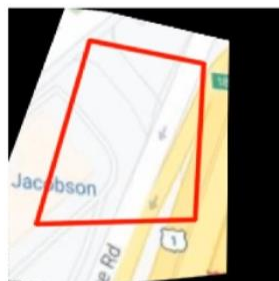
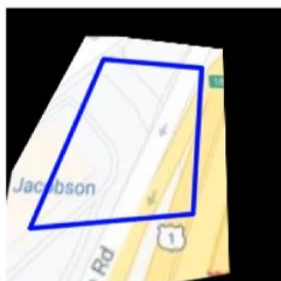
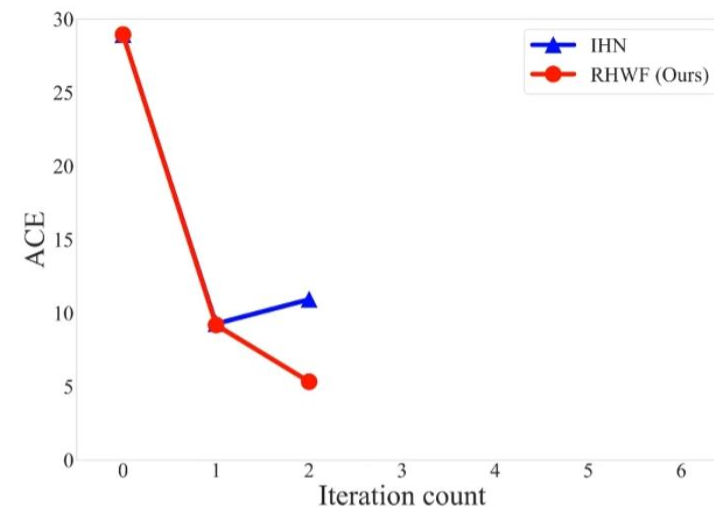
[1] Si-Yuan Cao, Jianxin Hu, Zehua Sheng, and Hui-Liang Shen. Iterative deep homography estimation. CVPR, 2022.

Performance: Cross-Modal Datasets

IHN^[1]



RHWF (Ours)



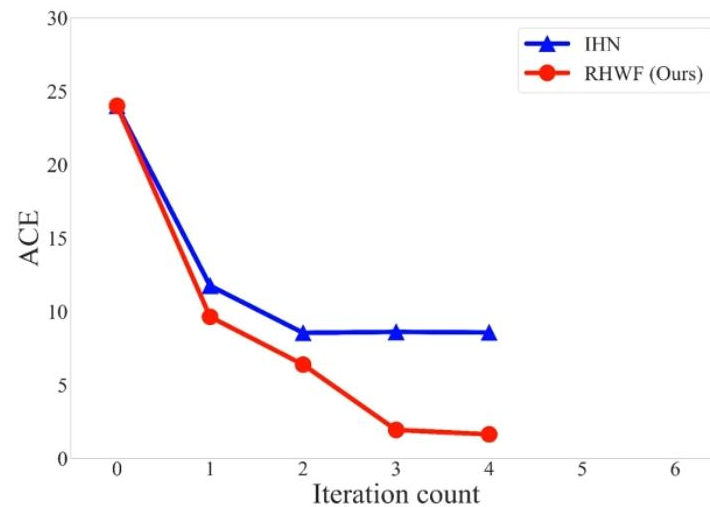
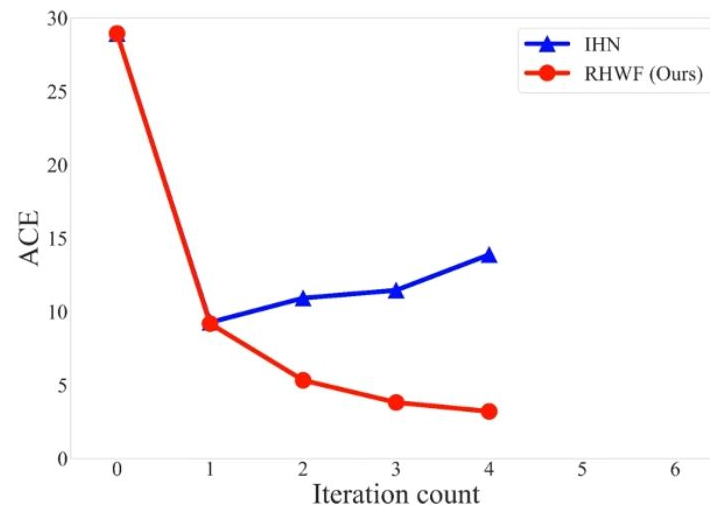
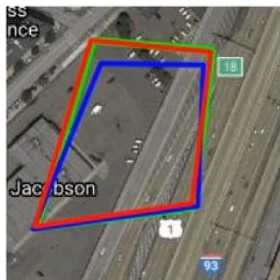
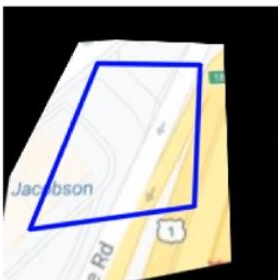
[1] Si-Yuan Cao, Jianxin Hu, Zehua Sheng, and Hui-Liang Shen. Iterative deep homography estimation. CVPR, 2022.

Performance: Cross-Modal Datasets

IHN^[1]



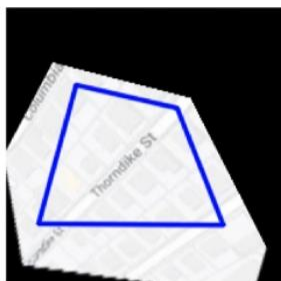
RHWF (Ours)



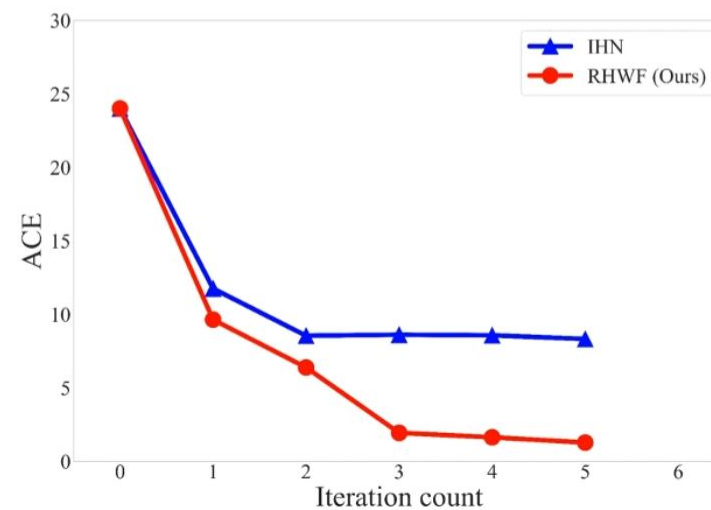
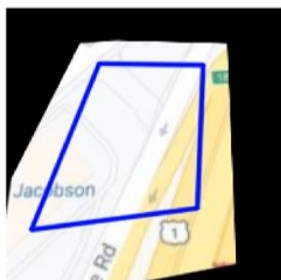
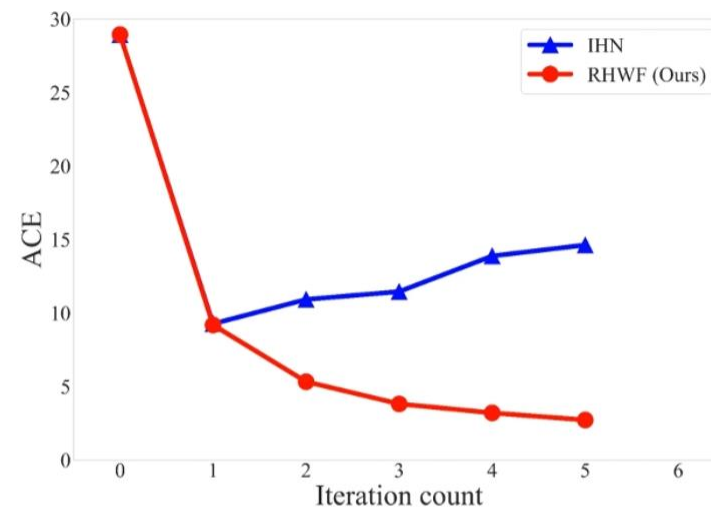
[1] Si-Yuan Cao, Jianxin Hu, Zehua Sheng, and Hui-Liang Shen. Iterative deep homography estimation. CVPR, 2022.

Performance: Cross-Modal Datasets

IHN^[1]



RHWF (Ours)



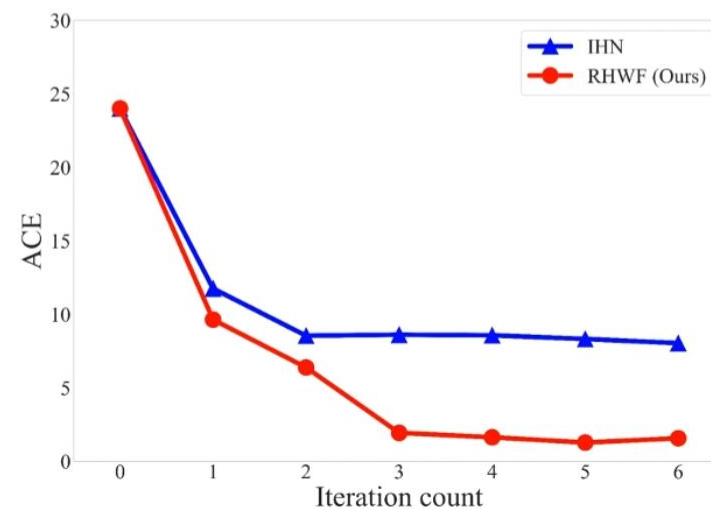
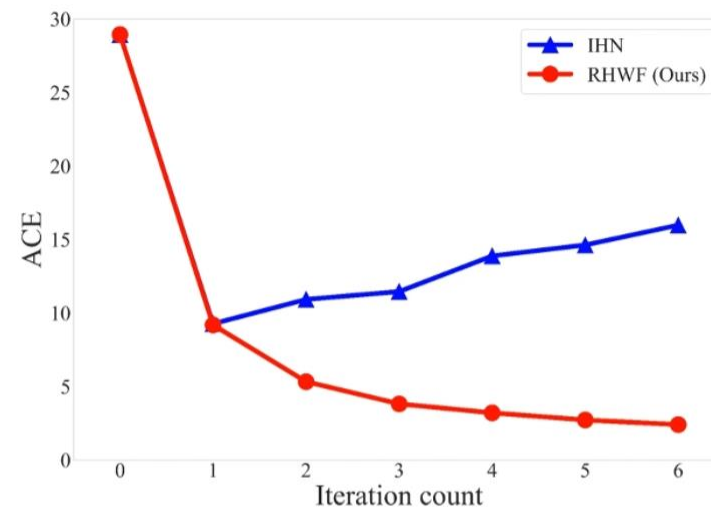
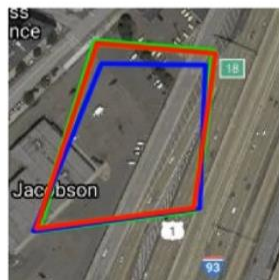
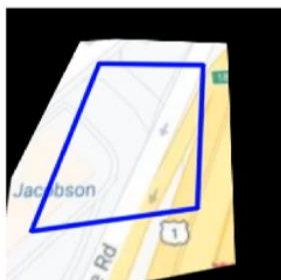
[1] Si-Yuan Cao, Jianxin Hu, Zehua Sheng, and Hui-Liang Shen. Iterative deep homography estimation. CVPR, 2022.

Performance: Cross-Modal Datasets

IHN^[1]



RHWF (Ours)



[1] Si-Yuan Cao, Jianxin Hu, Zehua Sheng, and Hui-Liang Shen. Iterative deep homography estimation. CVPR, 2022.

Ablation: on MSCOCO

Ablation part	Setting	MACE	Parameters
Warping	Feature warping	0.203(↑163.6%)	0.94 M
	Image warping	0.077	0.94 M
Attention	No	0.091(↑18.2%)	0.85 M
	Pure global	0.085(↑10.4%)	0.94 M
	Pure local	0.082(↑6.5%)	0.94 M
	Focus	0.077	0.94 M
Scale	1 scale	0.077(↑97.4%)	0.94 M
	2 scales	0.039	1.29 M

Parameter and FLOPs Comparison

Parameter comparison

RHWF	IHN	LocalTrans	DHN	MHN	UDHN	DLKFM
1.29 M	1.71 M	9.56 M	34.19 M	2.57 M	21.29 M	19.24 M

FLOPs of models with MACEs on MSCOCO

	RHWF	RHWF-3	IHN	IHN-mov	DLKFM
FLOPs	16.96 G	9.64 G	8.34 G	20.32 G	110.51 G
MACE	0.077	0.176	0.191	0.177	0.550

Conclusions

- We propose a novel Recurrent homography estimation framework using Homography-guided image Warping and FocusFormer, dubbed RHWF. The recurrent estimation, homography-guided image warping, and FocusFormer facilitate the functionality of each other.
- With the assistance of homography-guided image warping, the extracted features gradually converge into consistency, and hence boosting the homography estimation accuracy.
- The FocusFormer is proposed to be the fundamental block of the recurrent homography estimation. The attention mechanism in FocusFormer works in a global \rightarrow nonlocal \rightarrow local manner.
- RHWF ranks top on a variety of datasets, including the challenge scenes such as the cross-resolution and cross-modal ones.



浙江大學
ZHEJIANG UNIVERSITY



Thanks for watching!