InfoGeo: Information-Theoretic Object-Centric Learning for Cross-View Generalizable UAV Geo-Localization

ICML International Conference On Machine Learning

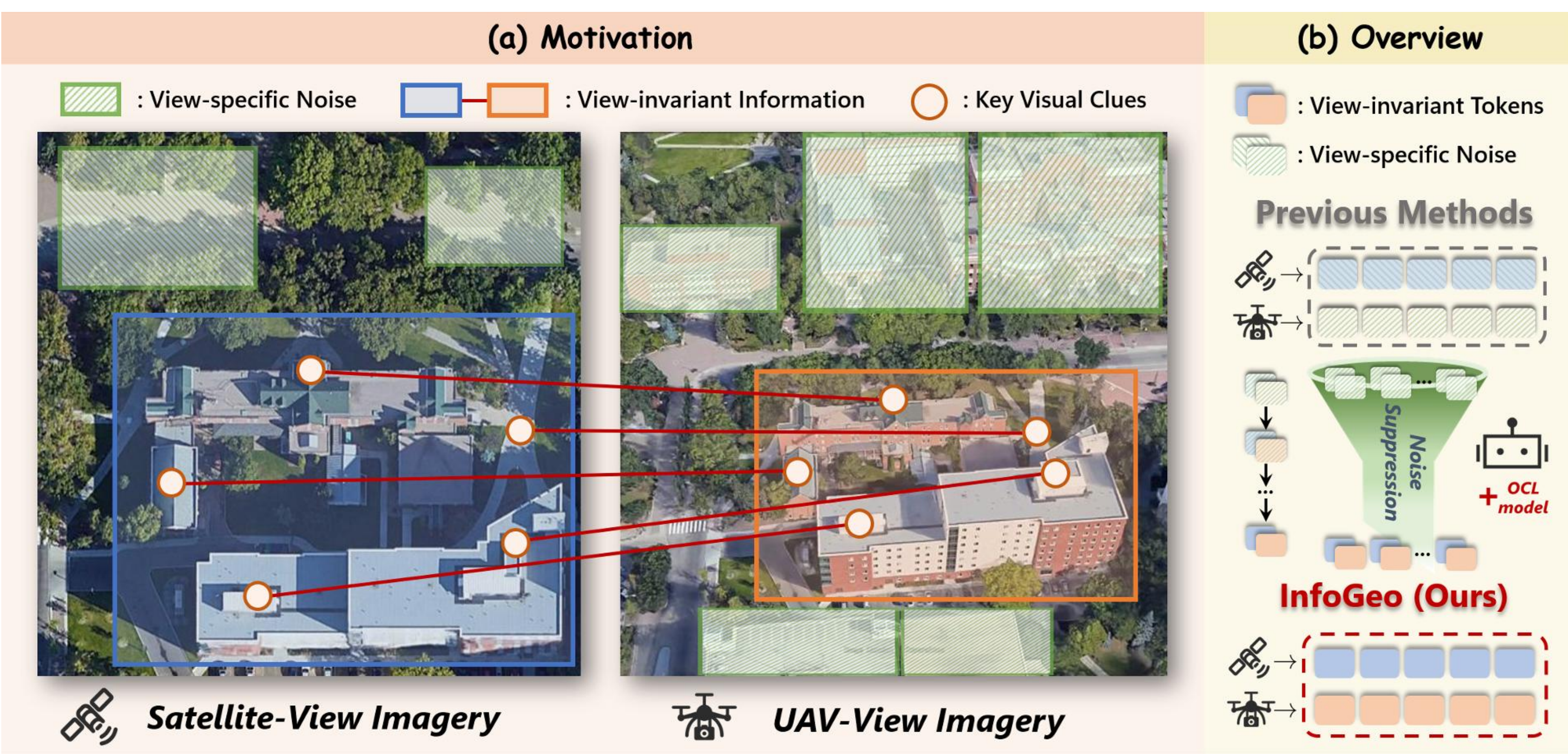
Paper GitHub Personal Website Lab Homepage

Hongyang Zhang[1, *], Maonan Wang[2,3,*], Ziyao Wang[1], Hongrui Yin[1, 2], Man On Pun[†]

[*]Equal Contribution, [†]Corresponding Author, [1]CUHK(SZ), [2]CUHK, [3]Shanghai AI Laboratory

上海人工智能实验室 Shanghai Artificial Intelligence Laboratory

Motivation



▶**(Limitation 1)** Most existing CVGL methods solely rely on global feature alignment, which is vulnerable to ***domain shifts*** caused by regional appearance, multi-weather, and viewpoint variations due to coarse-grained features.

▶**(Limitation 2)** UAV scenarios are more challenging because the broader aerial perspective introduces ***dense fine-grained objects, severe visual clutter and view-specific noisy signals***, leading to insufficient robustness.

Main Contribution

▶We propose **InfoGeo**, it reformulates the CVGL task as an ***information bottleneck (IB) process***: (1) maximizing view-invariant object-structural information, (2) suppressing view-specific nuisance factors.

▶The framework introduce ***Cross-view Adaptive Concept Selection*** and ***Concept Structural Relation Reasoning*** modules for cross-view object-centric learning.

▶Extensive experiments on four UAV benchmarks show that InfoGeo improves robustness and generalization under challenging settings.

Information Theory for CVGL

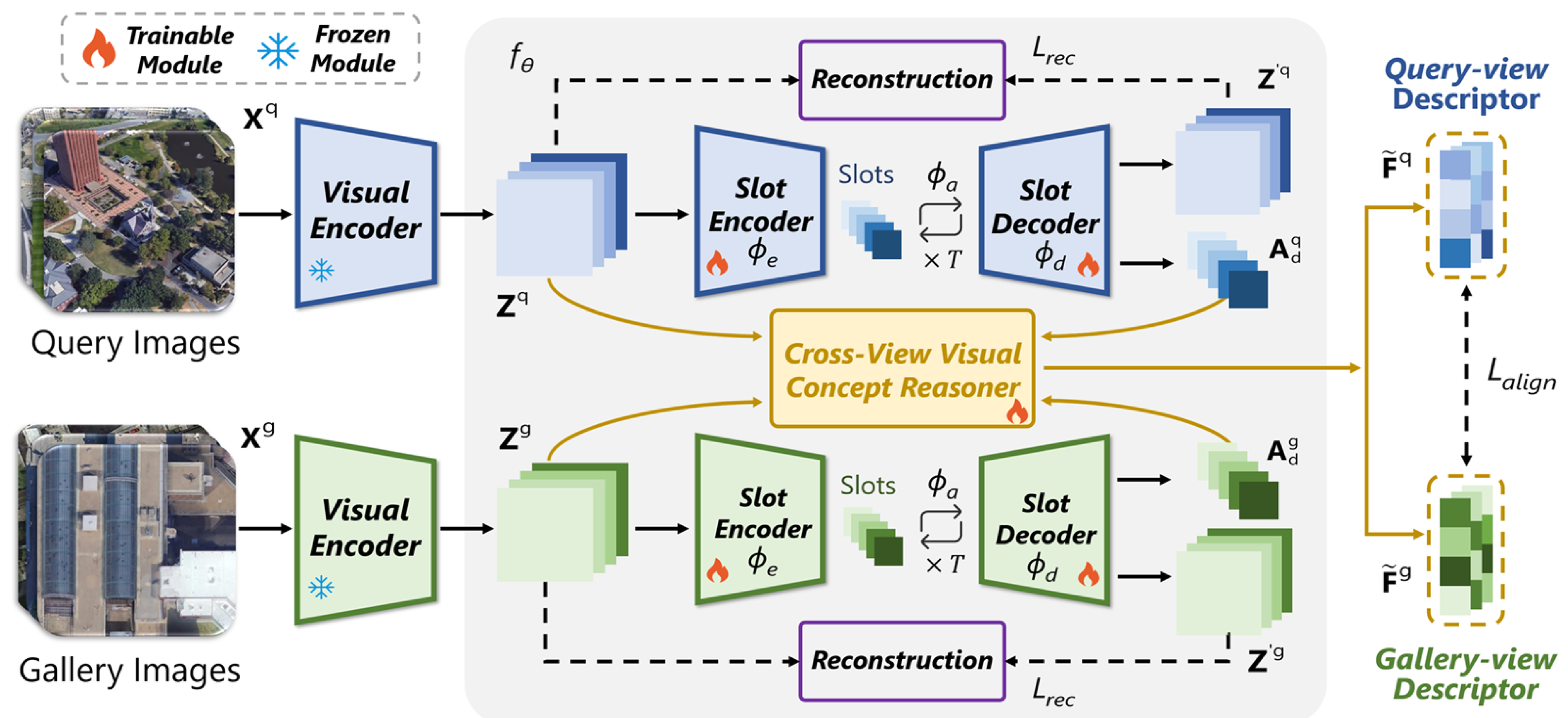
View-Invariant Information Bottleneck Principle: The core idea of IB theory is to maximize mutual information (MI) of CVGL while minimizing task-irrelevant information.

$$\max_{v\in\{q,g\}} \mathcal{L}_{IB}^{(v)} = I(\hat{\mathbf{Z}}^{(v)};\mathbf{Y}) - \beta I(\hat{\mathbf{Z}}^{(v)};\mathbf{Z}^{(v)} \mid \mathbf{Y})$$

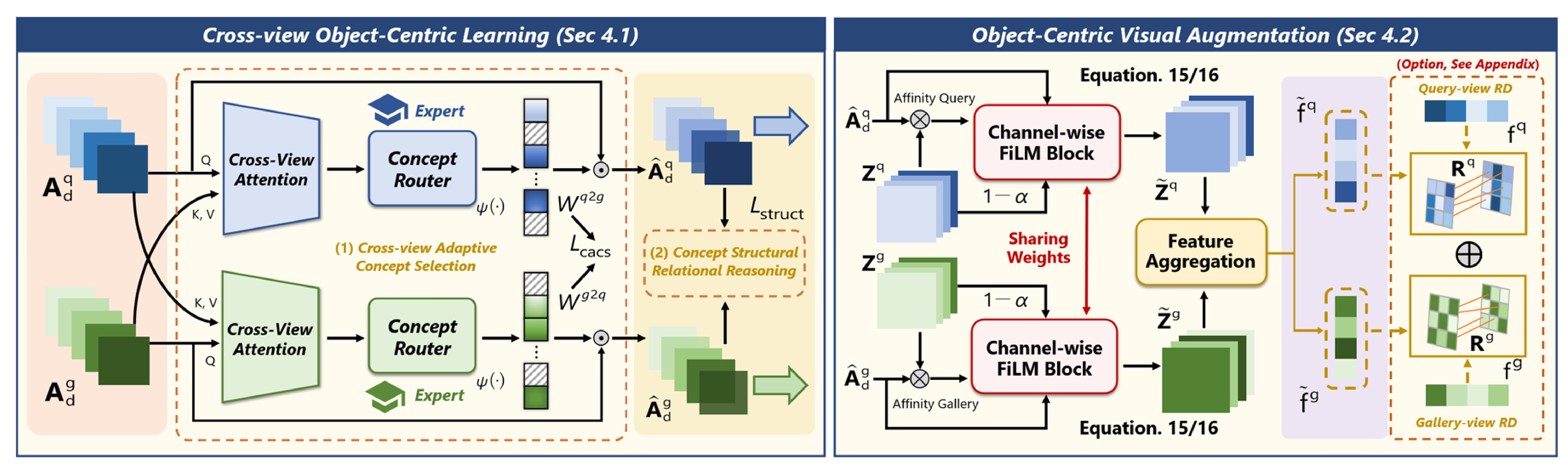
Then, the equation can be converted into the form of cross-view feature alignment (***Query→Gallery***):

$$\max \mathcal{L}_{IB}^{q\to g} = I(\hat{\mathbf{Z}}^{q};\hat{\mathbf{Z}}^{g}) - \beta I(\hat{\mathbf{Z}}^{q};\mathbf{Z}^{q} \mid \hat{\mathbf{Z}}^{g})$$

Methodology



Overview: We first introduce the Object-Centric Learning (OCL) modules to obtain object-centric features. Furthermore, ***Cross-View Visual Concept Reasoner*** is proposed in the perspective of the IB principle.



▶***Cross-View Object-Centric Learning***: We firstly employ Cross-view Adaptive Concept Selection (CACS) to filter out view-specific noisy signals. Then, Concept Structural Relation Reasoning (CSRR) is used to align concept graph.

$$\mathcal{L}_{\text{cacs}} = \mathbb{E}\left[\mathbf{W}_{\text{cv}} \odot (1-\mathbf{W}_{\text{cv}})\right] - \mathbb{E}\left[\text{Var}(\mathbf{W}_{\text{cv}})\right],$$

$$\mathcal{L}_{\text{struct}} = \min_{\mathbf{Q}\in\mathcal{O}(r)} \|\mathbf{U}^q - \mathbf{U}^g\mathbf{Q}\|_2^2,$$

▶***Object-Centric Visual Augmentation***: A FiLM-style channel feature-wise modulation module is proposed to incorporated object-centric representations into global descriptors.

Applying IB theory for Optimization: We reformulate the optimization of cross-view OCL modules with View-Invariant Information Bottleneck Principle:

$$\max \mathcal{L}_{\text{IB}} = \underbrace{I(\mathbf{U}^g;\mathbf{U}^q)}_{\mathcal{L}_{\text{struct}}} \underbrace{-\left[I(\hat{\mathbf{A}}_d^g;\mathbf{A}_d^q \mid \hat{\mathbf{A}}_d^q) + I(\hat{\mathbf{A}}_d^q;\mathbf{A}_d^g \mid \hat{\mathbf{A}}_d^g)\right]}_{\mathcal{L}_{\text{cacs}}}$$

Cross-View MI Maximization: The first term encourages the model to preserve view-invariant information.

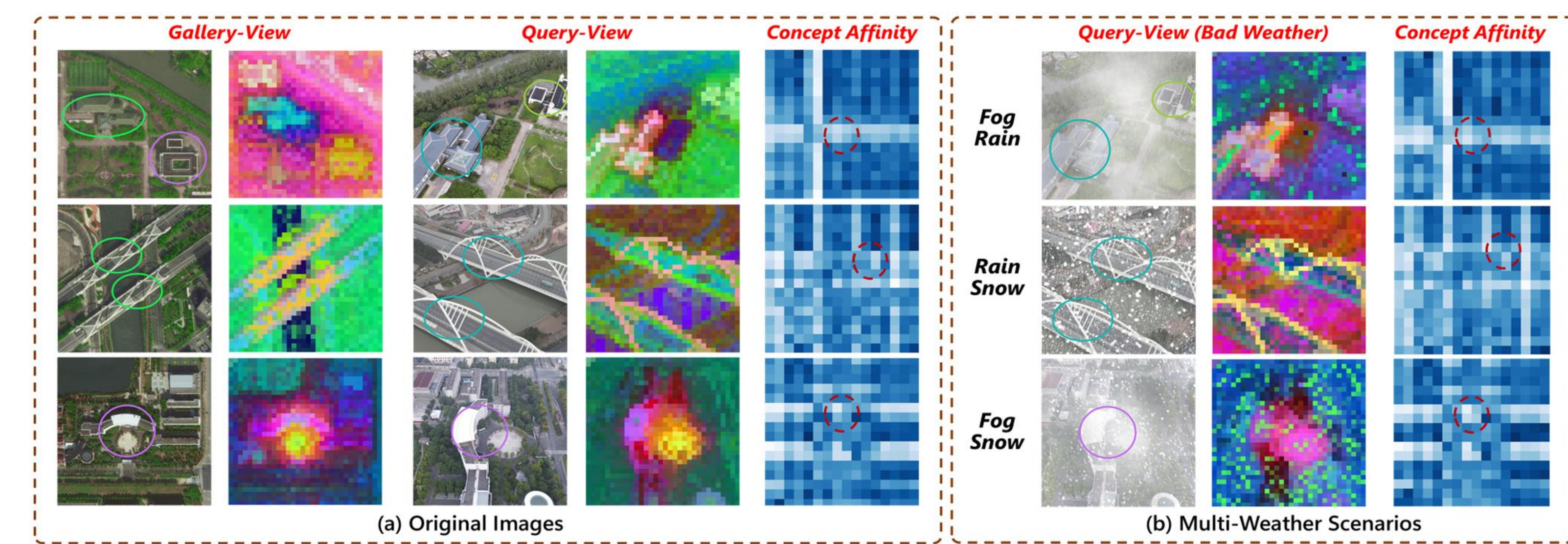
View-Specific Noise Compression: The second term aims to filter out task-irrelevant noise.

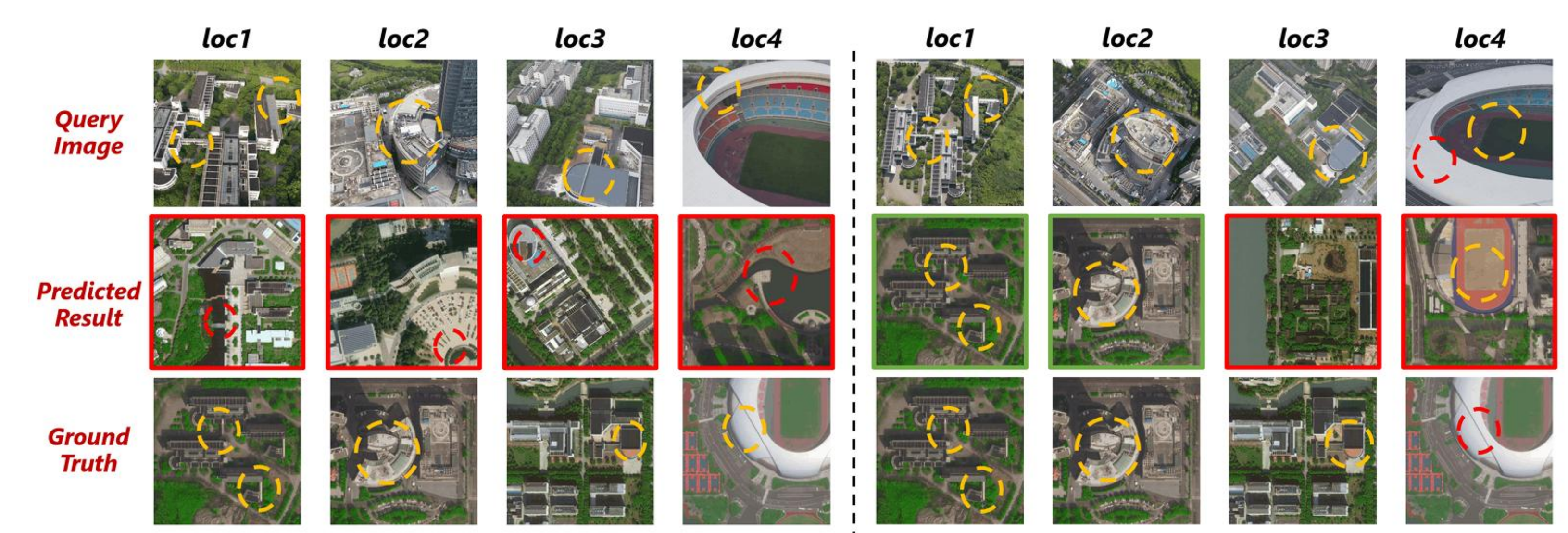
Experimental Results

▶**Main Results:** We conduct cross-area experiments on the UAV benchmarks with different scenarios.

Model	University→SUES		University → DenseUAV		DenseUAV→SUES		DenseUAV→University	
	R@1↑	AP↑	R@1↑	AP↑	R@1↑	AP↑	R@1↑	AP↑
MCCG (Shen et al., 2023)	70.31	74.58	17.54	13.01	80.09	83.18	50.52	55.35
Sample4Geo (Deuser et al., 2023)	82.03	85.12	33.46	22.28	78.94	81.86	37.90	42.54
DAC (Xia et al., 2024)	87.65	89.80	37.81	26.99	86.44	88.70	45.41	50.03
CAMP (Wu et al., 2024)	88.34	90.49	36.94	26.39	86.79	88.93	44.11	48.60
MFRGN (Wang et al., 2024)	78.28	81.75	23.60	16.08	67.22	71.14	37.95	41.73
EM-CVGL (Li et al., 2024b)	74.45	78.51	13.47	10.19	56.43	61.54	32.76	38.05
CVcities (Huang et al., 2024a)	90.27	91.77	38.87	28.44	89.67	91.34	66.08	70.01
Game4Loc (Ji et al., 2025a)	86.75	89.07	17.76	13.62	67.19	71.40	26.47	31.05
MMGeo (Ji et al., 2025b)	85.01	87.60	22.88	18.54	79.05	82.05	33.92	38.28
InfoGeo (Ours **w/o RD**)	**93.38** $_{+3.11}$	**94.35** $_{+2.58}$	**39.88** $_{+1.01}$	**29.13** $_{+0.69}$	**93.42** $_{+3.75}$	**94.55** $_{+3.21}$	**68.16** $_{+2.08}$	**72.15** $_{+2.11}$
InfoGeo* (Ours **w/ RD**)	**95.07** $_{+4.80}$	**95.85** $_{+4.08}$	**42.30** $_{+3.43}$	**31.78** $_{+3.34}$	**93.41** $_{+3.74}$	**94.52** $_{+3.18}$	**69.37** $_{+3.29}$	**73.21** $_{+3.20}$

▶**Qualitative Results:** PCA maps, Concept affinities and Retrieval results on cross-view and multi-weather settings.





(a) SUES-200 (150m) (b) SUES-200 (300m)

▶**Ablation Studies:** We further explore the effectiveness of cross-view OCL modules and hyper-parameters.

Model	University→SUES		DenseUAV→SUES	
	R@1↑	AP↑	R@1↑	AP↑
Baseline	82.73	85.60	80.20	83.24
Baseline[†]	86.10	88.28	83.28	86.14
OCVA-only	87.80	89.69	85.75	88.25
$+\mathcal{L}_{\text{cacs}}$	88.82	90.61	85.97	88.40
$+\mathcal{L}_{\text{struct}}$	90.87	92.41	88.20	89.33
+ RD	**91.80**	**93.18**	**88.70**	**90.73**

Method	University→SUES		DenseUAV→SUES	
	R@1↑	AP↑	R@1↑	AP↑
w/o CSRR	88.65	90.55	85.88	88.34
InfoNCE	88.20	90.03	86.64	88.77
$\mathcal{L}_{\text{struct}}\ (r=2)$	89.22	91.00	86.85	89.06
$\mathcal{L}_{\text{struct}}\ (r=4)$	**91.80**	**93.18**	**88.70**	**90.73**
$\mathcal{L}_{\text{struct}}\ (r=8)$	86.42	88.51	85.75	88.11