JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

TUE-PM-286

MISC210K: A Large-Scale Dataset for Multi-Instance Semantic Correspondence

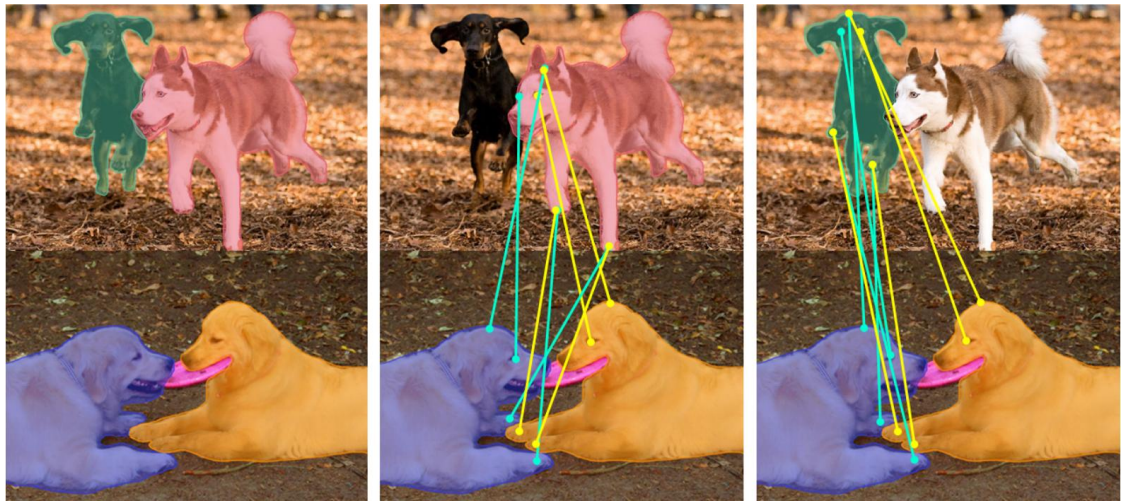
Yixuan Sun[1,*], Yiwen Huang[2,*], Haijing Guo[2], Yuzhou Zhao[2], Runmin Wu[3], Yizhou Yu[3], Weifeng Ge[2,†], Wenqiang Zhang[1,2,†]

[1]Academy of Engineering & Technology, Fudan University, Shanghai, China
[2]School of Computer Science, Fudan University, Shanghai, China
[3]The University of Hong Kong, Hong Kong, China
{wfge, wqzhang}@fudan.edu.cn

Introduction

- Existing semantic correspondence datasets mainly focus on one-to-one object matching.
- However, it is not suitable for real-world applications.

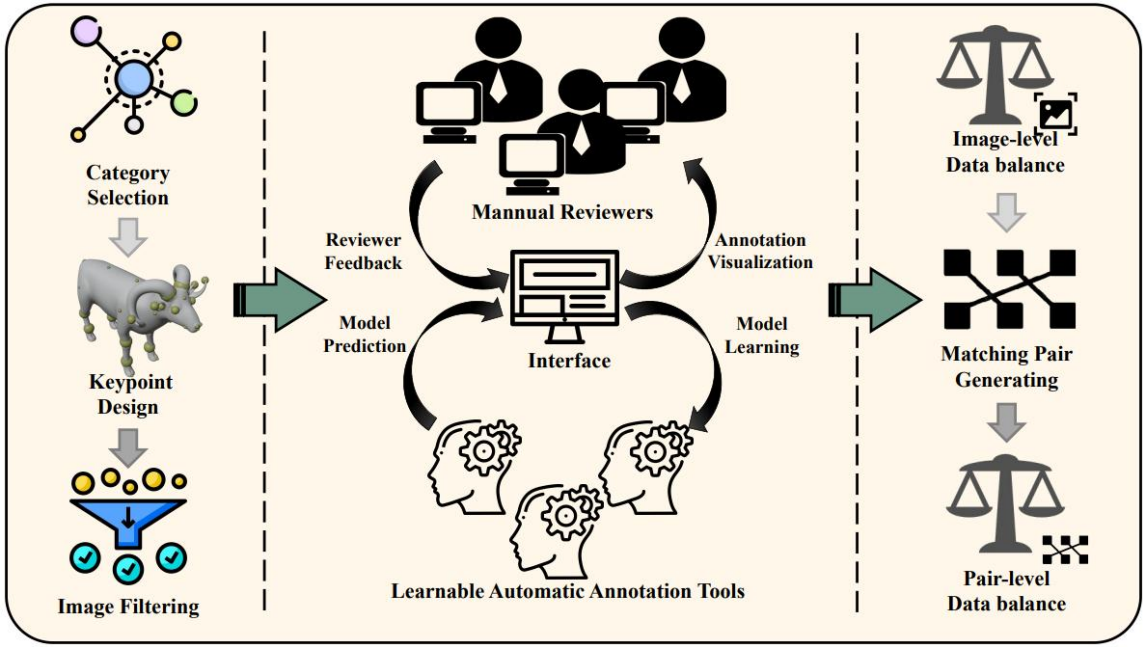
Spair-71k (Single to Single)



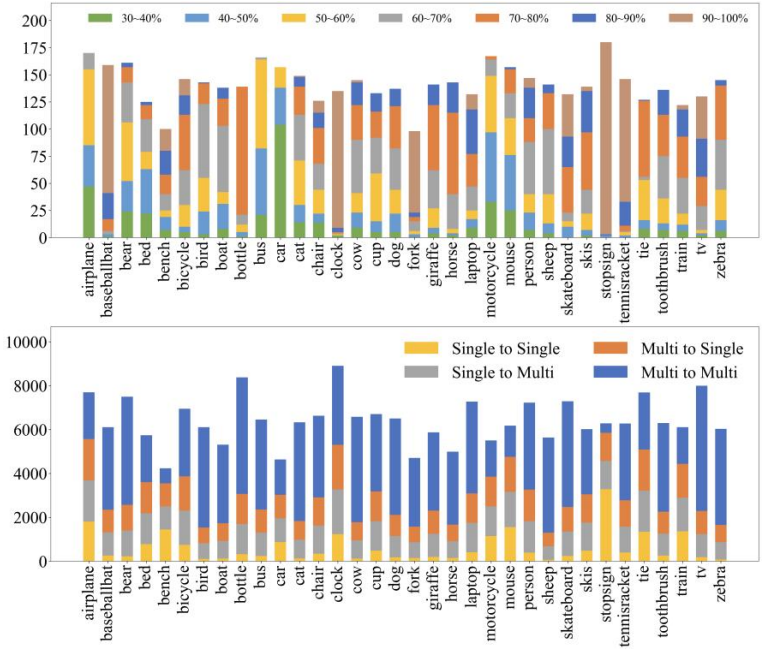
MISC210K (Multi-to-Multi)

Introduction

- MISC210K Dataset:
  - 218,179 image pairs across 34 object categories.
  - Multi-to-multi matching cases.
  - More complicated annotations, larger scale, and more challenging variations.



Overview of dataset construction pipeline

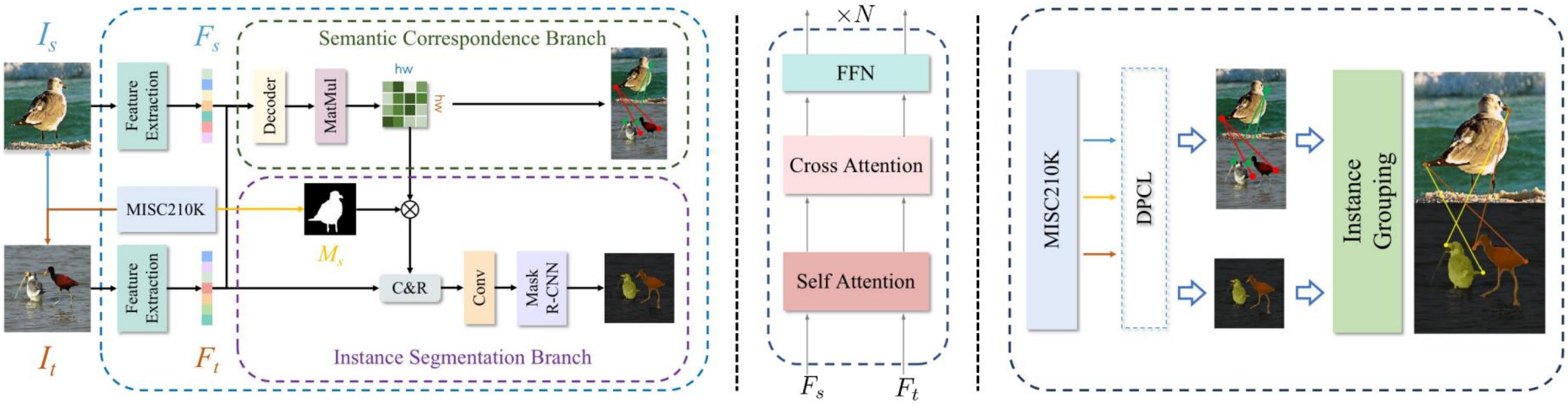


Statistics

Challenging Example

Introduction

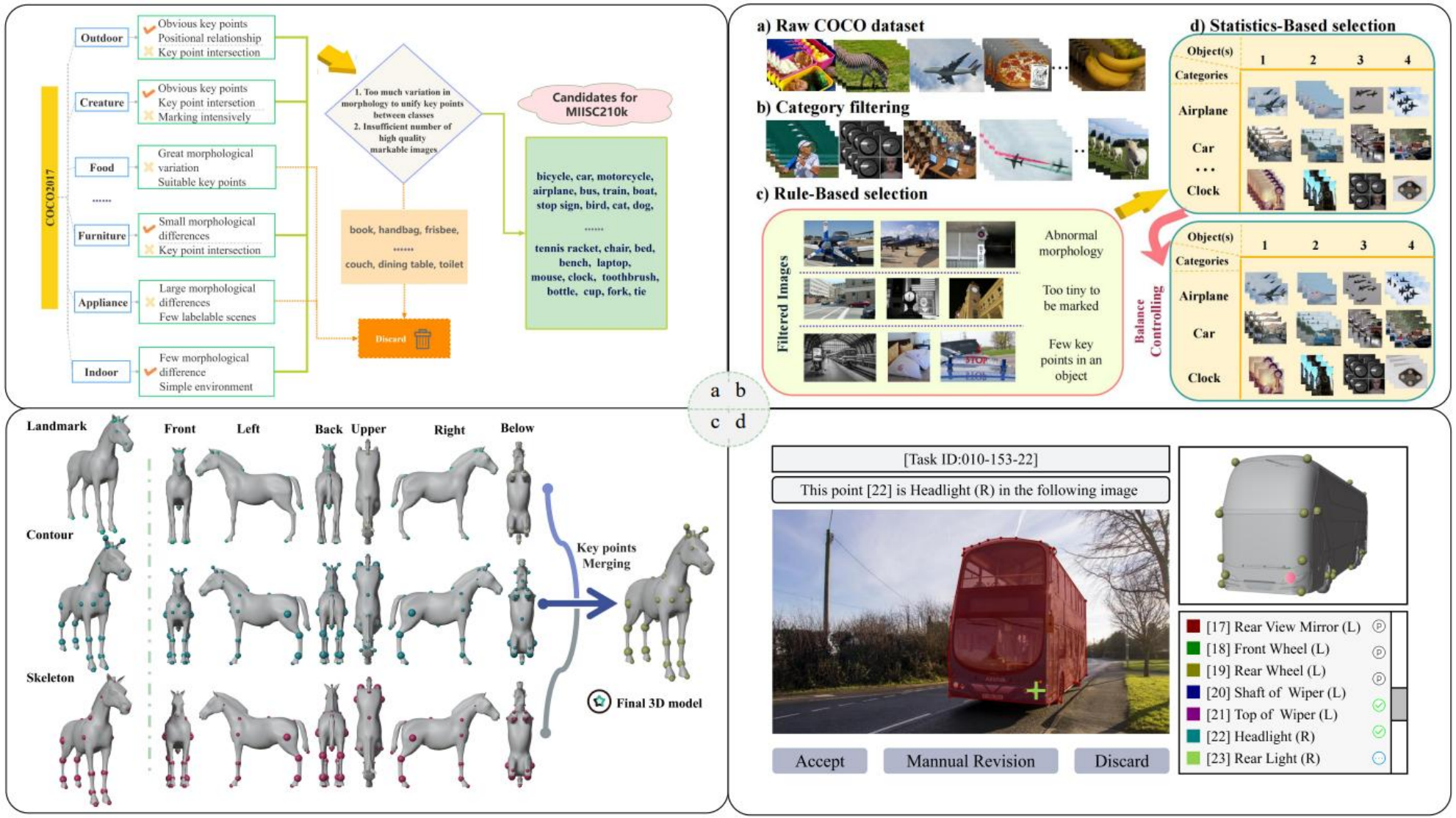
- Dual-Path Collaborative Learning (DPCL) Pipeline:
  - Extract discriminative features.
  - Alleviate uncertainty in the number of matching keypoints.
  - Handle occlusion and interlacing.



(a) Dual-Path Collaborative Learning Pipeline (b) Decoder (c) MISC Inference Pipeline

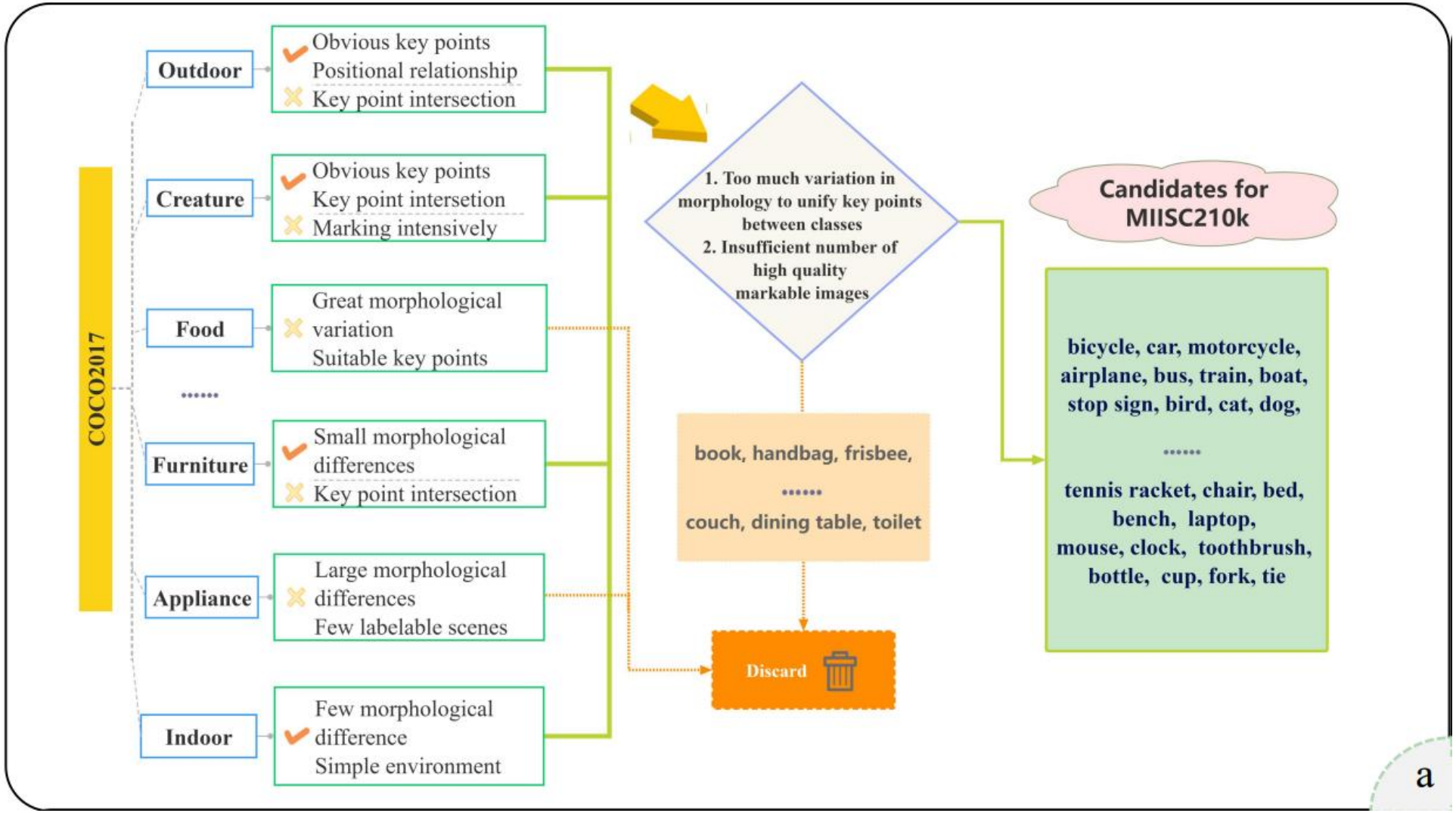
MISC210K Dataset

- Task Definition and Design Protocols:



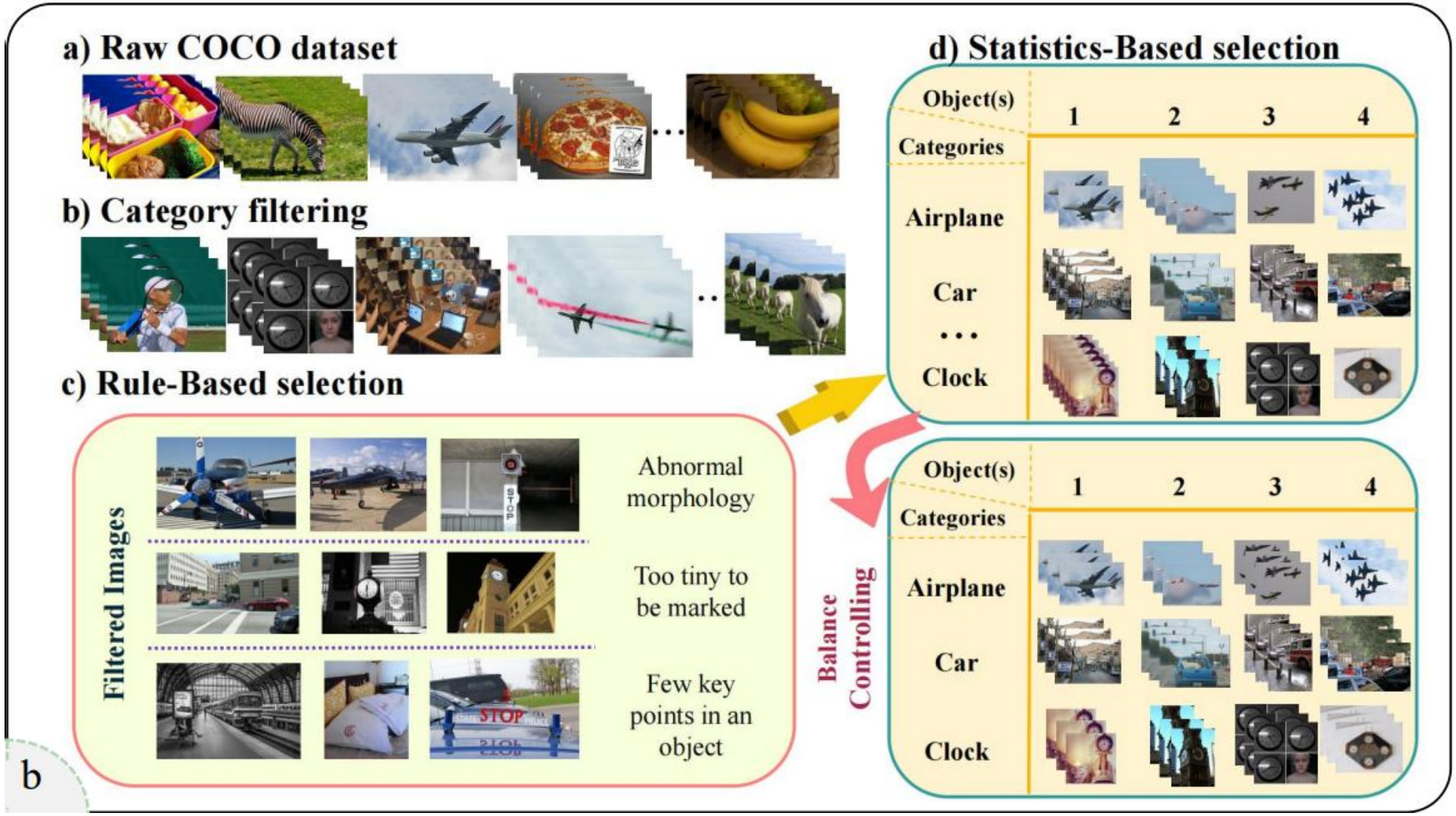
MISC210K Dataset

- Category and Image Selection:



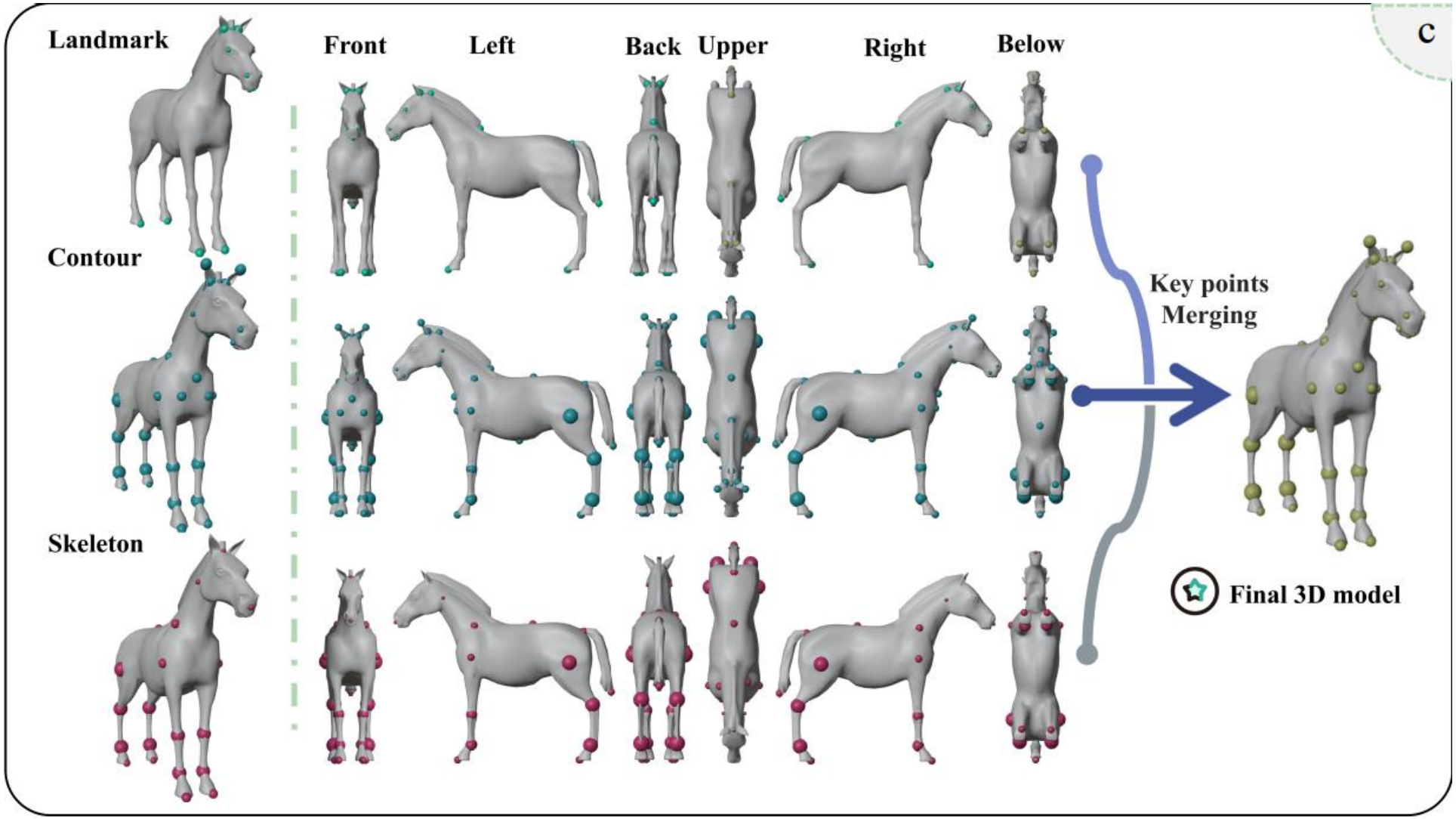
MISC210K Dataset

- Category and Image Selection:



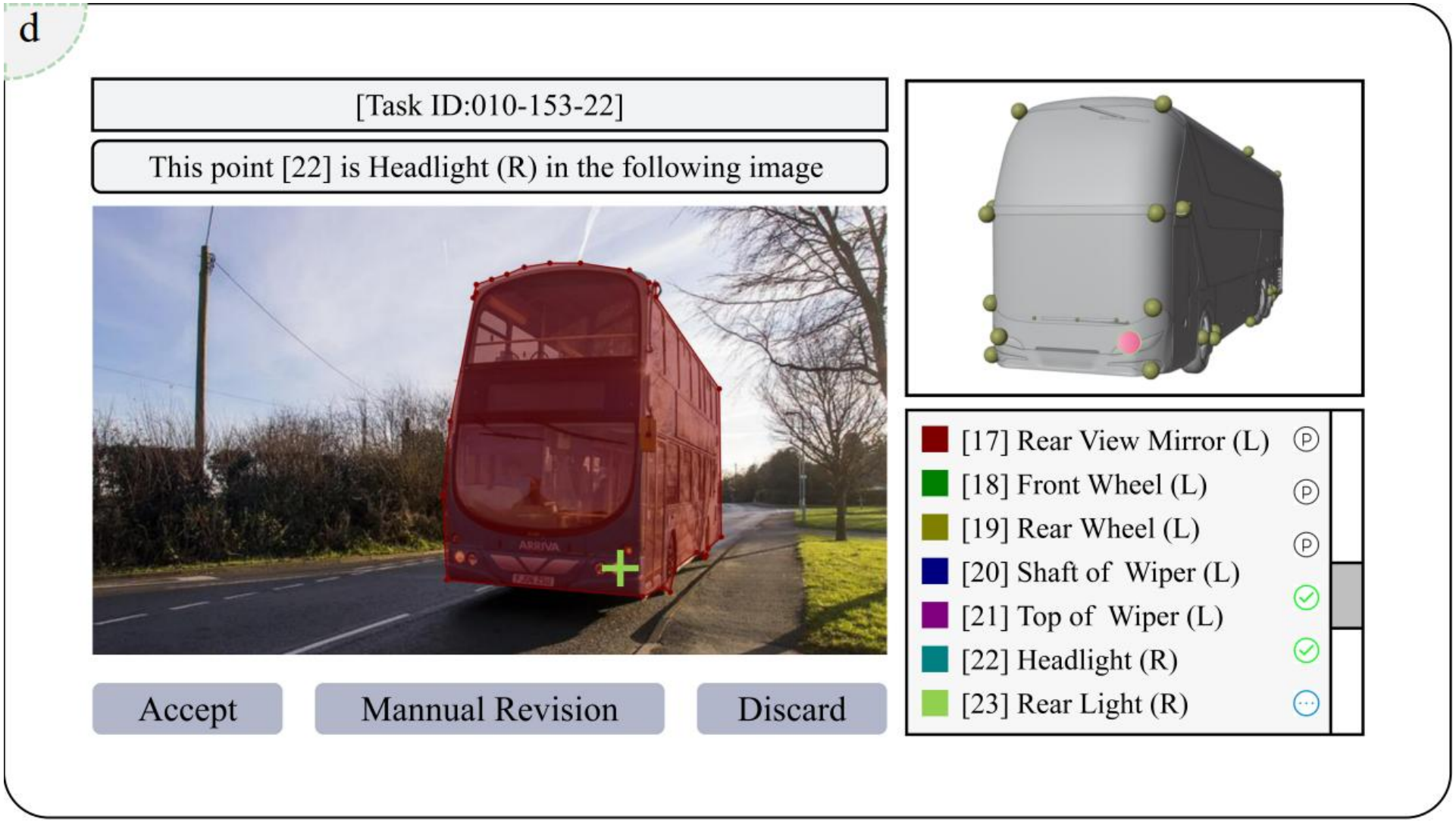
MISC210K Dataset

- Category Keypoint System Definition:



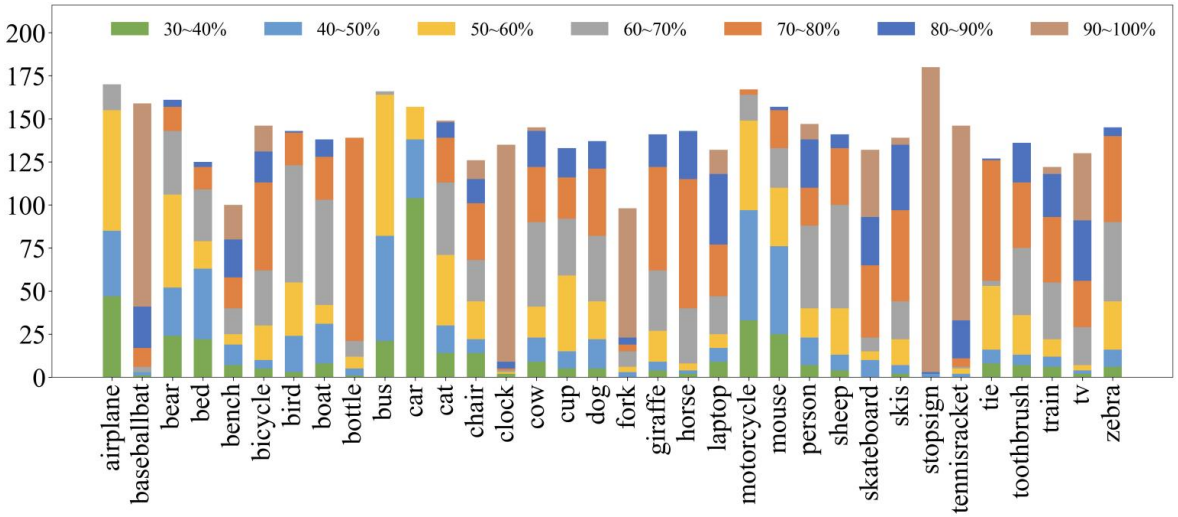
MISC210K Dataset

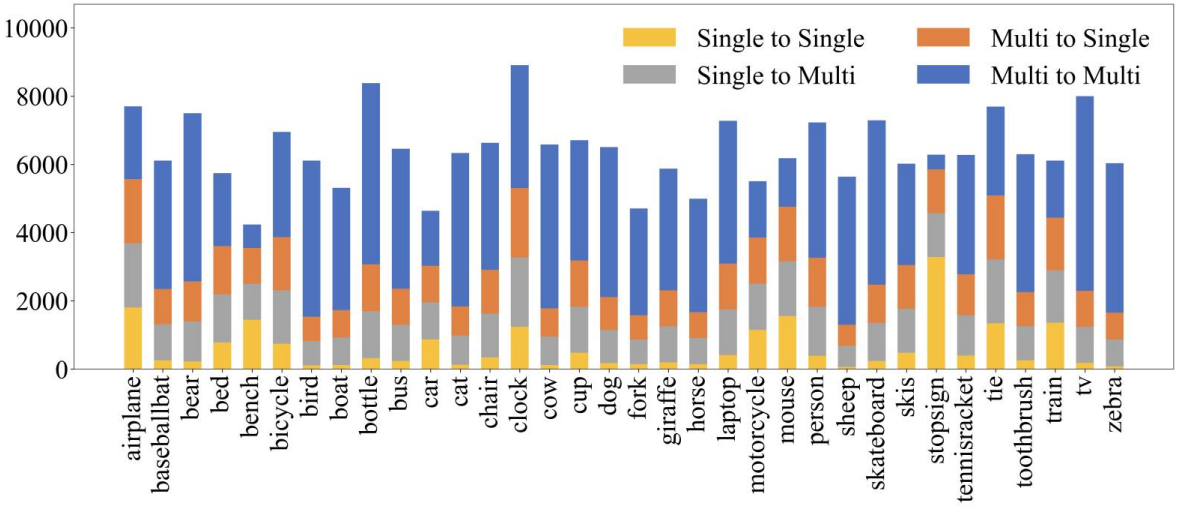
- Human-machine Collaborative Annotation:

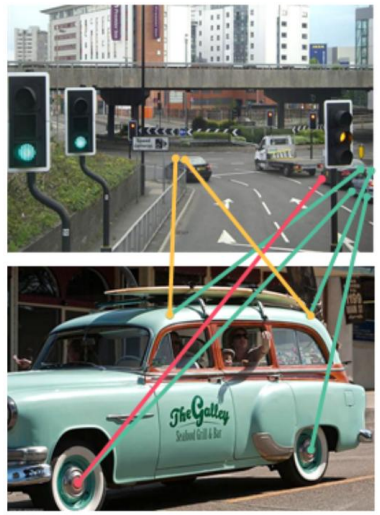


MISC210K Dataset

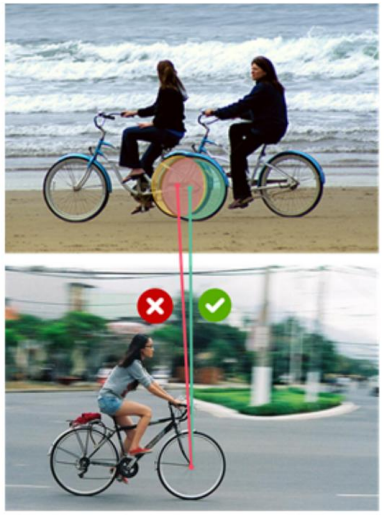
- Statistics and Examples:







Scale Variation

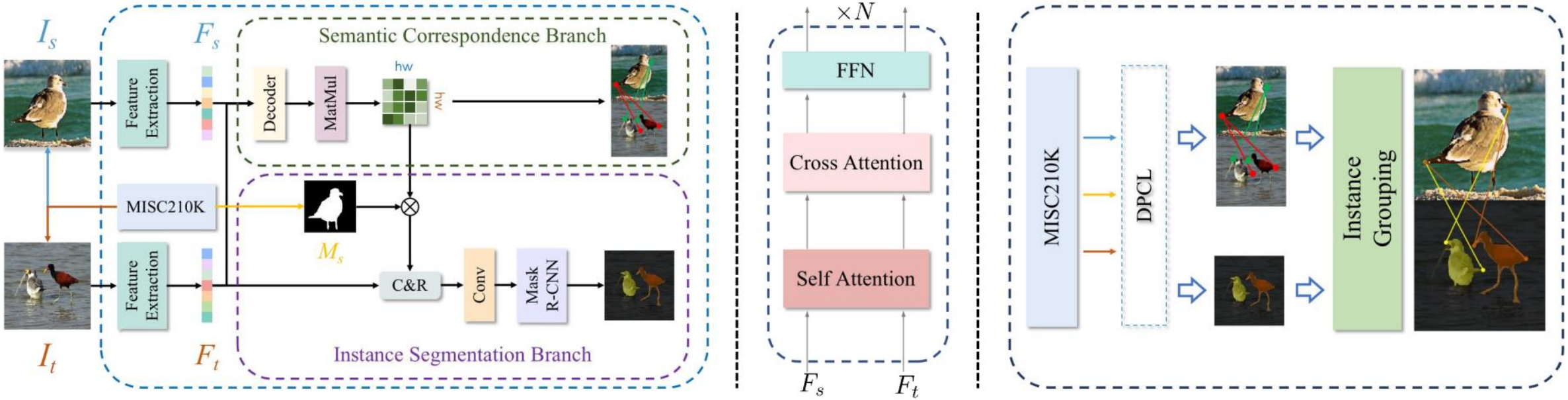


Heavy Occlusion

Shape Inconsistency

DPCL

- Dual-Path Collaborative Learning (DPCL) Pipeline:
  - Feature extraction backbone: ViT-B pre-trained with iBOT strategy.
  - Transformer decoder: 6 cascaded transformer blocks for generating 4D cost volume.
  - Semantic correspondence branch: employs a sigmoid block, non-maximum suppression, and static thresholding to obtain final predictions.
  - Instance segmentation branch: enables the grouping of same-instance matching key-points.



(a) Dual-Path Collaborative Learning Pipeline (b) Decoder (c) MISC Inference Pipeline

Experimental Results

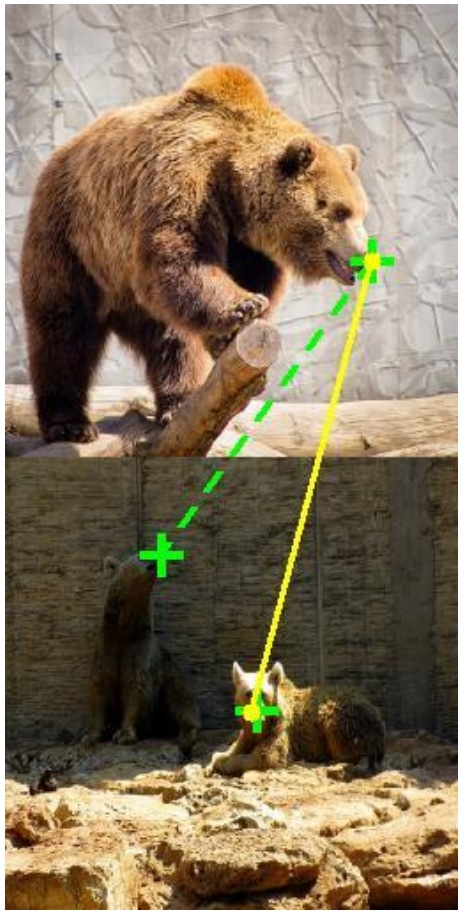
- Baseline evaluation:

Method	α	airplane	base-ball bat	bear	bed	bench	bicycle	bird	boat	bottle	bus	car	cat	chair	clock	cow	cup	dog	all
MMNet [49]	0.05	5.26	5.57	6.38	2.12	5.70	6.72	7.47	6.56	6.30	3.06	1.26	6.86	4.09	11.53	5.68	5.04	6.61	5.68
	0.10	20.38	21.81	25.95	9.86	18.03	26.80	27.98	23.64	20.50	11.85	5.96	24.03	15.10	39.78	23.92	18.79	26.70	21.58
	0.15	37.38	40.60	48.14	20.79	32.30	47.89	51.37	41.46	34.26	25.70	13.26	43.39	28.74	61.37	45.92	35.50	49.48	39.66
	1.00	99.76	99.94	99.98	99.39	98.63	99.71	99.92	100.00	99.99	99.73	99.89	99.57	99.91	99.96	99.85	99.47	99.69	99.76
CATs [4]	0.05	10.95	4.68	10.50	4.64	3.95	9.18	10.76	8.58	5.43	11.12	8.27	10.41	4.53	15.88	12.27	4.86	8.62	10.00
	0.10	25.55	14.47	27.09	13.02	12.40	23.68	24.11	20.86	14.55	27.13	19.87	24.44	13.28	33.29	27.83	15.07	22.79	23.88
	0.15	38.27	25.04	40.56	22.18	21.16	36.94	36.63	31.64	23.27	39.69	27.50	36.46	23.88	46.00	39.05	23.50	34.93	35.45
	1.00	88.02	91.85	88.41	84.60	83.53	88.90	89.56	88.96	91.70	90.74	83.21	89.17	87.97	91.41	90.80	87.53	88.60	89.15
DPCL	0.05	9.81	9.23	17.43	2.06	6.03	14.74	22.01	11.98	12.24	5.08	6.63	15.64	4.30	17.91	16.57	8.66	13.14	11.32
	0.10	22.96	23.50	36.83	7.37	16.24	32.90	43.04	27.93	23.47	14.37	15.93	38.96	13.62	37.37	35.74	20.24	31.45	25.21
	0.15	35.97	35.17	51.52	16.60	23.94	47.66	54.38	40.76	33.46	26.75	25.34	52.18	24.19	48.47	49.03	34.22	44.78	37.01
	1.00	93.90	94.04	97.29	95.25	94.67	92.79	96.77	96.29	96.23	96.32	92.06	95.37	94.17	97.27	94.65	95.46	93.60	95.07
	mIoU	21.39	1.74	44.36	27.81	32.69	24.92	24.59	21.37	4.42	52.27	16.17	33.03	3.94	15.57	37.50	0.93	30.17	22.80

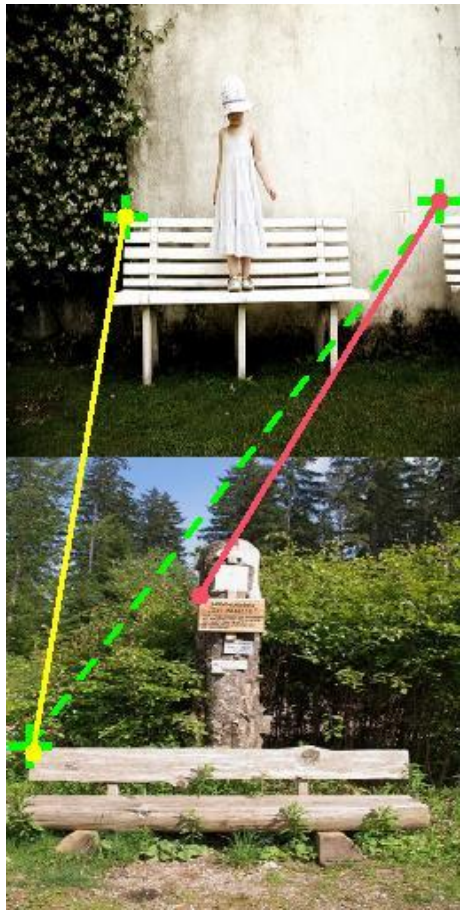
Method	α	fork	giraffe	horse	laptop	motor-cycle	mouse	person	sheep	skate-board	skis	stop-sign	tennis-racket	tie	tooth-brush	train	tv	zebra	all
MMNet [49]	0.05	4.19	3.76	5.39	5.20	5.22	4.26	3.80	7.78	5.96	5.54	7.21	7.21	5.68	4.85	7.59	4.64	4.35	5.68
	0.10	13.63	16.24	20.18	19.33	21.38	16.91	16.11	30.81	22.68	23.15	25.53	28.96	24.08	18.50	26.54	17.16	17.11	21.58
	0.15	24.93	32.49	38.71	34.95	42.19	33.02	32.80	55.63	41.13	44.39	44.90	51.55	47.83	35.18	46.41	31.58	34.30	39.66
	1.00	99.71	99.82	99.97	99.87	98.34	99.78	99.99	99.73	99.81	99.89	99.44	99.91	100.00	99.86	99.77	99.96	99.72	99.76
CATs [4]	0.05	5.34	15.85	14.82	5.34	11.43	8.19	13.22	17.82	7.22	11.47	18.19	8.56	15.73	4.94	11.66	10.42	14.93	10.00
	0.10	13.81	34.07	35.04	15.94	29.47	17.91	28.90	38.07	19.67	29.50	34.89	20.57	34.57	12.58	28.52	24.37	34.35	23.88
	0.15	21.97	47.42	49.38	27.41	44.62	26.78	41.59	51.80	31.79	42.09	46.48	32.19	47.00	19.92	41.49	36.20	48.57	35.45
	1.00	88.63	92.24	91.64	87.54	90.34	82.90	89.58	92.71	87.76	89.40	86.95	92.45	89.30	87.22	88.21	91.96	91.76	89.15
DPCL	0.05	11.66	14.77	11.68	4.18	11.41	8.74	8.99	21.38	10.24	6.15	6.93	15.13	17.30	16.26	10.10	2.48	13.46	11.32
	0.10	21.93	31.38	25.94	13.92	29.86	15.99	23.73	42.47	25.12	16.14	15.39	34.37	27.93	27.82	28.21	8.19	29.22	25.21
	0.15	28.92	41.82	40.18	22.90	45.19	27.00	38.62	54.26	37.44	24.24	26.69	48.97	38.57	35.24	41.54	20.10	43.36	37.01
	1.00	96.29	94.66	95.18	94.40	91.28	93.06	95.77	96.21	94.82	94.87	93.02	96.97	94.53	96.15	93.58	95.61	96.62	95.07
	mIoU	0.14	29.03	38.05	38.07	41.13	4.92	11.11	33.48	6.81	0.00	41.77	6.06	2.25	10.60	49.22	29.13	40.48	22.80

Experimental Results

- Challenges and Visualizations:



Missing Keypoint

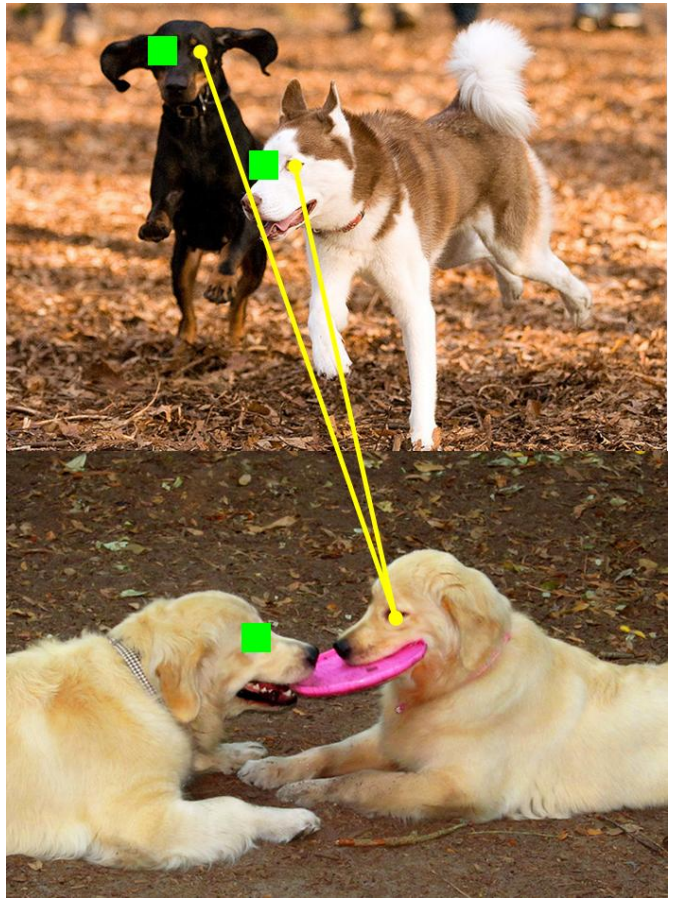


Inconsistent Prediction

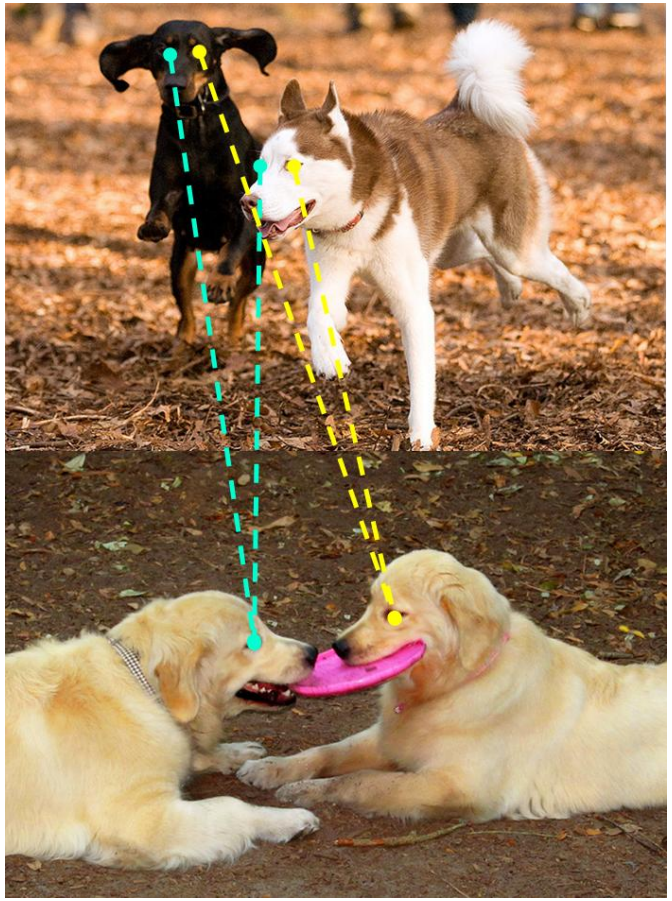
Perspective Distortion

Future Direction

- Unseen Key-point Discovery

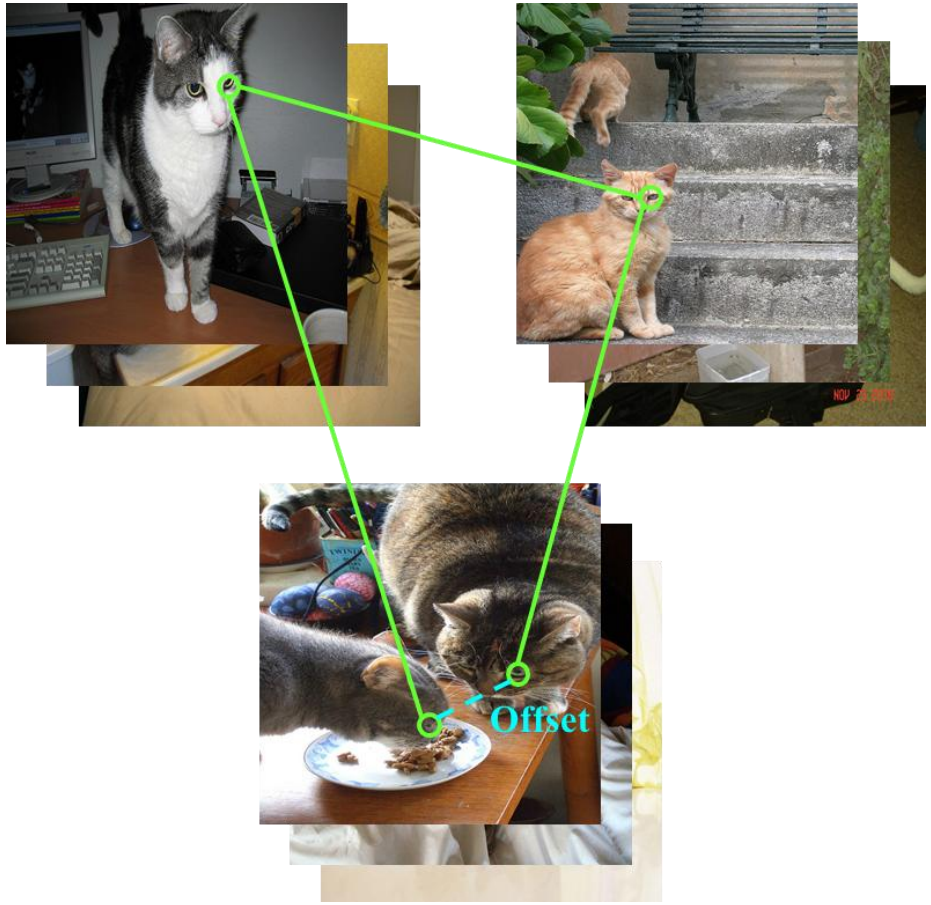


Training

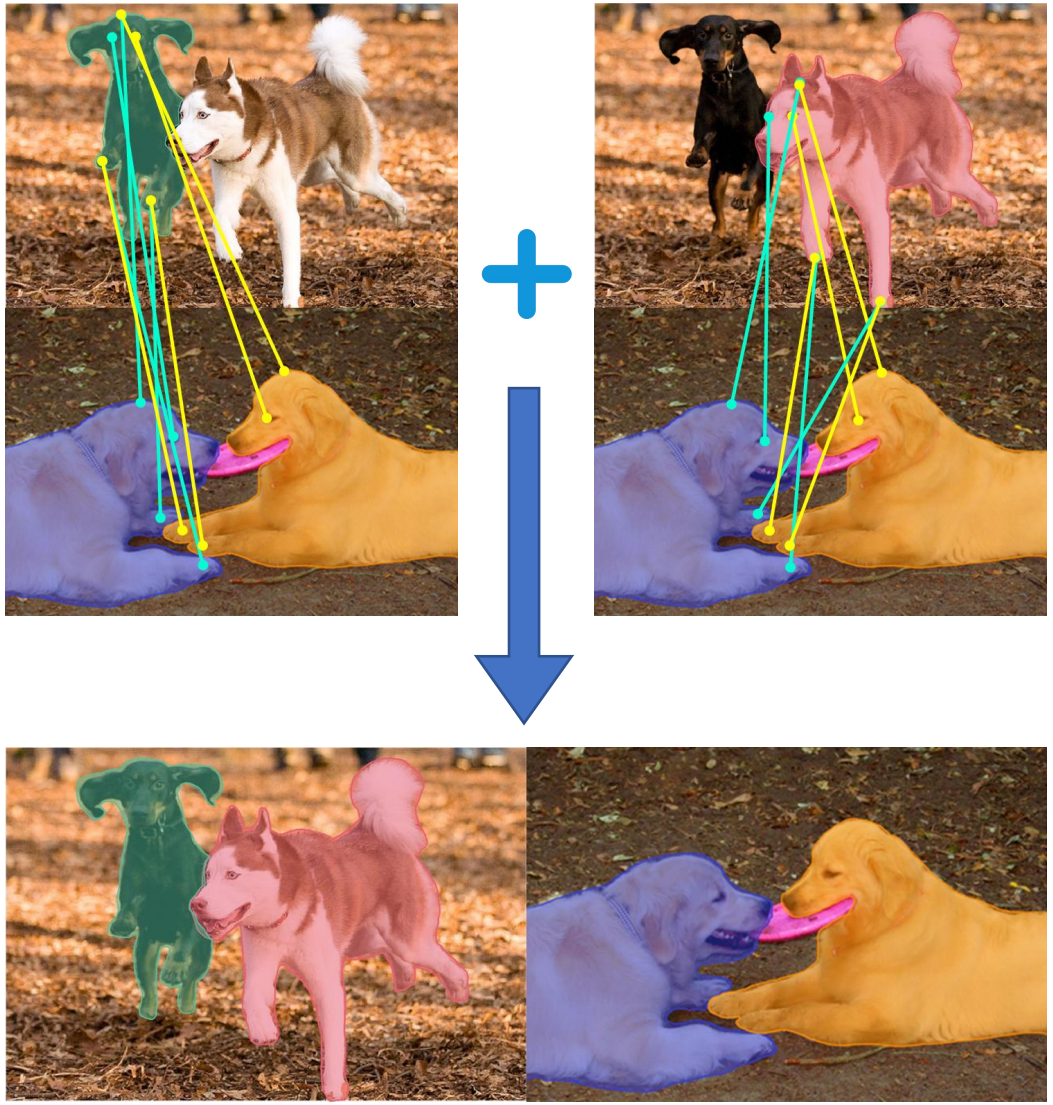


Validation/Testing

Future Direction



Matching Closed-loop Constraint



Correspondence based Recognition Tasks

Thank You

Academy of Engineering & Technology, Fudan University, Shanghai, China
School of Computer Science, Fudan University, Shanghai, China
The University of Hong Kong, Hong Kong, China
{wfge, wqzhang}@fudan.edu.cn