Matching Is Not Enough: A Two-Stage Framework for Category-Agnostic Pose Estimation

Min Shi[1, 2, *] Zihao Huang[1, *] Xianzheng Ma[2] Xiaowei Hu[2] Zhiguo Cao[ǂ]

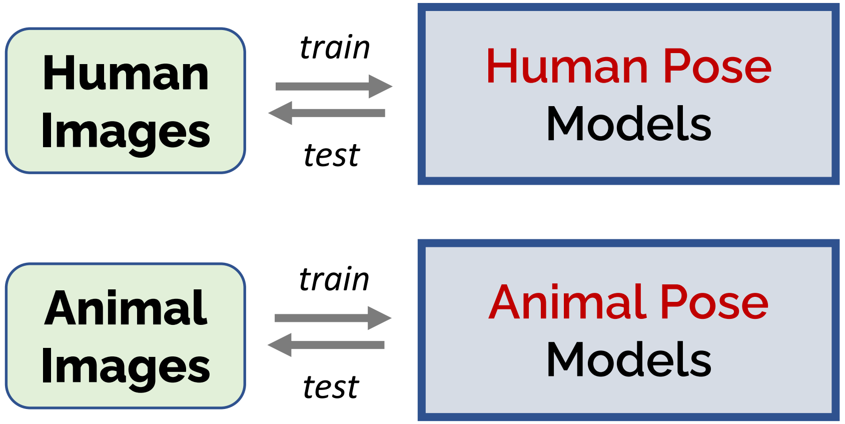
[1] Huazhong University of Science and Technology, China

[2] Shanghai AI Laboratory, China

*: Equal contribution. ǂ : Corresponding author.

Category-Agnostic Pose Estimation

Class-specific Pose Estimation



Human Images

train

test

Human Pose Models

Animal Images

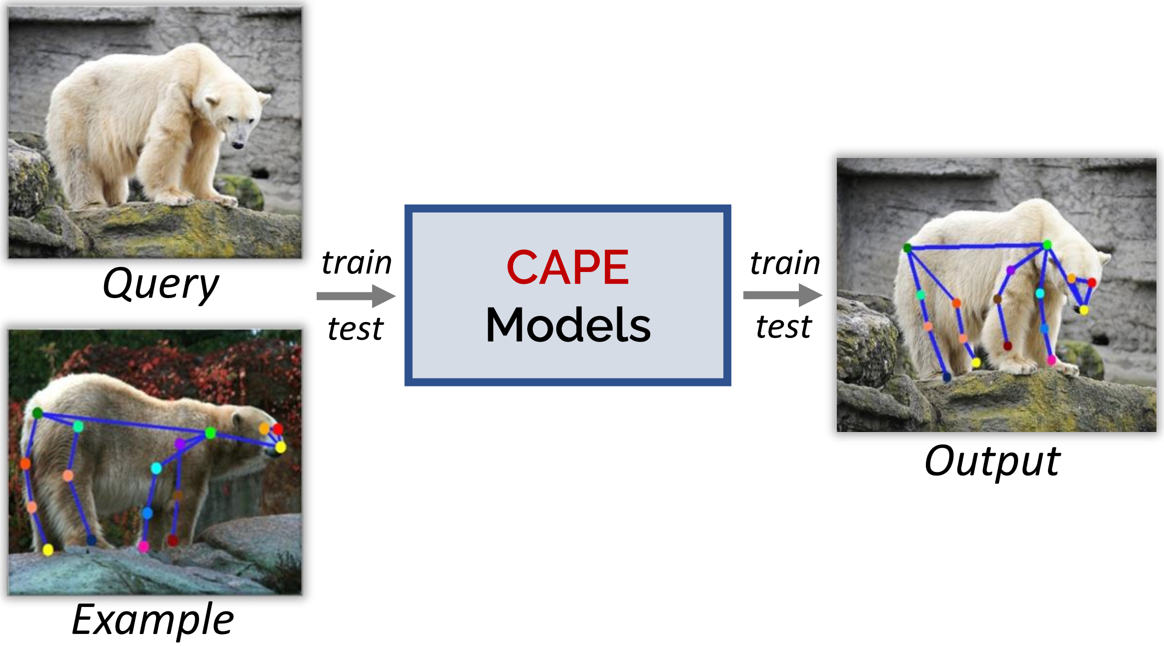
train

test

Animal Pose Models

Learning to localize the keypoints for specific categories

Class-agnostic Pose Estimation (CAPE)



Query

Example

train

test

CAPE Models

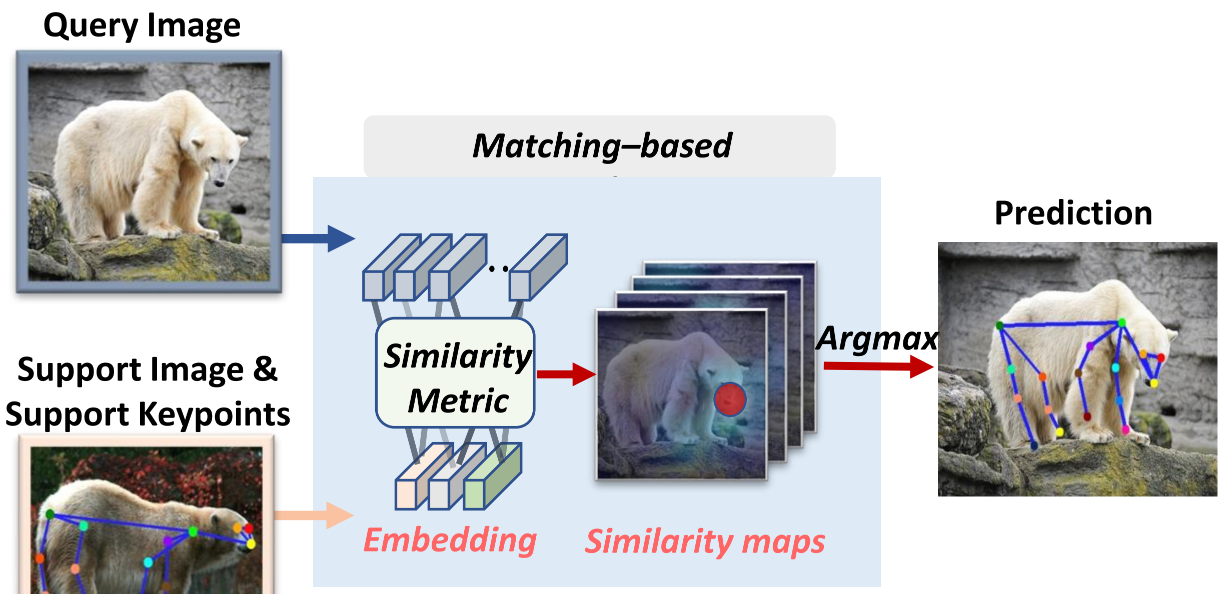
train

test

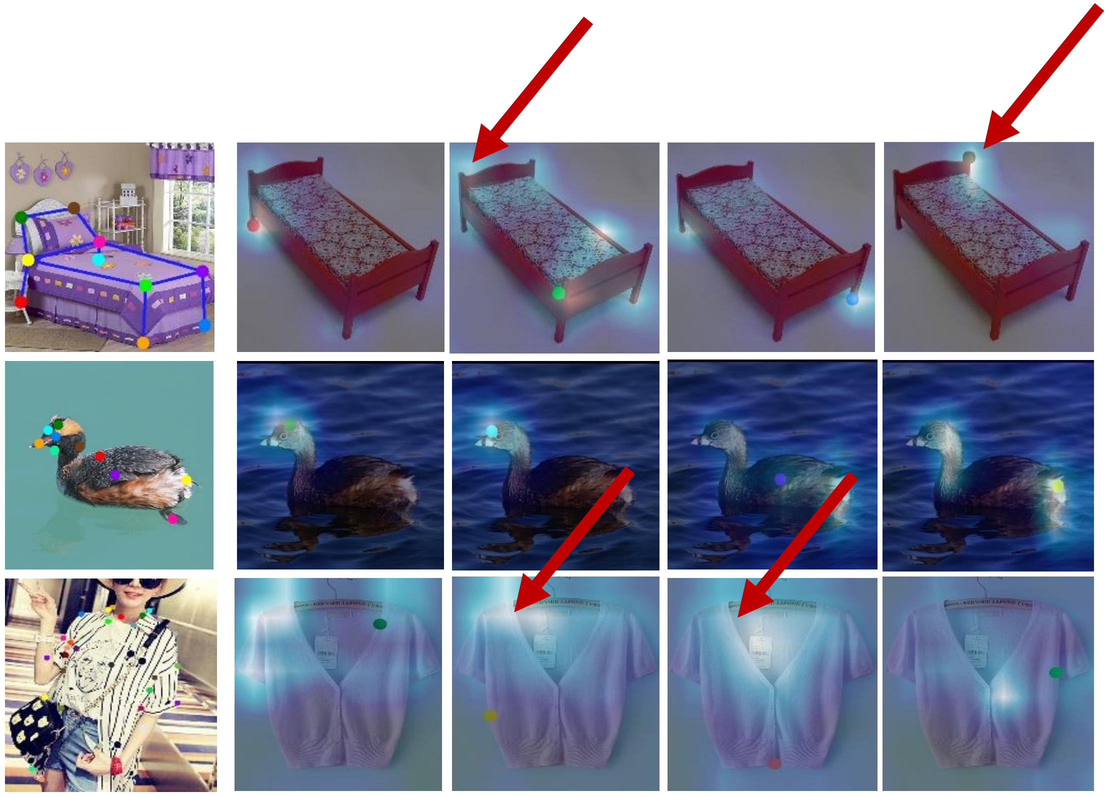
Output

Learning the way to localize keypoints given one or a few examples

Motivation

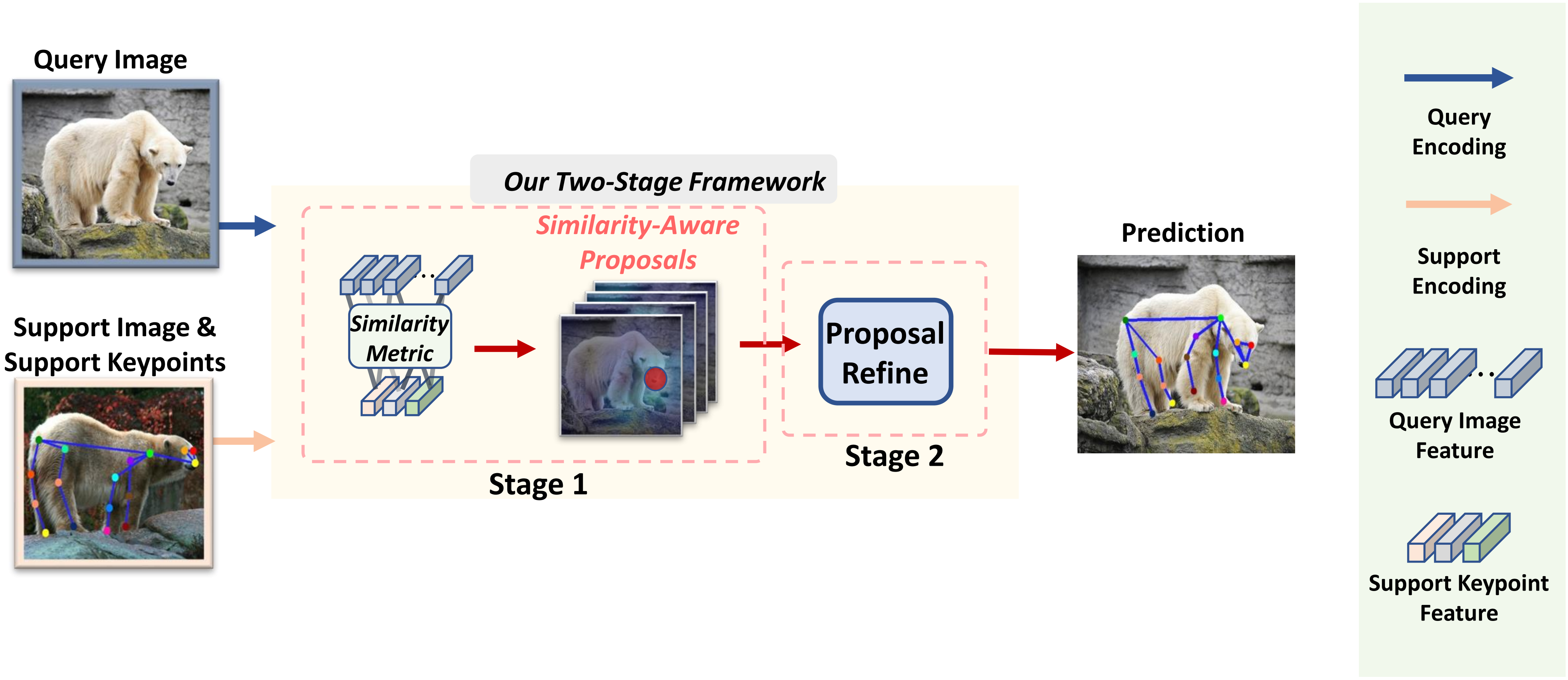


Previous Approaches

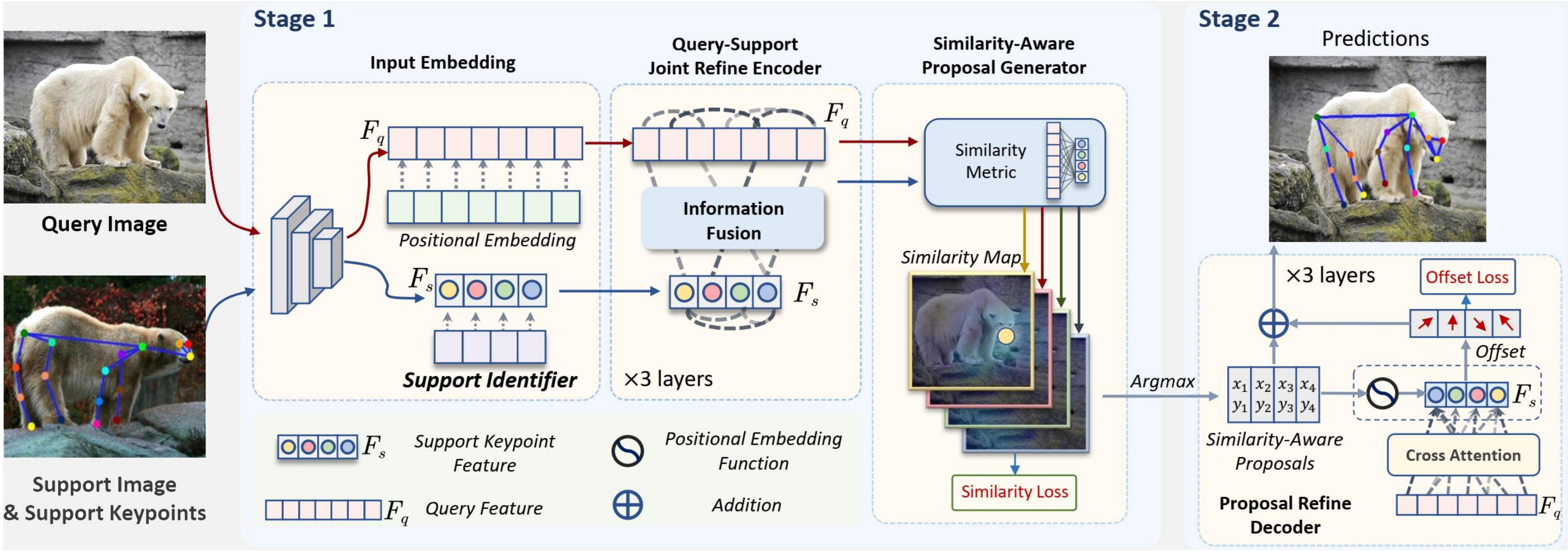


Noisy Similarity Map

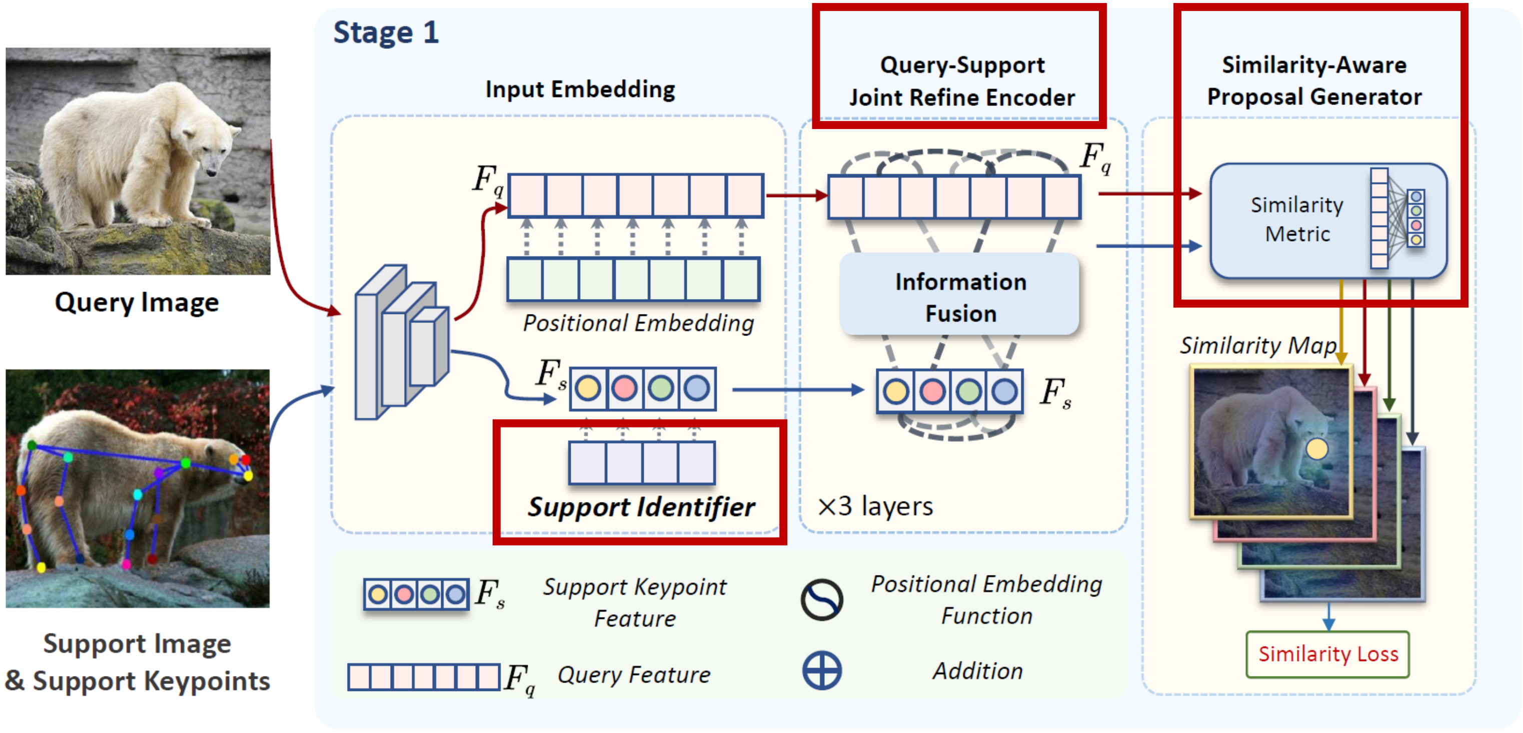
A Two-stage Framework for CAPE



CAPE Transformer (CapeFormer)



First Stage: Matching



- ➢ **Query-Support Refine Encoder**

  Transfer features among support keypoints and the query images.

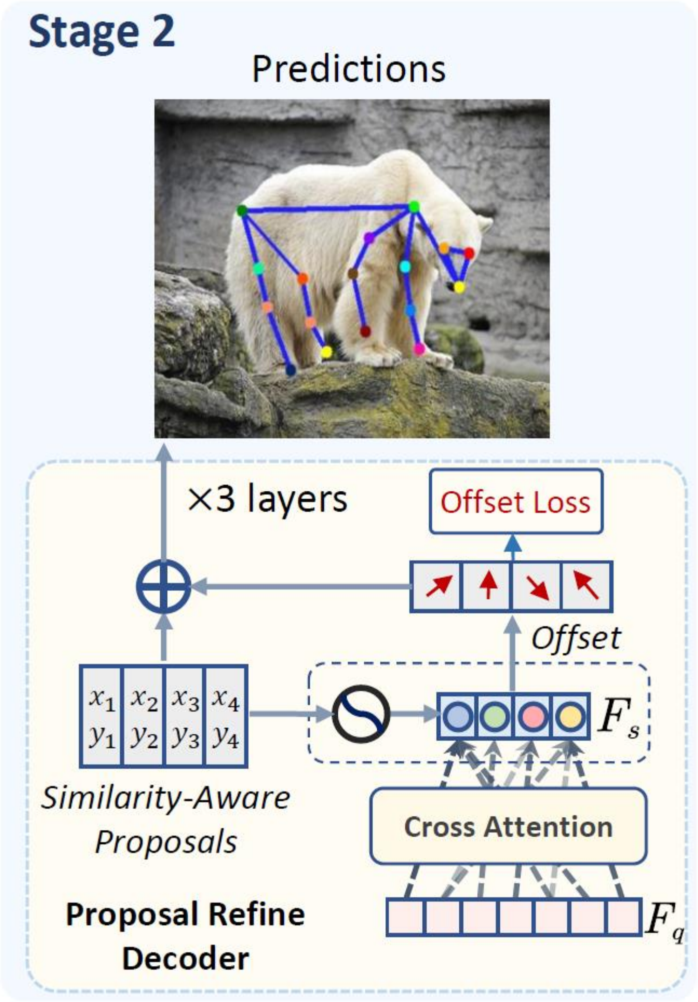
- ➢ **Support Identifier**

  Encode positional and context information of each support keypoint.

- ➢ **Similarity-Aware Proposal Generator**

  Generate position proposals from similarity maps via inner product.

Second Stage: Proposal Refine



- **Self-attention among support keypoints**

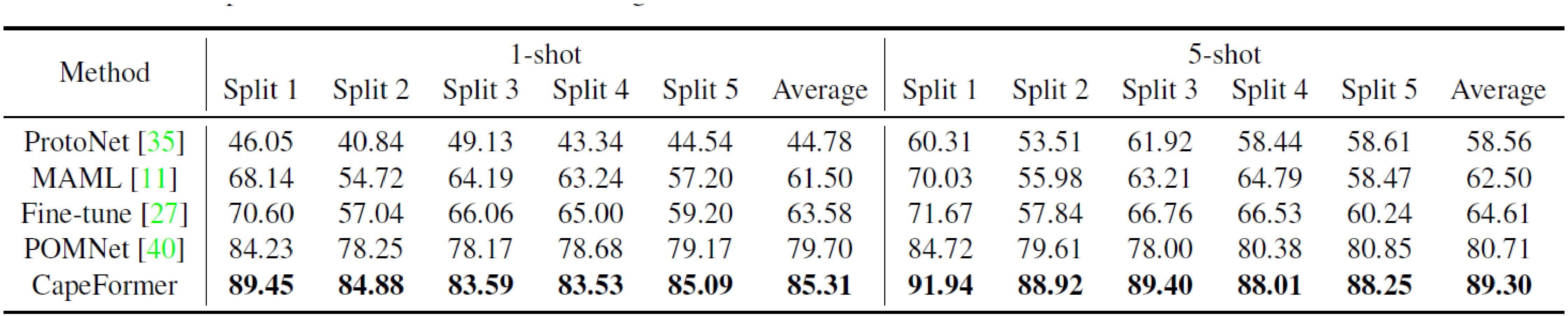
  Make each proposal aware of other keypoints' positions and contents.

- **Cross-attention between support and query features**

  Extract relevant contents from the query feature for each support keypoint.

Experiments

Quantitative comparisons on MP-100 Dataset



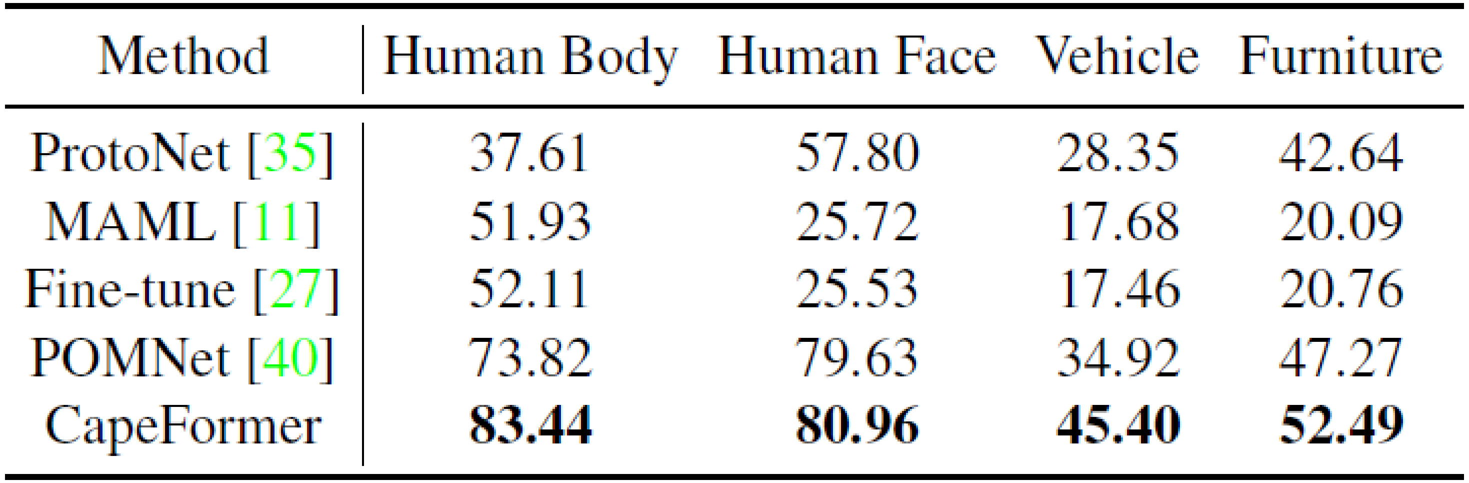
Method	1-shot						5-shot					
	Split 1	Split 2	Split 3	Split 4	Split 5	Average	Split 1	Split 2	Split 3	Split 4	Split 5	Average
ProtoNet [35]	46.05	40.84	49.13	43.34	44.54	44.78	60.31	53.51	61.92	58.44	58.61	58.56
MAML [11]	68.14	54.72	64.19	63.24	57.20	61.50	70.03	55.98	63.21	64.79	58.47	62.50
Fine-tune [27]	70.60	57.04	66.06	65.00	59.20	63.58	71.67	57.84	66.76	66.53	60.24	64.61
POMNet [40]	84.23	78.25	78.17	78.68	79.17	79.70	84.72	79.61	78.00	80.38	80.85	80.71
CapeFormer	**89.45**	**84.88**	**83.59**	**83.53**	**85.09**	**85.31**	**91.94**	**88.92**	**89.40**	**88.01**	**88.25**	**89.30**

CapeFormer *vs.* POMNet

+7.04 % on 1-shot Average PCK

+10.64 % on 5-shot Average PCK

Cross Super-category Experiments



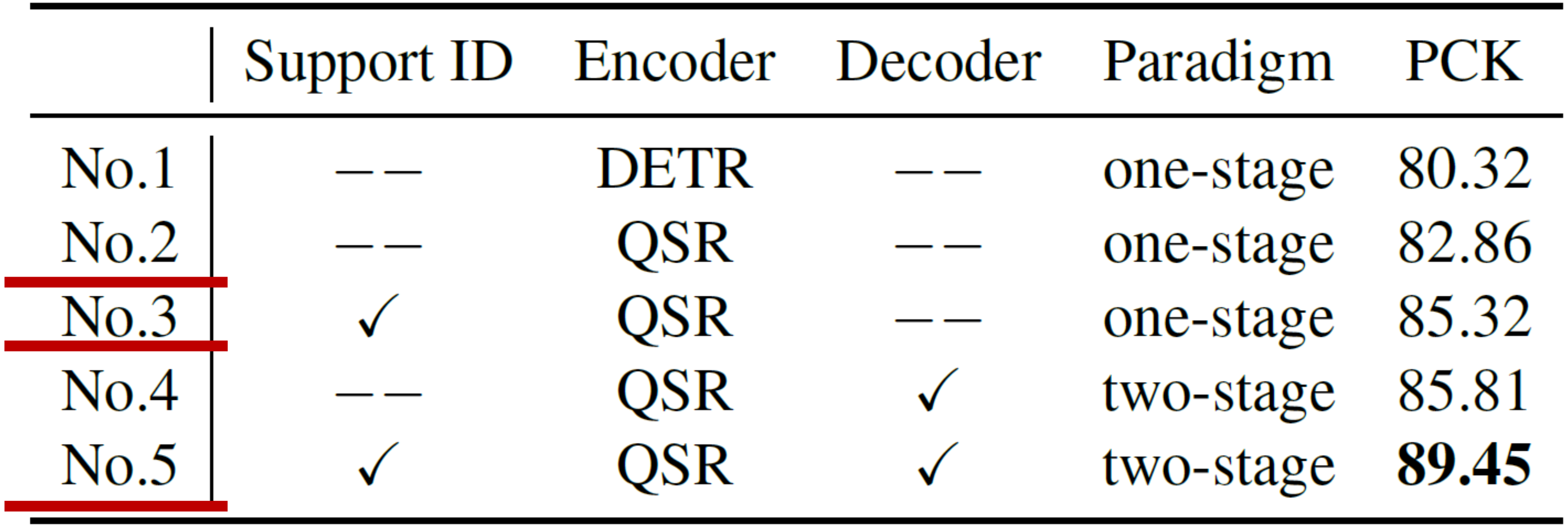
Method	Human Body	Human Face	Vehicle	Furniture
ProtoNet [35]	37.61	57.80	28.35	42.64
MAML [11]	51.93	25.72	17.68	20.09
Fine-tune [27]	52.11	25.53	17.46	20.76
POMNet [40]	73.82	79.63	34.92	47.27
CapeFormer	**83.44**	**80.96**	**45.40**	**52.49**

- **cross-super category**
  - Leave-One Out Strategy
  - Each super-category are treated as the test set, while the others are used as training samples.

CapeFormer *vs.* POMNet

+13.03 % on PCK of *Human Body*
+1.67 % on PCK of *Human Face*
+30.01 % on PCK of *Vehicle*
+11.04 % on PCK of *Furniture*

Ablation Study

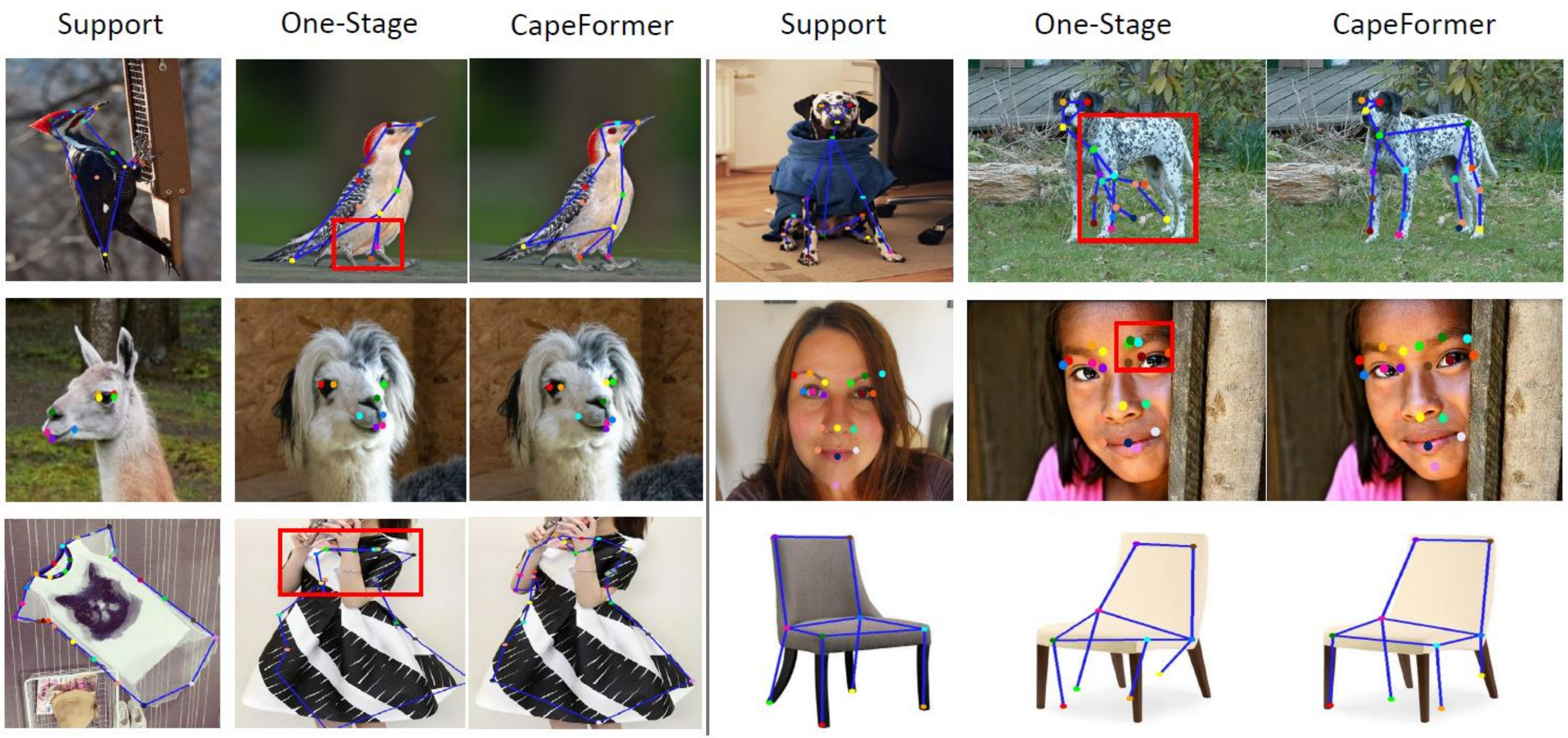


	Support ID	Encoder	Decoder	Paradigm	PCK
No.1	−−	DETR	−−	one-stage	80.32
No.2	−−	QSR	−−	one-stage	82.86
No.3	✓	QSR	−−	one-stage	85.32
No.4	−−	QSR	✓	two-stage	85.81
No.5	✓	QSR	✓	two-stage	**89.45**

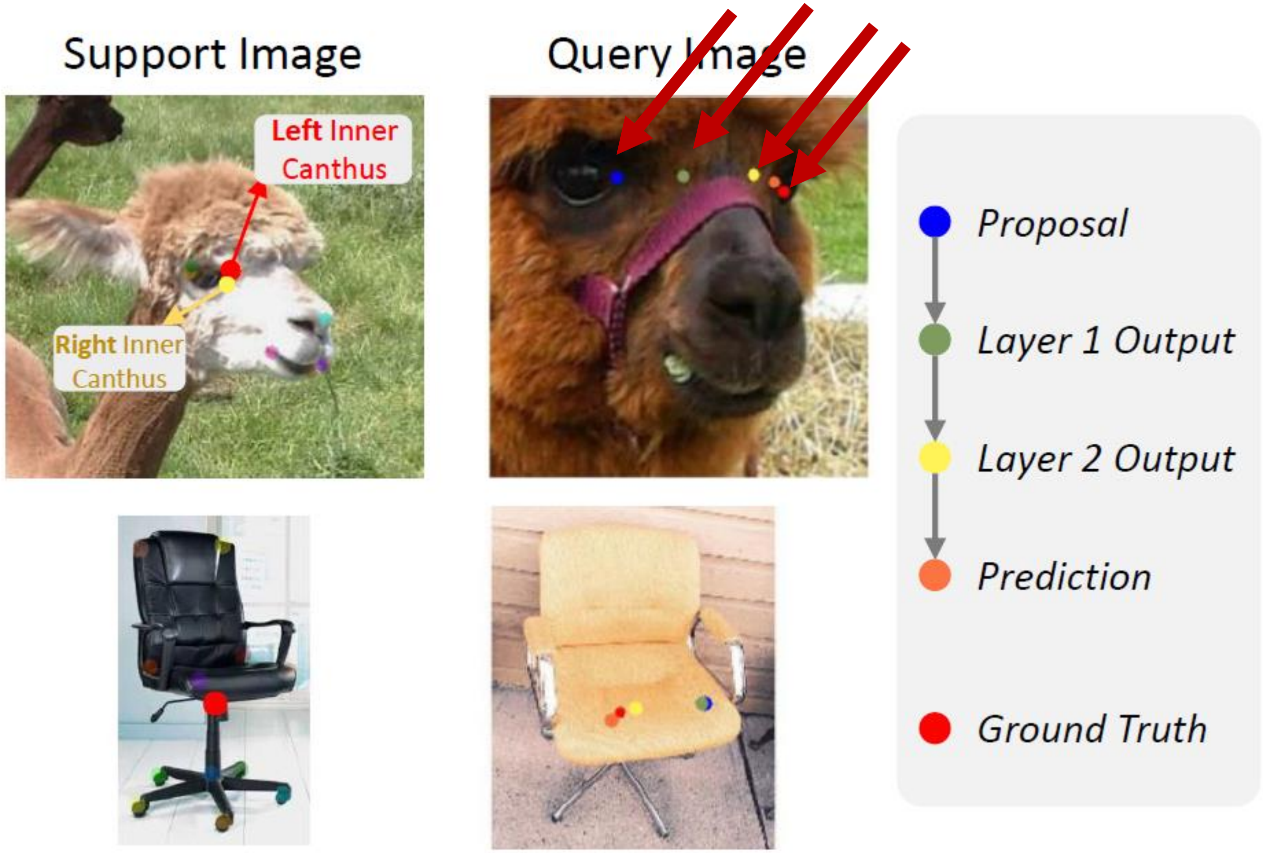
Support ID: Similarity Loss

QSR: query-support joint refine encoder

Visualizations



Proposal Correction Process



Thanks for watching!

- There are still many challenges: data shortage, multi-scale modeling, feature fusion … We look forward to more attention on this topic.
- Full training and test code has been released at tiny.one/BMNet.
- Feel free to contact us by min_shi@hust.edu.cn.