Pixel-level Semantic Correspondence through Layout-aware Representation Learning and Multi-scale Matching Integration

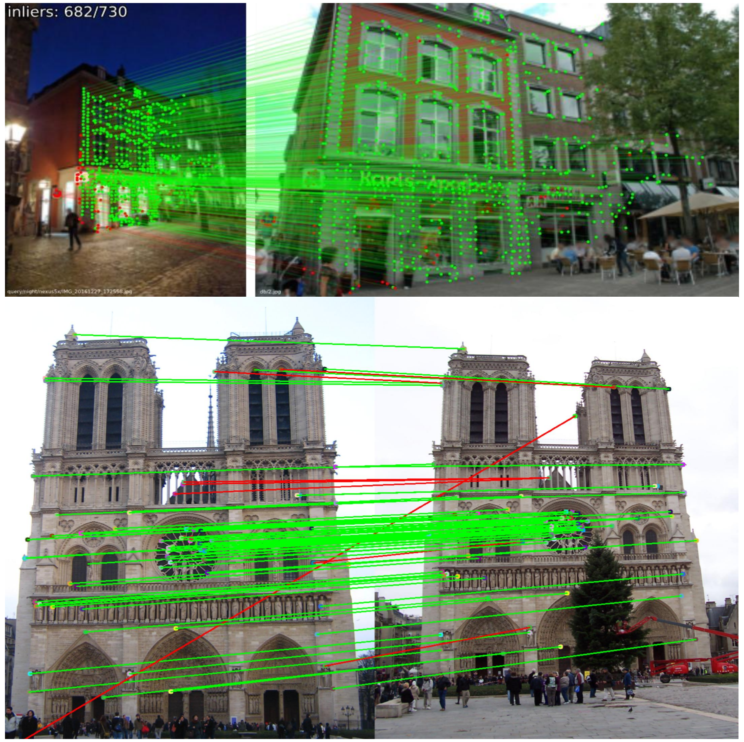
Yixuan Sun[1,2,*], Zhangyue Yin[3,*], Haibo Wang[3], Yan Wang[1,2], Xipeng Qiu[3] , Weifeng Ge[3,†] and Wenqiang Zhang[1,2,3,†]
[1]Academy for Engineering and Technology, Fudan University; [2]Engineering Research Center of AI & Robotics, Ministry of Education, China; [3]School of Computer Science, Fudan University
{wfge, wqzhang}@fudan.edu.cn

https://github.com/YXSUNMADMAX/LPMFlow

CVPR SEATTLE, WA JUNE 17-21, 2024

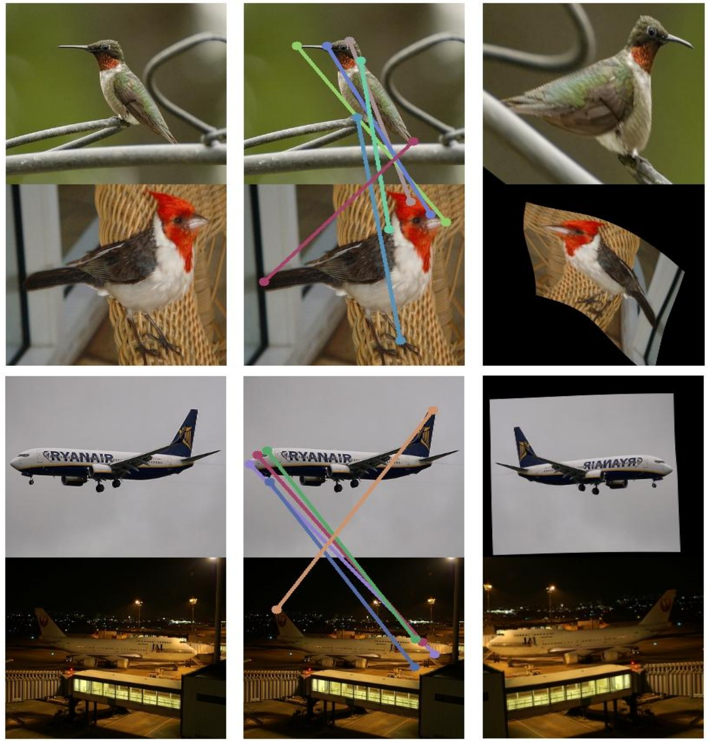
Introduction

- Semantic Correspondence Aims To Establish Pixel-level Correspondence between Semantically Adjacent Image Pair.
- Requires high-quality patch-level representations with aligned semantic spaces; Requires matching representation in high resolution.



Task of Feature Matching

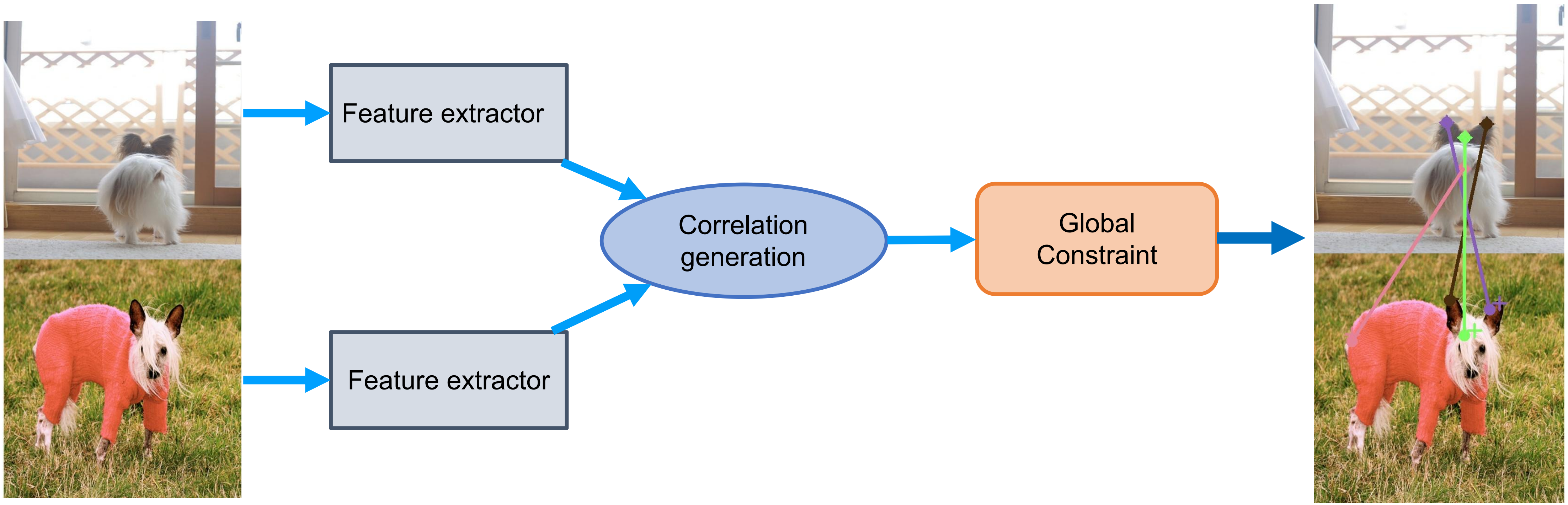
Task of Dense Matching



Task of Semantic Correspondence

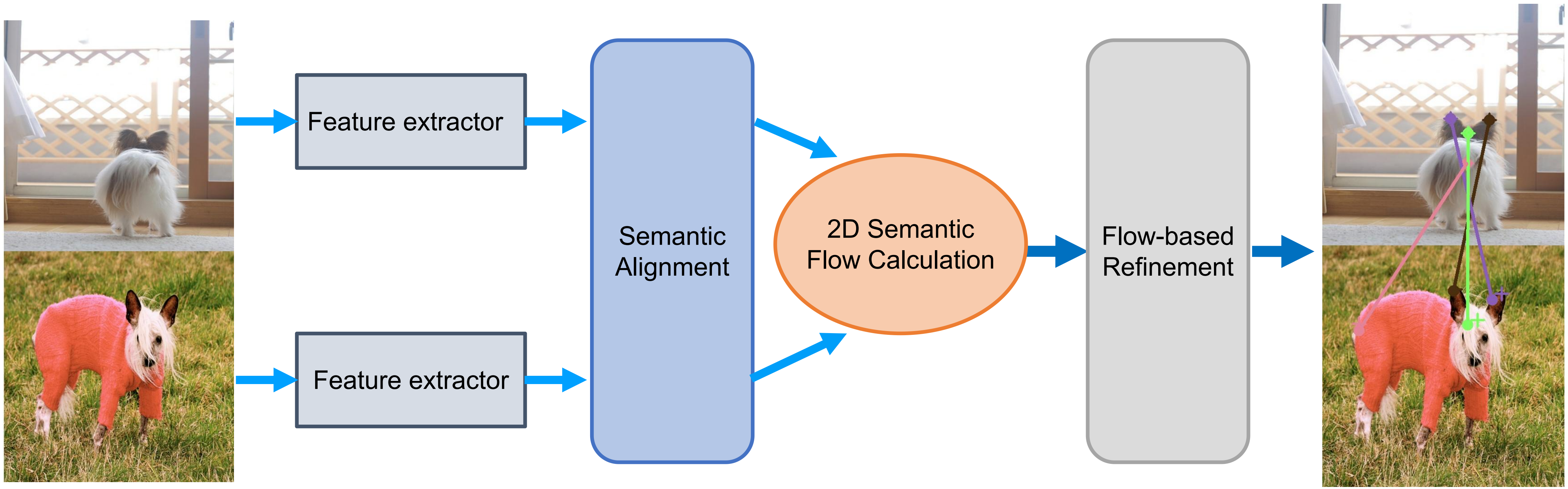
Previous Frameworks

- Siamese Backbone
- 4D Matrix-based Refinement.



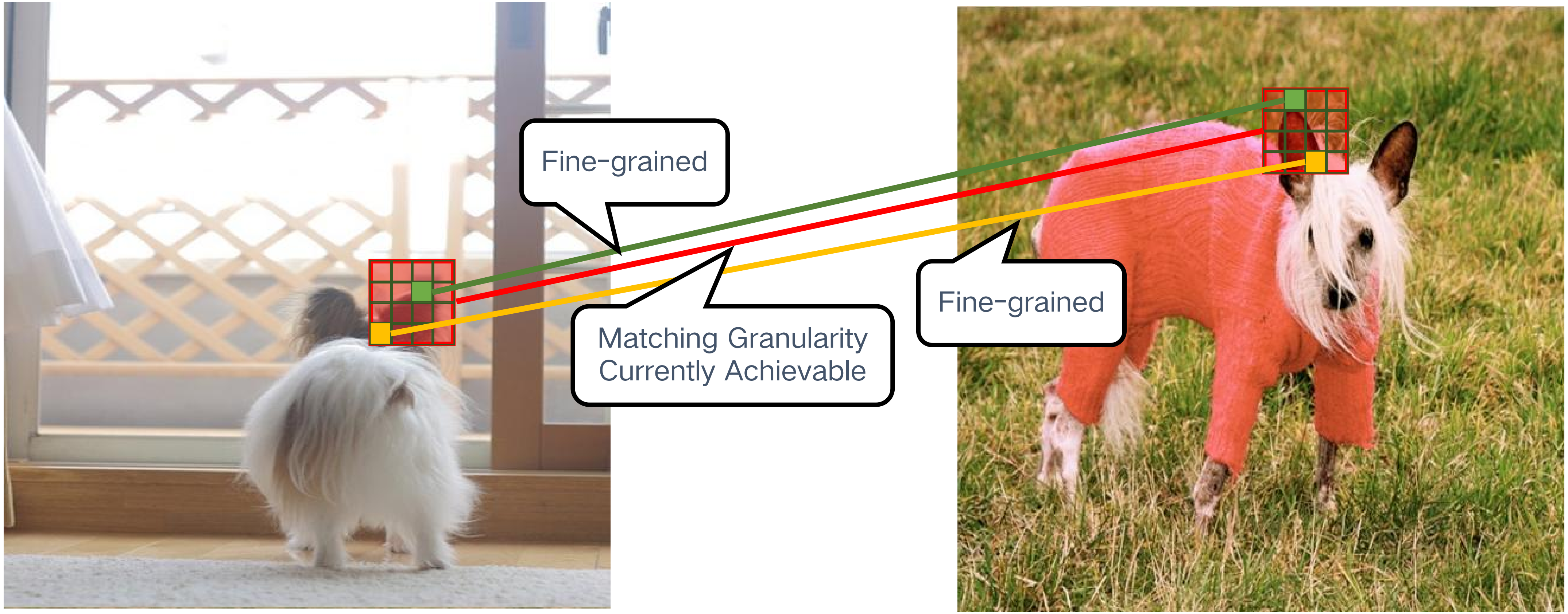
Recent Frameworks

- Siamese Backbone + Semantic Alignment.
- 2D Semantic Flow based Refinement.



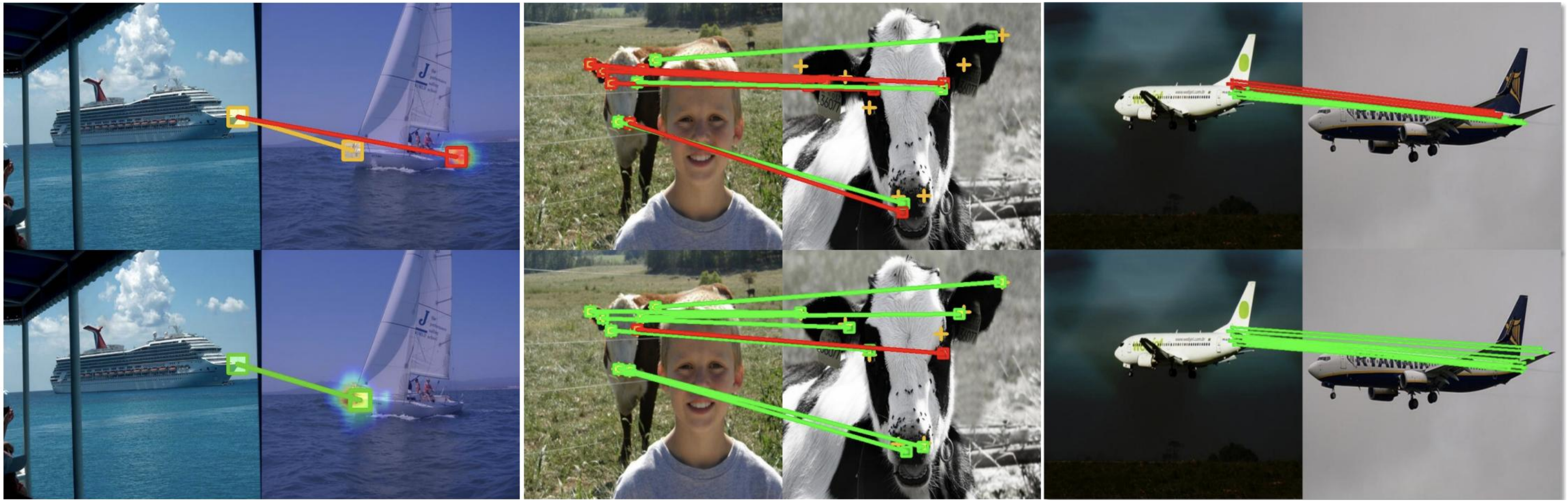
Purpose

- Build up Semantic Correspondence in high resolution (Pixel-level)
- 1/2 or 1 as Original Input Resolution Maximum



New Challenges

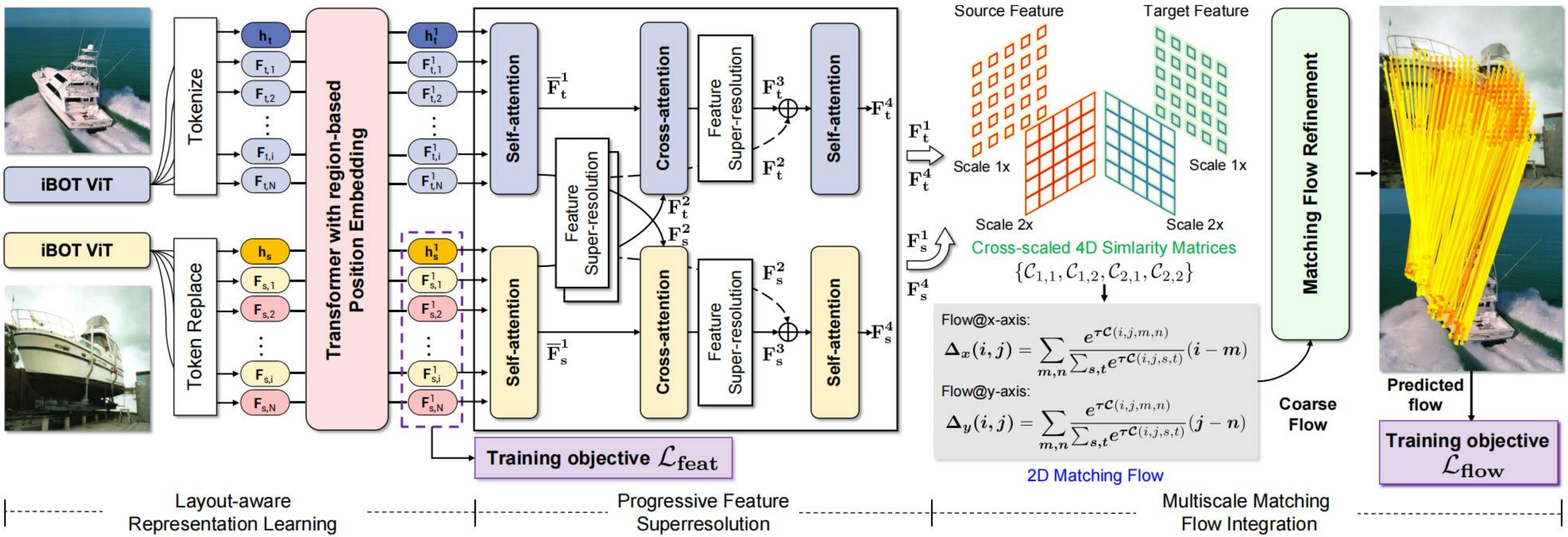
- Semantic regions that share similar appearances are often confused
- Objects in different scales present a challenge in establishing correlations for details
- Nearby pixels are hard to be distinguished



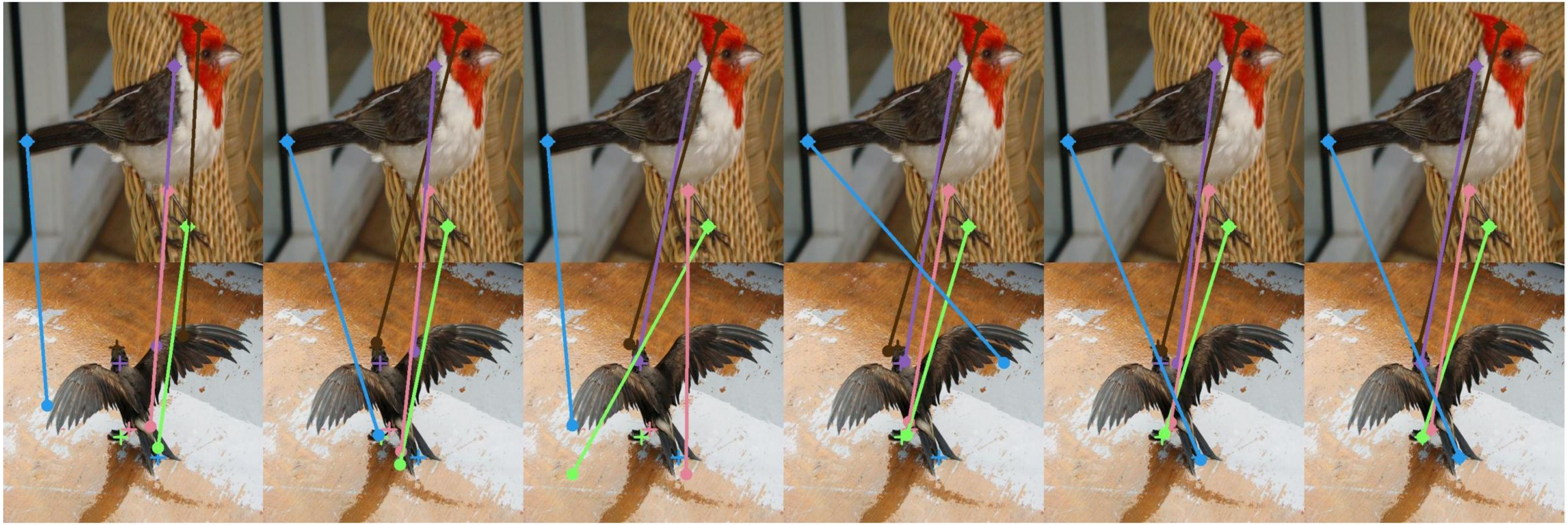
Visualization for the effectiveness of three designed modules on three challenges. Ground truth is indicated in yellow, successes in green, and failures in red.

Our Approach

- Layout-aware Representation Learning
- Progressive Feature Super-Resolution
- Multi-scale Matching Flow Integration



Results (Comparison with other methods)



SCOT CATs MMNet ACTR LPMFlow(ours) Groundtruth

LPMFlow can clearly overcome the significant geometric appearance changes and distinguish local areas with similar appearance based on geometric information.

Results (Comparison with other methods)

Input Images | CATs | MMNet | ACTR | LPMFlow(ours)

LPMFlow can provide better fine-grained dense correspondence.

Results (Comparison with other methods)

Source Target SCOT CATs MMNet ACTR LPMFlow(ours) Groundtruth

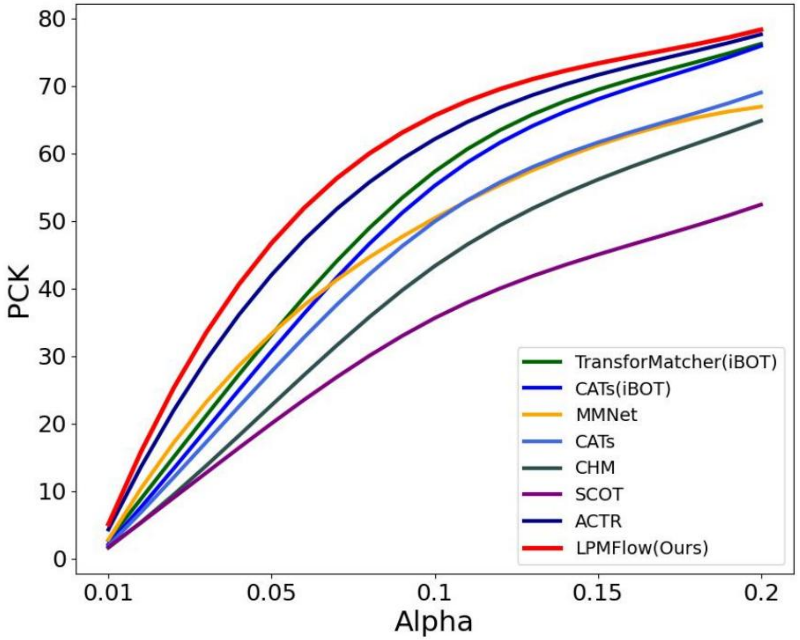
Liu et al. 2020; Cho et al. 2021; Zhao et al. 2021; Sun et al. 2023;

LPMFlow can provide better fine—grained dense correspondence.

Evaluation

The Input Resolution of Our LPMFlow is 256x256.

Method	Description		Performance					Generalizability		Efficiency			
			Spair-71K		PF-PASCAL			PF-WILLOW		TITAN RTX: 24GB			
	Multi Scale	Corr Format	α: bbox		α: img			α: bbox	α: bkp	Params(M)		Mem	Time
			0.05	0.1	0.05	0.1	0.15	0.1	0.1	Head	Total	(GB)	(ms)
NC-Net[32]	✗	4D Mtrx	-	20.1	54.3	78.9	86.0	-	67.0	0.2	27.6	1.2	222.9
SCOT[21]	✗	4D Mtrx	20.0	35.6	63.1	85.4	92.7	-	76.0	-	44.5	4.6	133.5
DHPF[27]	✓	4D Mtrx	-	37.3	75.7	90.7	95.0	77.6	71.0	5.8	50.3	1.6	58.2
CHM[25]	✗	4D Mtrx	22.7	46.3	80.1	91.6	94.9	79.4	69.6	7.1	94.1	1.7	55.3
CATs[3]	✓	2D Flow	27.7	49.9	75.4	92.6	96.4	79.2	69.0	4.7	49.2	2.0	45.4
MMNet-FCN[48]	✓	4D Mtrx	33.3	50.4	81.1	91.6	95.9	-	-	10.3	64.7	5.4	258.6
TransMatcher[16]	✓	4D Mtrx	-	53.7	80.8	91.8	-	65.3	76.0	0.9	87.9	2.7	54.2
CATs* [3]	✓	2D Flow	30.7	55.2	77.8	93.1	96.8	86.3	79.5	5.7	90.7	2.8	54.2
TransMatcher* [16]	✓	4D Mtrx	33.1	57.9	77.3	93.3	96.6	84.3	78.3	1.6	86.6	2.4	48.5
ACTR* [38]	✗	2D Flow	42.0	62.1	81.2	94.0	97.0	87.2	79.9	87.8	172.8	3.9	84.1
LPMFlow*	✓	2D Flow	**46.7**	**65.6**	**82.4**	**94.3**	**97.2**	**87.6**	**81.0**	93.9	178.9	3.8	85.7



Yields large Improvements over several benchmarks. Having the best Generalizability.

Evaluation

The Input Resolution of Our LPMFlow is 256x256.

Methods	aero.	bike	bird	boat	bott.	bus	car	cat	chai	cow	dog	hors.	mbik.	pers.	plan.	shee.	trai.	tv	all
NC-Net[12]	17.9	12.2	32.1	11.7	29.0	19.9	16.1	39.2	9.9	23.9	18.8	15.7	17.4	15.0	14.8	9.6	24.2	31.1	20.1
SCOT [6]	34.9	20.7	63.8	21.1	43.5	27.3	21.3	63.1	20.0	42.9	42.5	31.1	29.8	35	27.7	24.4	48.4	40.8	35.6
DHPF [11]	38.4	23.8	68.3	18.9	42.6	27.9	20.1	61.6	22.0	46.9	46.1	33.5	27.6	40.1	27.6	28.1	49.5	46.5	37.3
CHM [8]	49.6	29.3	68.7	29.7	45.3	48.4	39.5	64.9	20.3	60.5	56.1	46.0	33.8	44.3	38.9	31.4	72.2	55.5	46.3
CATs [2]	52.0	34.7	72.2	34.3	49.9	57.5	43.6	66.5	24.4	63.2	56.5	52.0	42.6	41.7	43.0	33.6	72.6	58	49.9
MMNet[15]	55.9	37.0	65.0	35.4	50	63.9	45.7	62.8	28.7	65.0	54.7	51.6	38.5	34.6	41.7	36.3	77.7	62.5	50.4
TMatcher[5]	59.2	39.3	73.0	41.2	52.5	66.3	55.4	67.1	26.1	67.1	56.6	53.2	45.0	39.9	42.1	35.3	75.2	68.6	53.7
CATs* [2]	56.7	41.3	77.8	35.0	54.8	59.8	45.2	69.9	31.4	63.7	57.6	62.5	46.7	49.1	43.2	43.5	76.4	64.1	55.2
TMatcher* [5]	57.1	47.4	**83.5**	42.3	56.8	57.0	55.4	75.3	34.5	66.1	64.2	60.2	52.8	55.2	40.5	46.0	75.1	65.8	57.9
ACTR* [13]	65.1	48.5	82.3	**50.4**	55.9	65.3	63.1	72.8	35.8	74.1	70.3	68.9	58.6	**57.1**	46.8	49.5	84.4	73.3	62.1
LPMFlow*	**71.4**	**54.8**	83.2	50.3	**57.0**	**75.4**	**68.9**	**79.3**	**41.1**	**78.4**	**74.1**	**73.7**	**58.7**	56.9	**48.7**	**54.7**	**87.5**	**74.6**	**65.6**

Yields large Improvements on a challenging dataset.

Reach best result on 15/18 sub-classes.

Ablation Results

The Input Resolution of Our LPMFlow is 256x256.

LARL	PFSR	MMFI	SPair-71K $\alpha_{bbox} = 0.1$
✓	✓	✓	65.6
✗	✓	✓	63.2 (2.4↓)
✓	✗	✓	62.0 (3.6↓)
✓	✓	✗	63.9 (1.7↓)

Methods	SPair-71K $\alpha_{bbox} = 0.1$
LPMFlow	65.6
w/o Gradual Guidance of RPTC	64.5 (1.1↓)
w/o Self Contrastive Loss	63.9 (1.7↓)
w/o Region-based PE	64.8 (0.8↓)

Methods	SPair-71K $\alpha_{bbox} = 0.1$
LPMFlow	65.6
w/o Interactive Super-Resolution	64.1 (1.5↓)
w/o Internal Super-Resolution	63.8 (1.8↓)
w/o Feature Super-Resolution block	63.4 (2.2↓)

Methods	SPair-71K $\alpha_{bbox} = 0.1$
LPMFlow	65.6
w/o Multi-Scale Flow Integration	64.3 (1.3↓)
w/o C2F Refinement (16×16)	64.6 (1.0↓)
w/o C2F Refinement (4×4)	64.0 (1.6↓)

Pixel-level Semantic Correspondence through Layout-aware Representation Learning and Multi-scale Matching Integration

Thank You

Academy for Engineering and Technology, Fudan University; Engineering Research Center of AI & Robotics, Ministry of Education, China; School of Computer Science, Fudan University
{wfge, wqzhang}@fudan.edu.cn