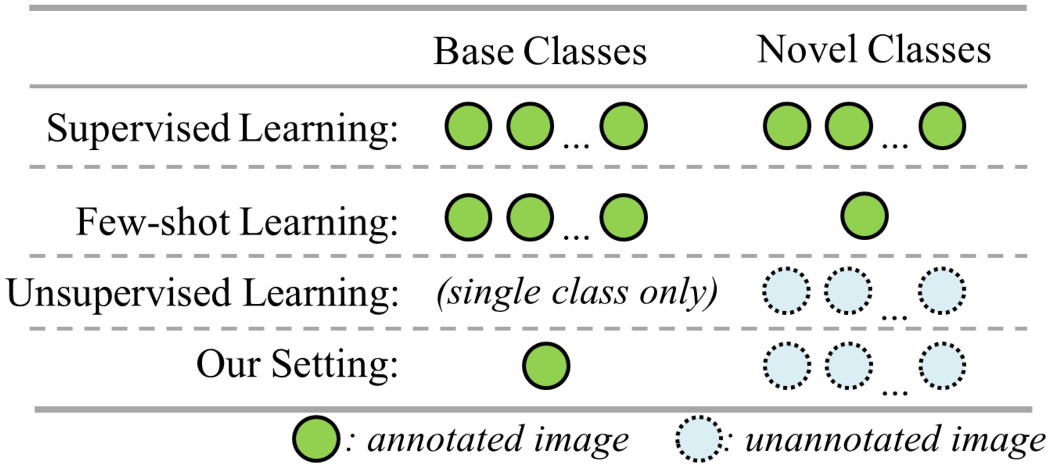
Weak-Shot Keypoint Estimation via Keyness and Correspondence Transfer

Junjie Chen[1], Zeyu Luo[1], Zezheng Liu[1], Wenhui Jiang[1],
Li Niu[2*], Yuming Fang[1]
[1]Jiangxi University of Finance and Economics,
[2]Shanghai Jiao Tong University

Background

Few-shot and unsupervised keypoint estimation are prevalent economical paradigms, but the former still requires annotations for extensive novel classes while the latter only supports for single class.

- we focus on the task of weak-shot keypoint estimation, where multiple novel classes are learned from unlabeled images with the help of labeled base classes.
- Data comparison among fully supervised learning, few-shot learning, unsupervised learning and our weak-shot learning. Our setting is more economical and practicable.



	Base Classes	Novel Classes
Supervised Learning:	● ● ... ●	● ● ... ●
Few-shot Learning:	● ● ... ●	●
Unsupervised Learning:	*(single class only)*	◌ ◌ ... ◌
Our Setting:	●	◌ ◌ ... ◌

●*: annotated image* ◌*: unannotated image*

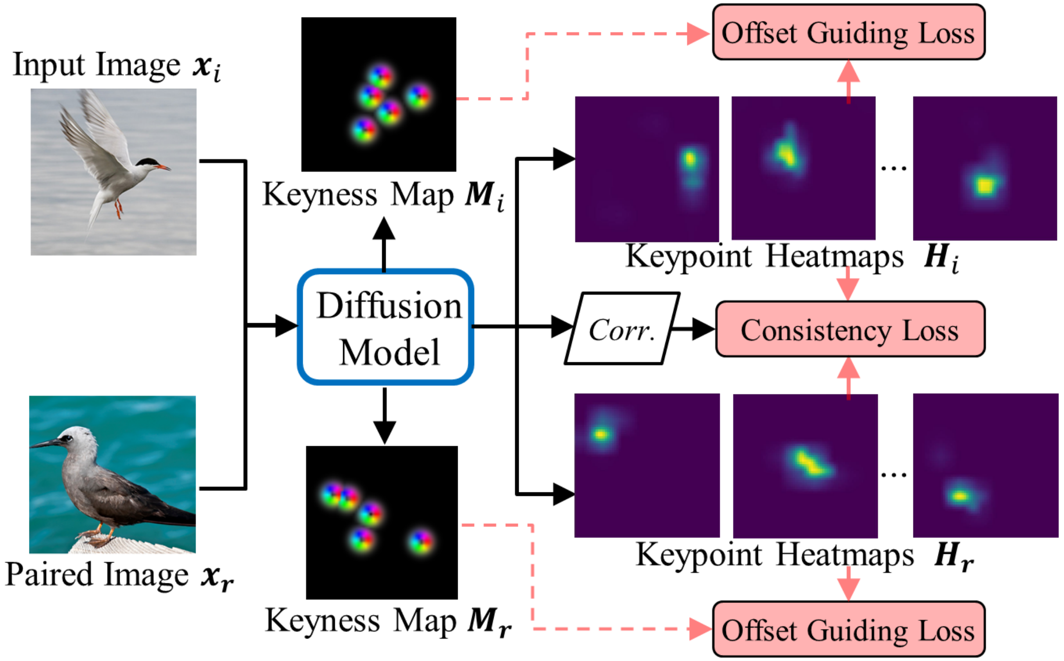
Comparison of various task settings.

Background

The key problem is what to transfer from base classes to novel classes,and we propose to transfer keyness and correspondence, which essentially belong to comparing entities and thus are class-agnostic and class-wise transferable.

▸The keyness compares which pixel in the local region is more key, which can guide the keypoints of novel classes to move towards the local maximum.

▸The correspondence compares whether the two pixels belongs to the same semantic part, which can activate the keypoints of novel classes by reinforcing the consistency between two paired images.


Input Image $\boldsymbol{x}_i$
Paired Image $\boldsymbol{x}_r$
Keyness Map $\boldsymbol{M}_i$
Diffusion Model
Keyness Map $\boldsymbol{M}_r$
Offset Guiding Loss
Keypoint Heatmaps $\boldsymbol{H}_i$
...
Corr.
Consistency Loss
Keypoint Heatmaps $\boldsymbol{H}_r$
Offset Guiding Loss

Overview of our framework.

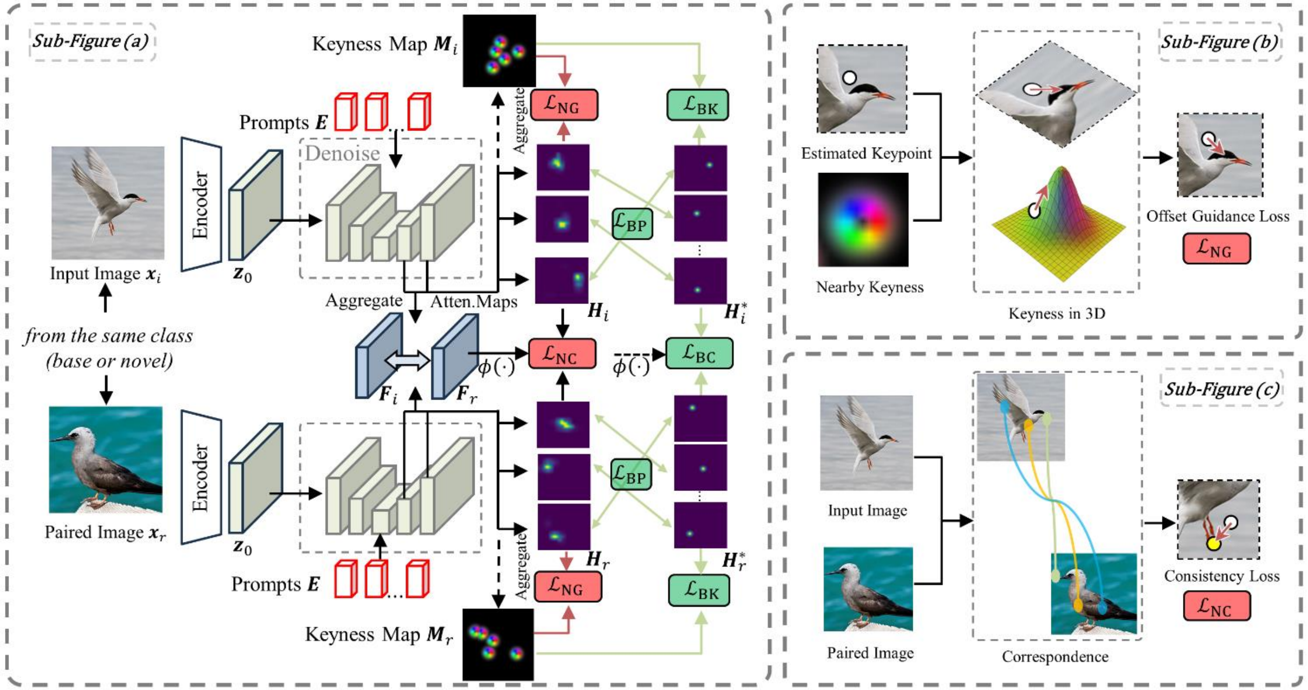
Contributions

- We are the first to explore weak-shot keypoint estimation, where we could use the knowledge transferred from base classes to facilitate the unsupervised learning of multiple novel classes.

- We propose a well-tailored framework to learn keyness and correspondence from base classes, and transfer them to provide effective supervision for the unsupervised learning of novel classes.

- Extensive experiments on the large-scale dataset demonstrate the effectiveness of our method.

Figure: The detailed illustration of our framework in the training stage.



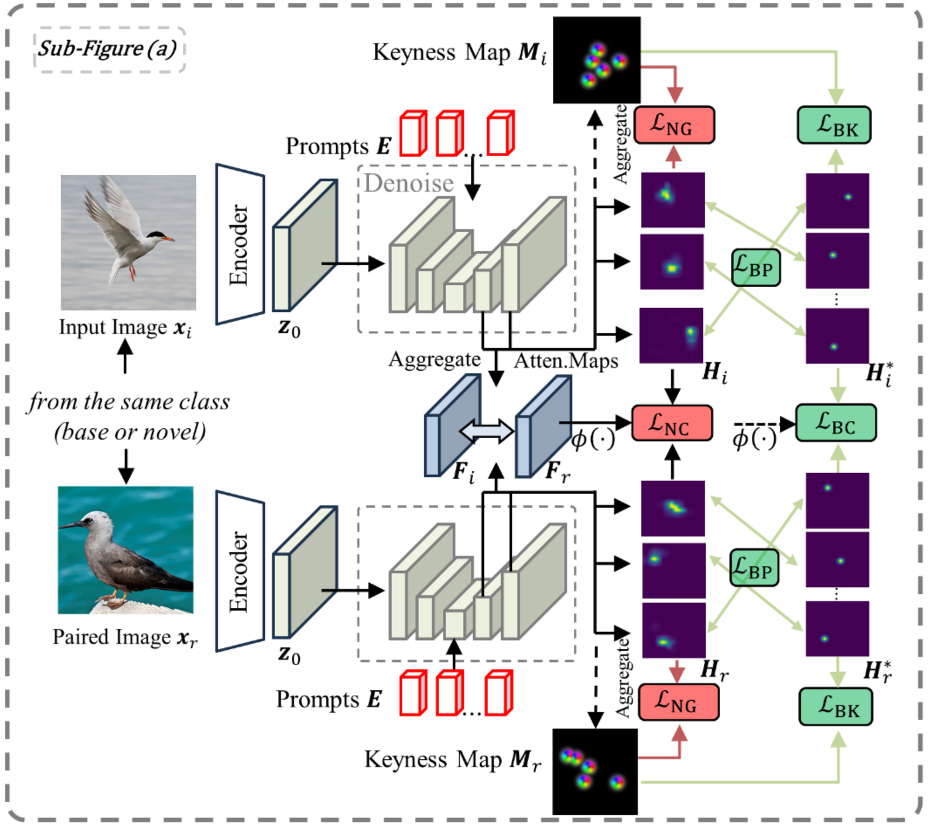
Method

The Forward Pipeline of Our Framework

Given image pairs from the same base or novel class, we employ keypoint prompts to estimate keypoint heatmaps and aggregate feature maps to estimate keyness and extract correspondences.

For labeled image pairs from base classes, we learn keypoint prompts, keyness and correspondence via $\mathcal{L}_{BP}$, $\mathcal{L}_{BK}$ and $\mathcal{L}_{BC}$.

For unlabeled image pairs from novel classes, we transfer them to learn valid keypoints via $\mathcal{L}_{NG}$, $\mathcal{L}_{NC}$ and $\mathcal{L}_{NU}$.



Sub-Figure (a)
Keyness Map $\boldsymbol{M}_i$
$\mathcal{L}_{\mathrm{NG}}$
$\mathcal{L}_{\mathrm{BK}}$
Aggregate
Prompts $\boldsymbol{E}$
...
Denoise
Encoder
Input Image $\boldsymbol{x}_i$
$\boldsymbol{z}_0$
$\mathcal{L}_{\mathrm{BP}}$
Aggregate
Atten.Maps
$\boldsymbol{H}_i$
$\boldsymbol{H}_i^*$
from the same class
(base or novel)
$\boldsymbol{F}_i$
$\boldsymbol{F}_r$
$\phi(\cdot)$
$\mathcal{L}_{\mathrm{NC}}$
$\phi(\cdot)$
$\mathcal{L}_{\mathrm{BC}}$
Encoder
Paired Image $\boldsymbol{x}_r$
$\boldsymbol{z}_0$
$\mathcal{L}_{\mathrm{BP}}$
$\boldsymbol{H}_r$
$\boldsymbol{H}_r^*$
Aggregate
Prompts $\boldsymbol{E}$
...
$\mathcal{L}_{\mathrm{NG}}$
$\mathcal{L}_{\mathrm{BK}}$
Keyness Map $\boldsymbol{M}_r$

Method

Learning from Base Classes

- To learn keypoint prompts from base classes, we apply following loss over the keypoint heatmaps.

$$\mathcal{L}_{BP} = \frac{1}{K}\sum_{k}^{K}\|H_i[\delta(k)] - H_i^*[k]\|^2 + \|H_r[\delta(k)] - H_r^*[k]\|^2 \quad (1)$$

- To learn valid keyness, we apply the loss analogous to Eqn. 1 due to the same “map” representation.

$$\mathcal{L}_{BK} = \|M_i - M_i^*\|^2 + \|M_r - M_r^*\|^2 \qquad (2)$$

- To learn valid correspondences, we compute cosine similarity between every possible pair of points and apply a symmetric cross entropy loss according to GT keypoints, which is denoted as $\mathcal{L}_{BC}$

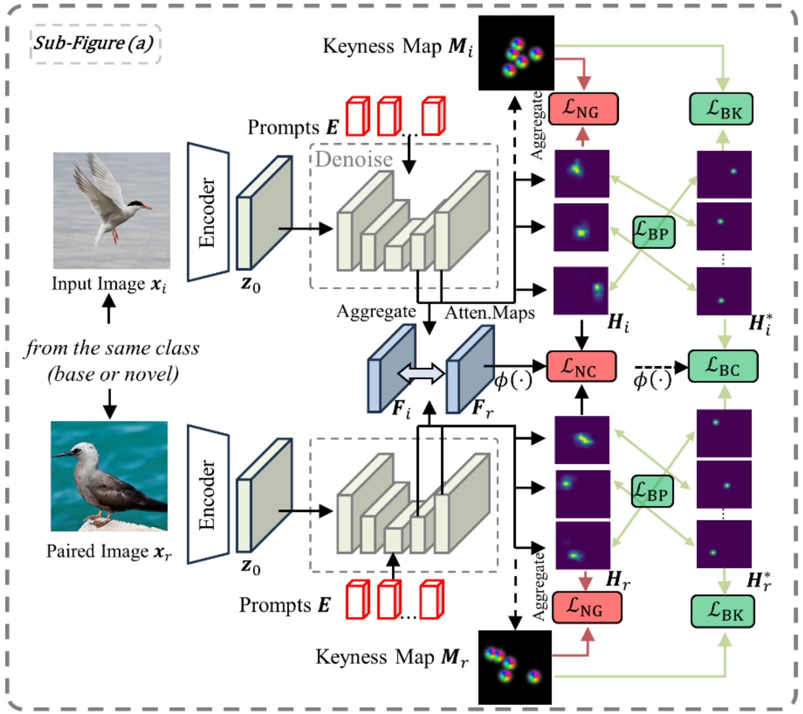


Figure: The illustration of our framework in the training stage.

Method

Transferring to Novel Classes

To bridge the domain gap, we propose to apply offset guiding loss and consistency loss according to the transferred keyness and correspondences.

- we firstly convert the heatmaps H_i and H_r to the keypoint coordinates and visibilities $[P_i; V_i]$ and $[P_i; V_i] \in \mathbb{R}^{N\times(2+1)}$ via the soft argmax and max operators.

$$\mathcal{L}_{NG} = \frac{1}{N}\sum_k^N \left\|P_i[k] - dt\left(P_i[k] + M_i\left[P_i[k]\right]\right)\right\|^2 + \left\|P_r[k] - dt\left(P_r[k] + M_r\left[P_r[k]\right]\right)\right\|^2 \quad (3)$$

- Then, we use below consistency loss to facilitate the keypoint learning of novel classes.

$$\mathcal{L}_{NC} = \sum_n^N V_r[n] \cdot \left\|H_i[n] - \mathcal{H}\left(\tilde{P}_i[n]\right)\right\|^2 + V_i[n] \cdot \left\|H_r[n] - \mathcal{H}\left(\tilde{P}_r[n]\right)\right\|^2 \quad (4)$$

- we apply the equivariance and localization loss to complement our unsupervised learning on novel classes, which is denoted as $\mathcal{L}_{NU}$

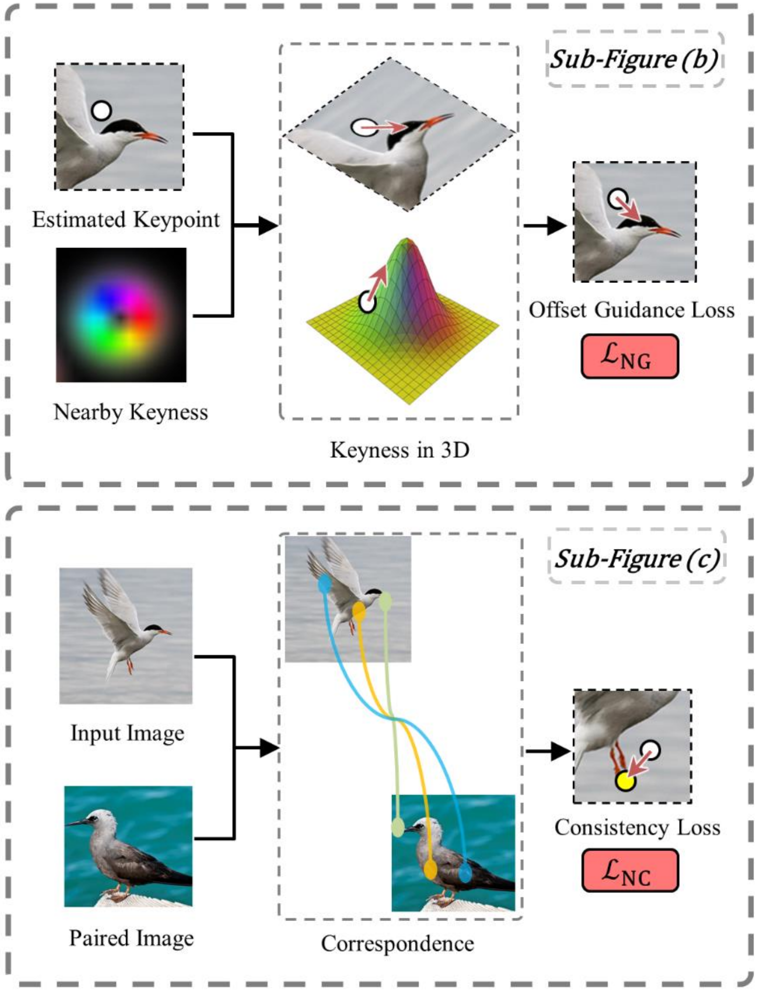


Figure: The insight of $\mathcal{L}_{NG}$ and $\mathcal{L}_{NC}$.

Method

Summary of Training and Inference

- Our framework is end-to-end learnable, which can jointly learn keypoint prompts, keyness and correspondence from base classes and transfer them to novel classes. In the training stage, our full loss for the i-th image x_i can be summarized as:

$$\mathcal{L}_{FULL} = \mathbf{1}_{x_i \in \mathcal{B}} \cdot (\mathcal{L}_{BP} + \mathcal{L}_{BK} + \alpha \cdot \mathcal{L}_{BC}) + \mathbf{1}_{x_i \in \mathcal{N}} \cdot (\beta \cdot \mathcal{L}_{NG} + \gamma \cdot \mathcal{L}_{NC} + \lambda \cdot \mathcal{L}_{NU}) \quad (6)$$

- In the test stage, we remove the additional modules estimating keyness and correspondence, which are only used to provide effective supervisions for learning novel classes without annotations.

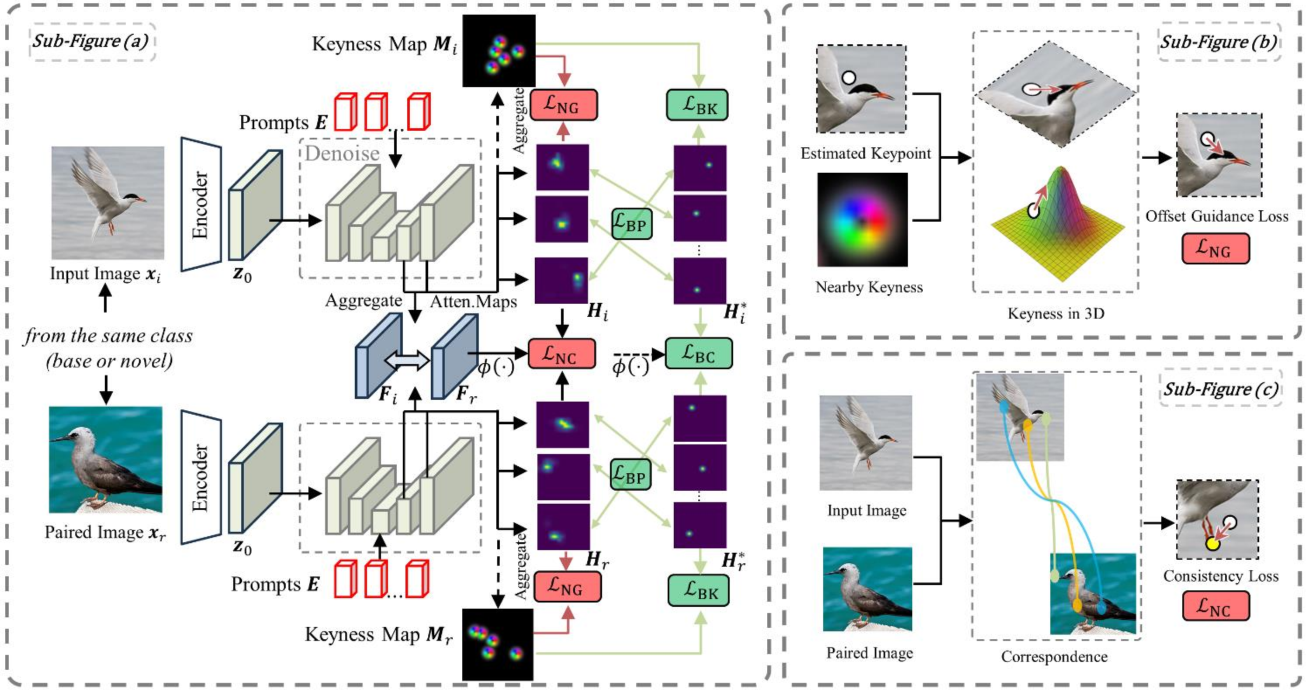


Figure: The detailed illustration of our framework in the training stage.

Experiments

Comparison with prior works. Quantitative comparison

Our proposed model has relatively robust mAPs on different splits using both evaluation protocols.

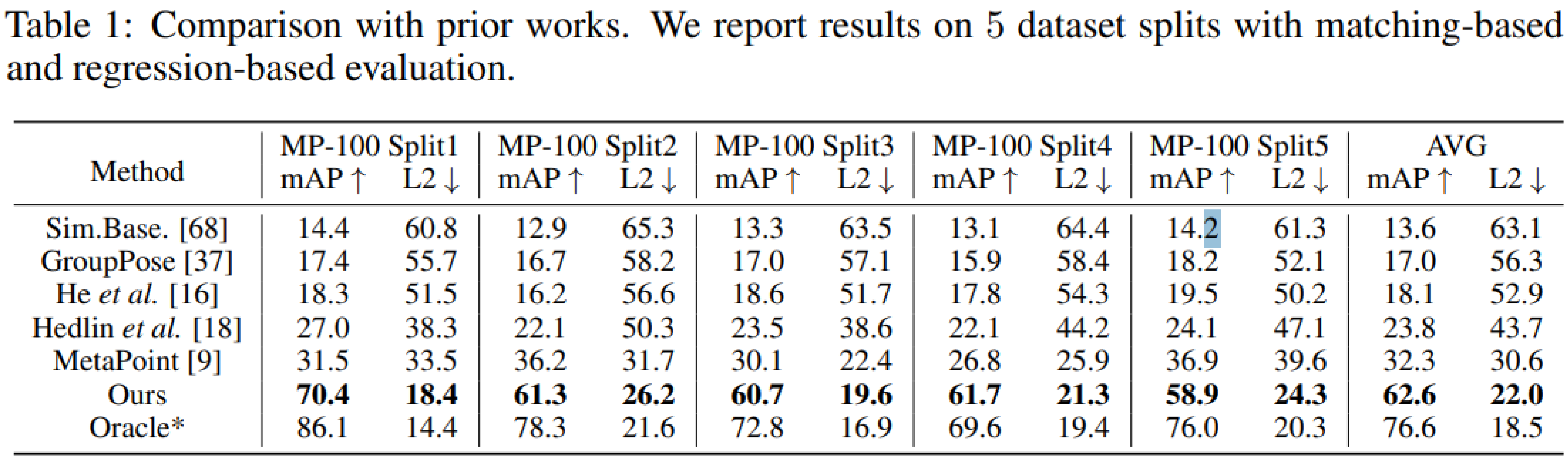


Table 1: Comparison with prior works. We report results on 5 dataset splits with matching-based and regression-based evaluation.

Method	MP-100 Split1		MP-100 Split2		MP-100 Split3		MP-100 Split4		MP-100 Split5		AVG	
	mAP ↑	L2 ↓	mAP ↑	L2 ↓	mAP ↑	L2 ↓	mAP ↑	L2 ↓	mAP ↑	L2 ↓	mAP ↑	L2 ↓
Sim.Base. [68]	14.4	60.8	12.9	65.3	13.3	63.5	13.1	64.4	14.2	61.3	13.6	63.1
GroupPose [37]	17.4	55.7	16.7	58.2	17.0	57.1	15.9	58.4	18.2	52.1	17.0	56.3
He *et al.* [16]	18.3	51.5	16.2	56.6	18.6	51.7	17.8	54.3	19.5	50.2	18.1	52.9
Hedlin *et al.* [18]	27.0	38.3	22.1	50.3	23.5	38.6	22.1	44.2	24.1	47.1	23.8	43.7
MetaPoint [9]	31.5	33.5	36.2	31.7	30.1	22.4	26.8	25.9	36.9	39.6	32.3	30.6
Ours	**70.4**	**18.4**	**61.3**	**26.2**	**60.7**	**19.6**	**61.7**	**21.3**	**58.9**	**24.3**	**62.6**	**22.0**
Oracle*	86.1	14.4	78.3	21.6	72.8	16.9	69.6	19.4	76.0	20.3	76.6	18.5

Our model can leverage pretrained diffusion model to adaptively learn and transfer keypoint prompts and correspondences, resulting in preferable performances. Besides, our model can reach about 75% of the upper-bound mAP represented by Oracle*, showing promising potential.

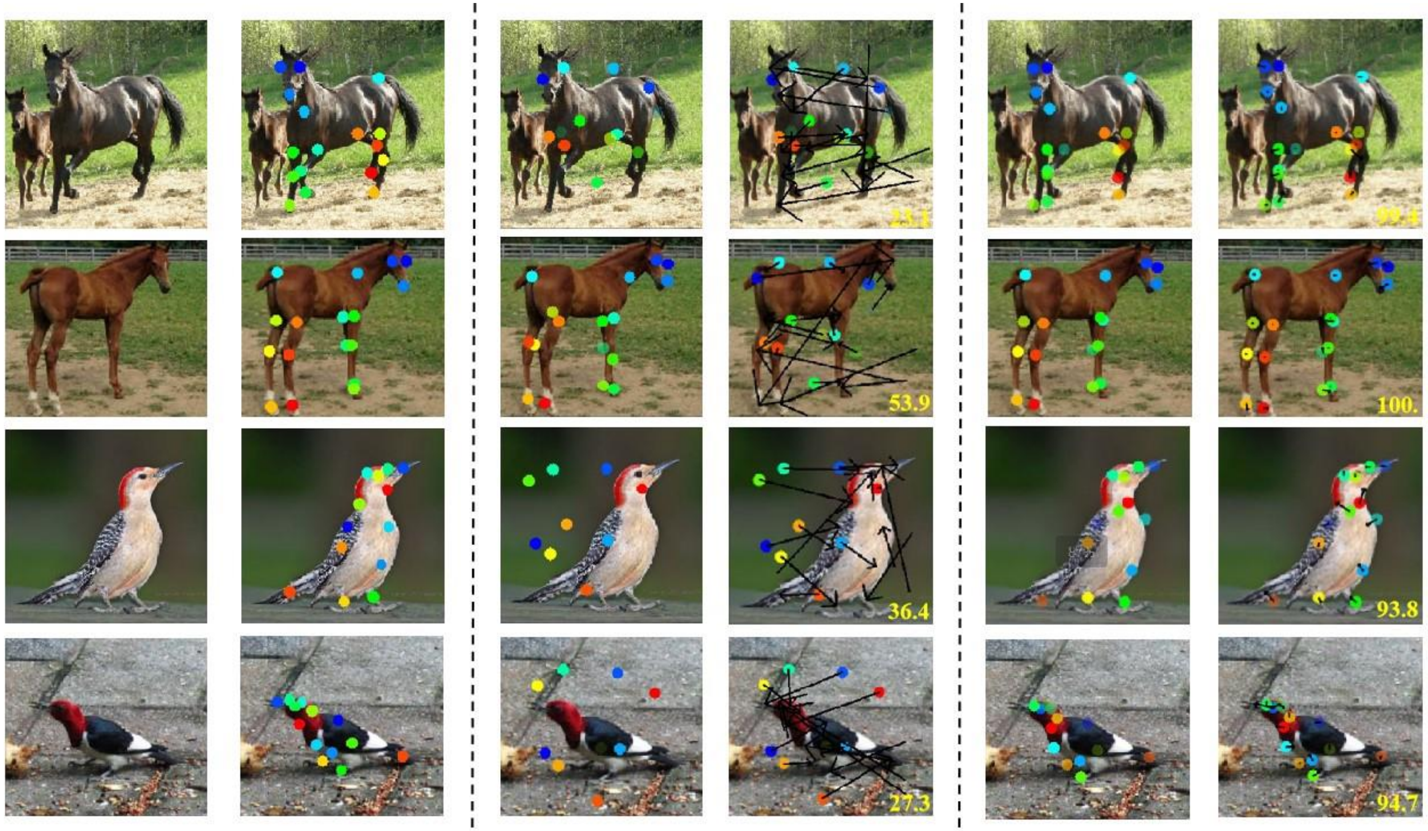
Experiments

Comparison with prior works.

The keypoints located by our model could better fit the object structures and keep semantic consistency across images, which is a general requirement in keypoint estimation.

Figure: The qualitative comparison against the most competitive baseline. The first two columns show the image and GT keypoints. The mid two columns display the localized keypoints and matched keypoints of MetaPoint

Qualitative comparison



1) Image 2) GT Keypoints 3) MetaPoint 4) MetaPoint-$\delta(\cdot)$ 5) Ours 6) Ours-$\delta(\cdot)$

Experiments

Comparison with Other Settings

▶Comparison with few-shot setting:

our method has the unique advantage on transferring knowledge upon diffusion models to facilitate the unsupervised learning of multiple novel classes, which is irreplaceable against few-shot methods.

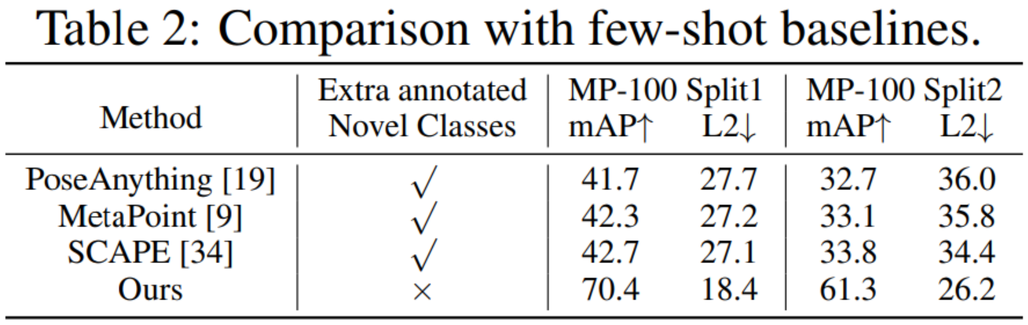


Table 2: Comparison with few-shot baselines.

Method	Extra annotated Novel Classes	MP-100 Split1 mAP↑	MP-100 Split1 L2↓	MP-100 Split2 mAP↑	MP-100 Split2 L2↓
PoseAnything [19]	√	41.7	27.7	32.7	36.0
MetaPoint [9]	√	42.3	27.2	33.1	35.8
SCAPE [34]	√	42.7	27.1	33.8	34.4
Ours	×	70.4	18.4	61.3	26.2

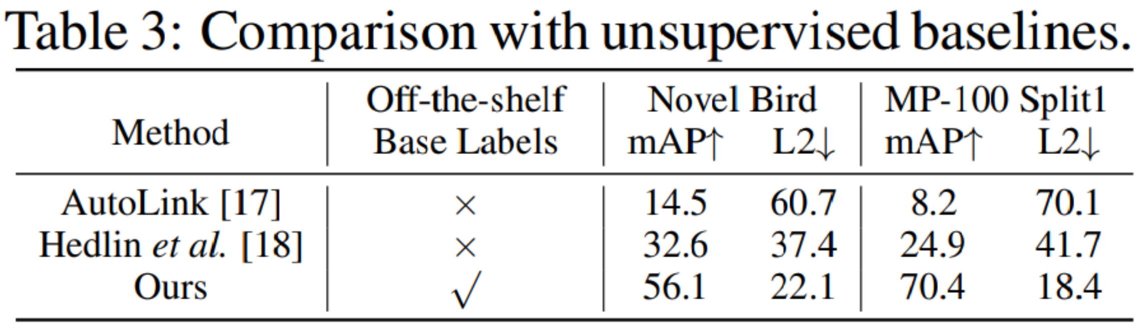


Table 3: Comparison with unsupervised baselines.

Method	Off-the-shelf Base Labels	Novel Bird mAP↑	Novel Bird L2↓	MP-100 Split1 mAP↑	MP-100 Split1 L2↓
AutoLink [17]	×	14.5	60.7	8.2	70.1
Hedlin *et al.* [18]	×	32.6	37.4	24.9	41.7
Ours	√	56.1	22.1	70.4	18.4

▶Comparison with unsupervised setting:

Our method outperforms unsupervised method dramatically in both splits, because our method could simultaneously tackle multiple novel classes and leverage the transferred knowledge.

Experiments

- **Analysis on correspondence transfer**

The learned and transferred correspondences provide beneficial regularization for the unsupervised learning of novel classes.

the PCKs of novel classes are relatively satisfactory, indicating that the transferred correspondences can provide valid regularization.

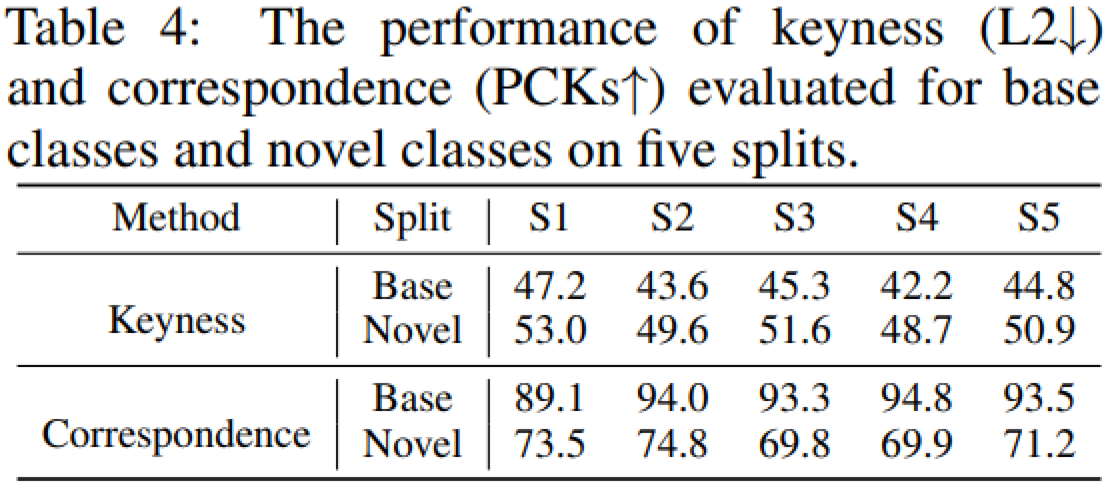


Table 4: The performance of keyness (L2↓) and correspondence (PCKs↑) evaluated for base classes and novel classes on five splits.

Method	Split	S1	S2	S3	S4	S5
Keyness	Base	47.2	43.6	45.3	42.2	44.8
	Novel	53.0	49.6	51.6	48.7	50.9
Correspondence	Base	89.1	94.0	93.3	94.8	93.5
	Novel	73.5	74.8	69.8	69.9	71.2

Experiments

- **Analysis on keyness transfer**

To further explore the transferability of keyness, we visualize the estimated keypoints and keyness map in the intermediate training stage to see whether the keyness could provide beneficial guidance.

Generally, the estimated keyness maps are more approximate to their GT than the estimated keypoints. Therefore, some biased keypoints (e.g., the right eye of horse in the top row) could be guided to offset to the most key pixels, leading to precise keypoints of novel classes.

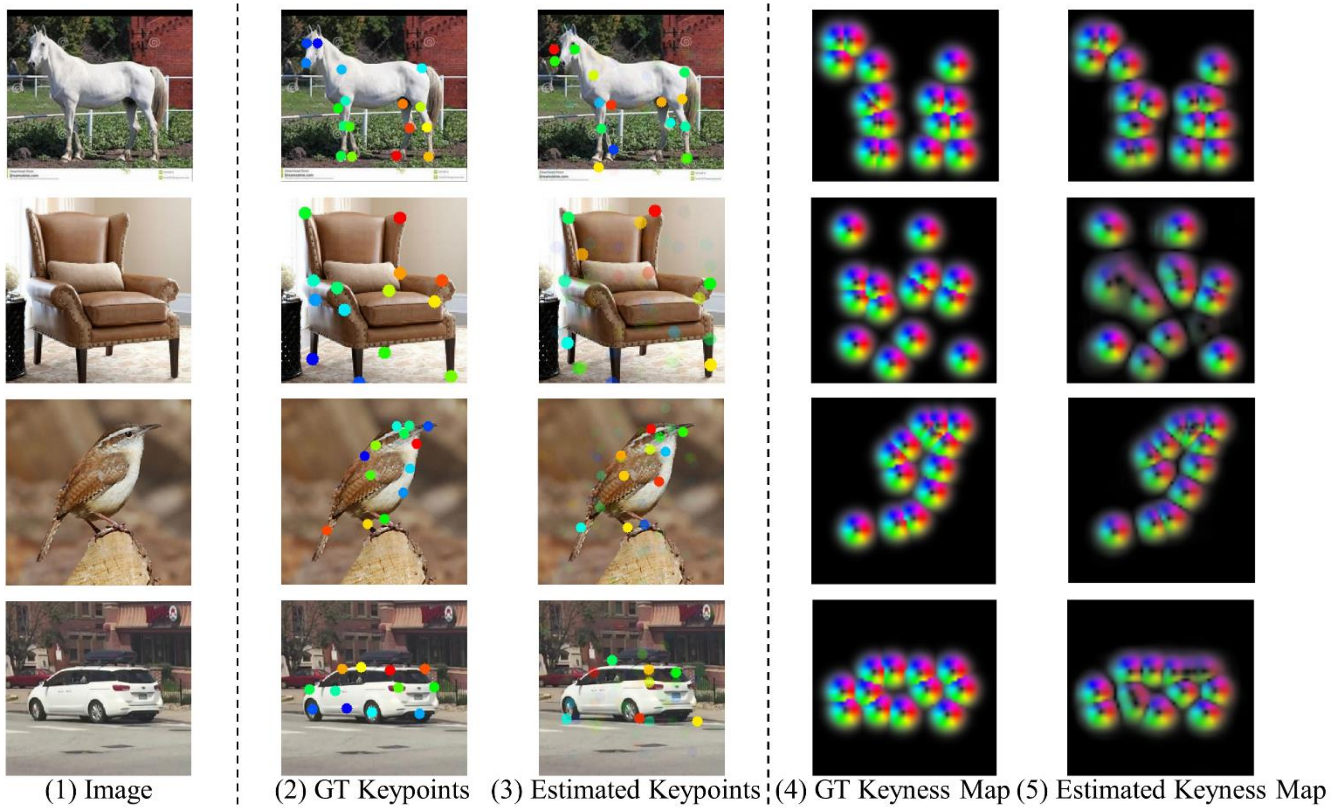


Figure: Qualitative analysis of estimated keypoints and keyness

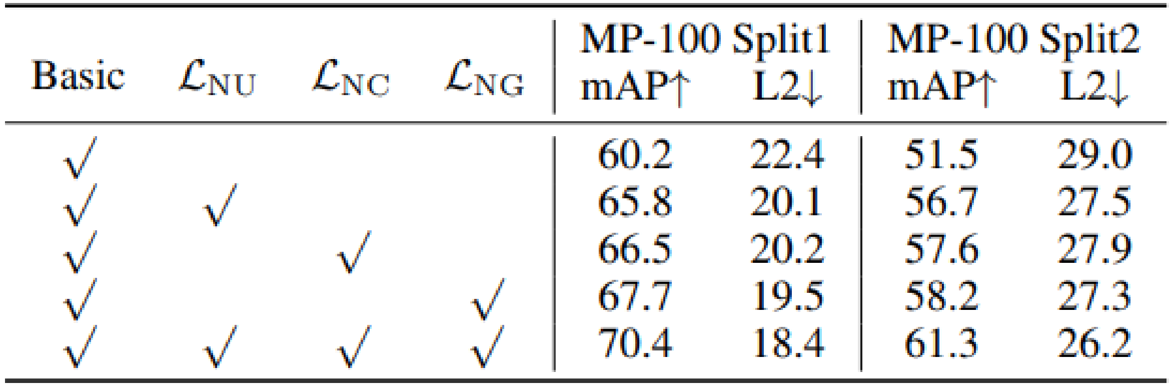
Experiments

‣ Ablation study

To study the contributions of our loss terms, we evaluate the combinations of our basic model, $\mathcal{L}_{NG}, \mathcal{L}_{NC}, \mathcal{L}_{NU}$.

With all losses enabled, our method achieves the best results. Thus, our loss terms are effective and complementary.

Table 5: Ablation results of our proposed framework on two splits.



Basic	$\mathcal{L}_{\mathrm{NU}}$	$\mathcal{L}_{\mathrm{NC}}$	$\mathcal{L}_{\mathrm{NG}}$	MP-100 Split1		MP-100 Split2	
				mAP↑	L2↓	mAP↑	L2↓
√				60.2	22.4	51.5	29.0
√	√			65.8	20.1	56.7	27.5
√		√		66.5	20.2	57.6	27.9
√			√	67.7	19.5	58.2	27.3
√	√	√	√	70.4	18.4	61.3	26.2

Conclusion

- we have explored weak-shot keypoint estimation, where multiple novel classes are learned from unlabeled images with the support of labeled base classes.

- We have proposed a novel framework transferring keyness and correspondence to facilitate the unsupervised learning of novel classes. By transferring keyness and correspondence, our framework has achieved favourable performance for weak-shot keypoint estimation.

- Extensive experiments and analyses on large-scale benchmark MP-100 demonstrate our effectiveness.

Thank you !