Relational Proxies: Emergent Relationships as Fine-Grained Discriminators

Abhra Chaudhuri

Massimiliano Mancini

Zeynep Akata

Anjan Dutta

University of Exeter

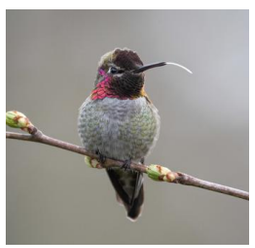
EBERHARD KARLS UNIVERSITÄT TÜBINGEN

UNIVERSITY OF SURREY

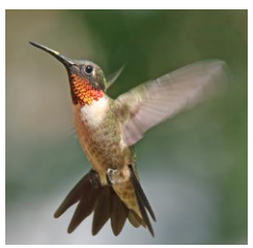
Fine-Grained Visual Categorization (FGVC)

Coarse-Grained

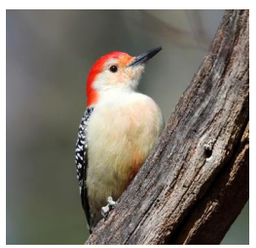
Birds



Anna's Hummingbird



Ruby-Throated Hummingbird



Red-Bellied Woodpecker

Fine-Grained

Plants

Common Dandelion

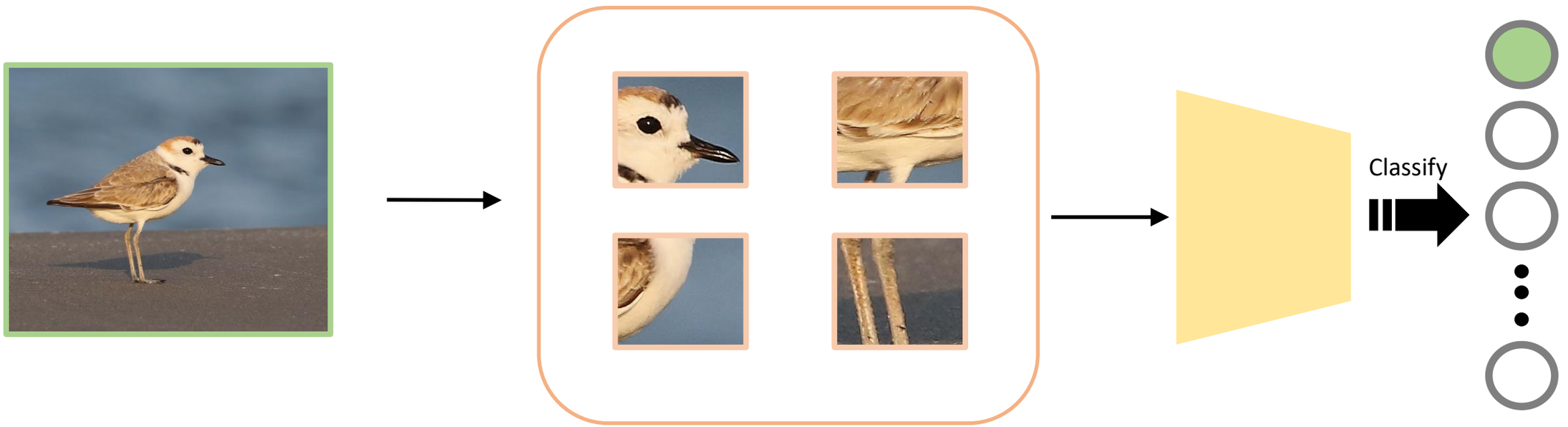
Colt's-Foot

Common Sow-Thistle

Fine-Grained

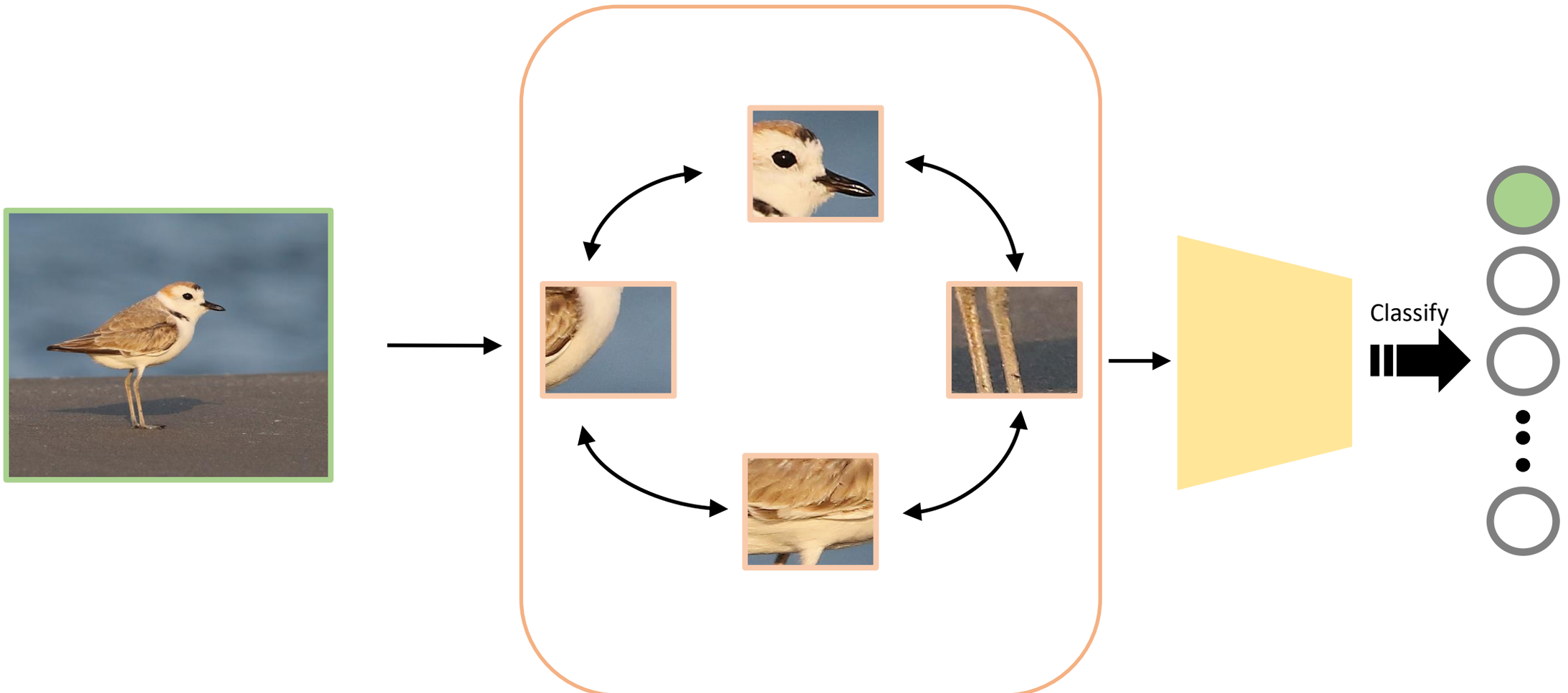
Source: iNaturalist

Part-to-class mapping



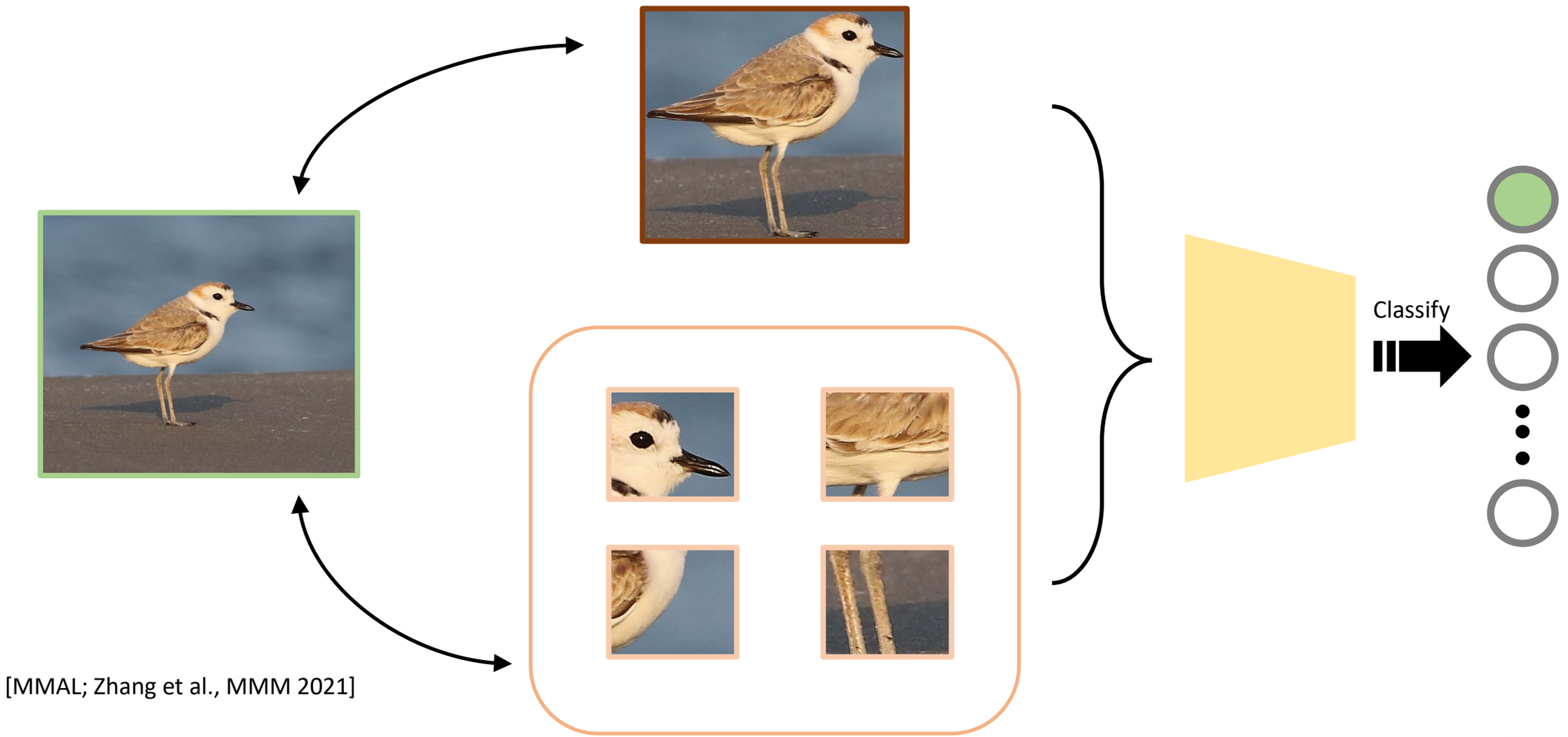
[Part based R-CNN; Zhang et al., ECCV 2014], [Part-stacked CNN; Huang et al., CVPR 2016],
[TransFG; He et al., AAAI 2022]

Cross Locality Relationship-to-class mapping



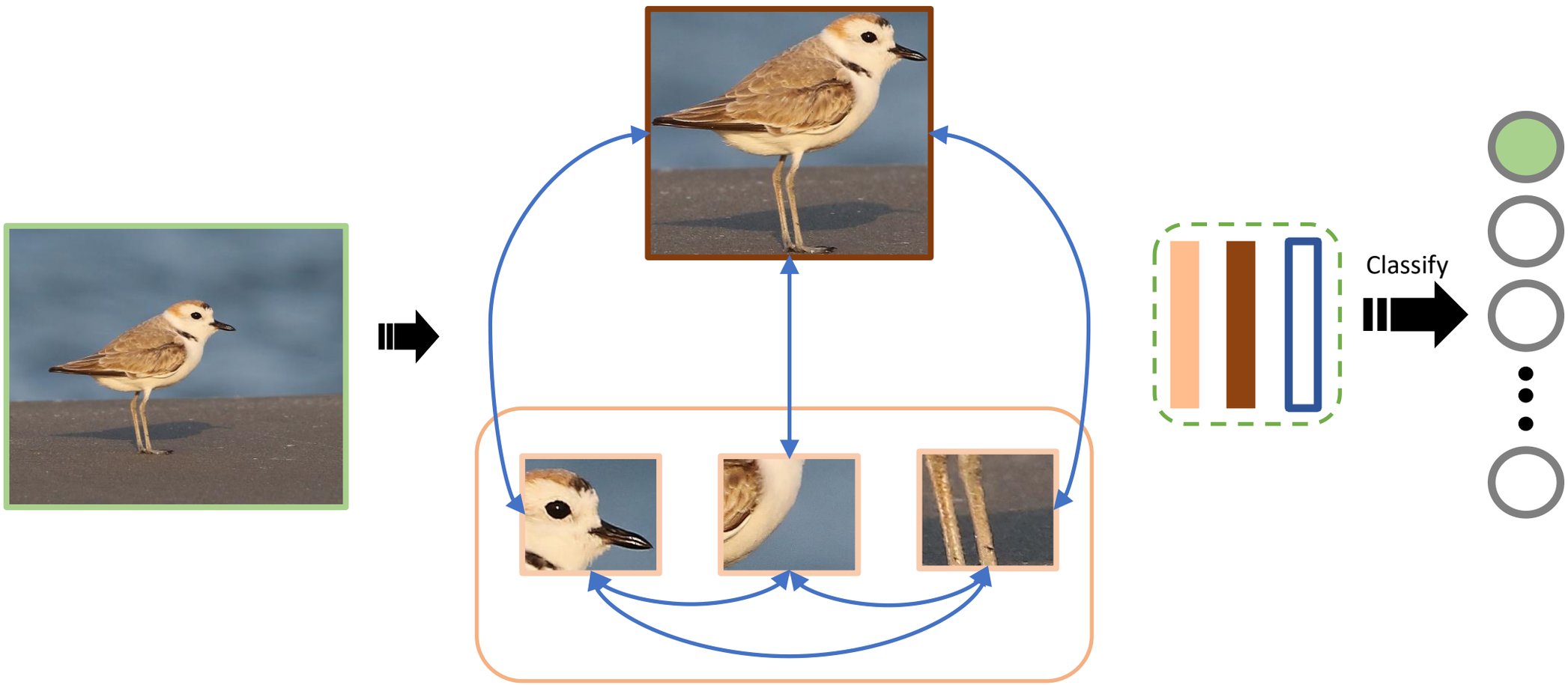
[CAP; Behera et al., AAAI 2021]

Local + Global View-to-class mapping



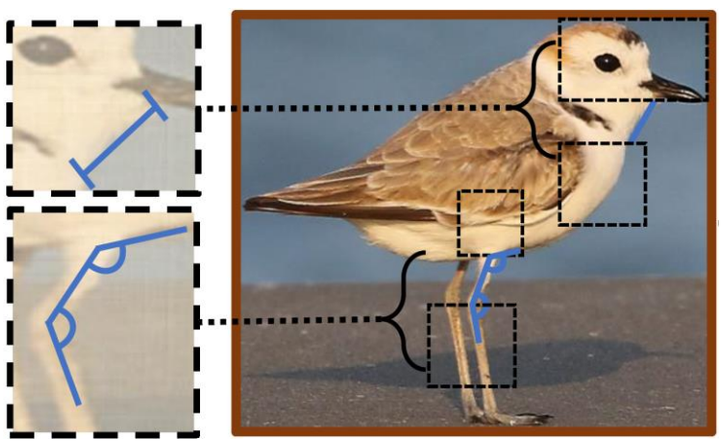
[MMAL; Zhang et al., MMM 2021]

The Missing Bit – Cross-View Relationships

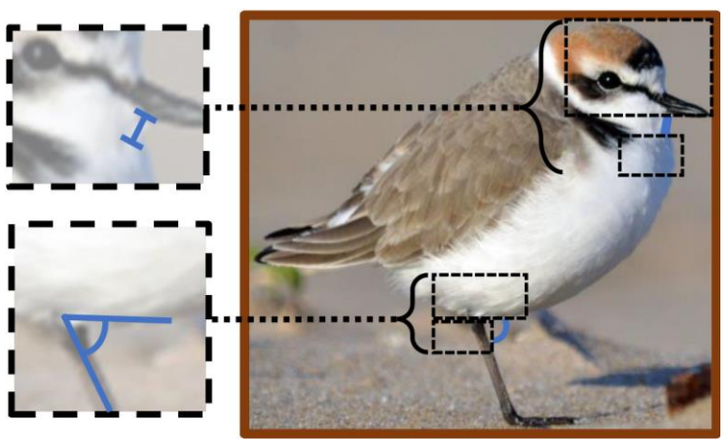


[Unsupervised Parts; Choudhury et al., NeurIPS 2021]
[DiNo; Caron et al., ICCV 2021]

Intuition – Why Emergent Relationships?

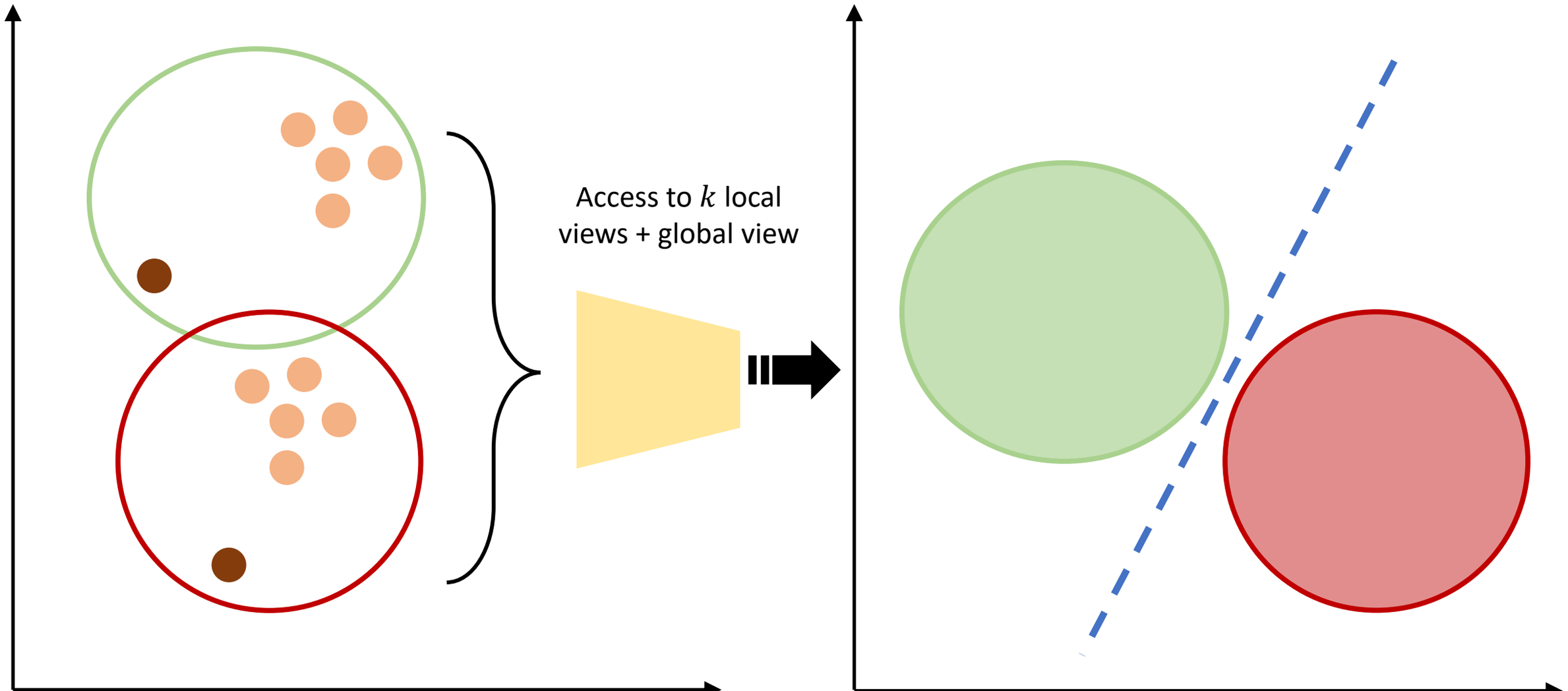


White-Faced Plover

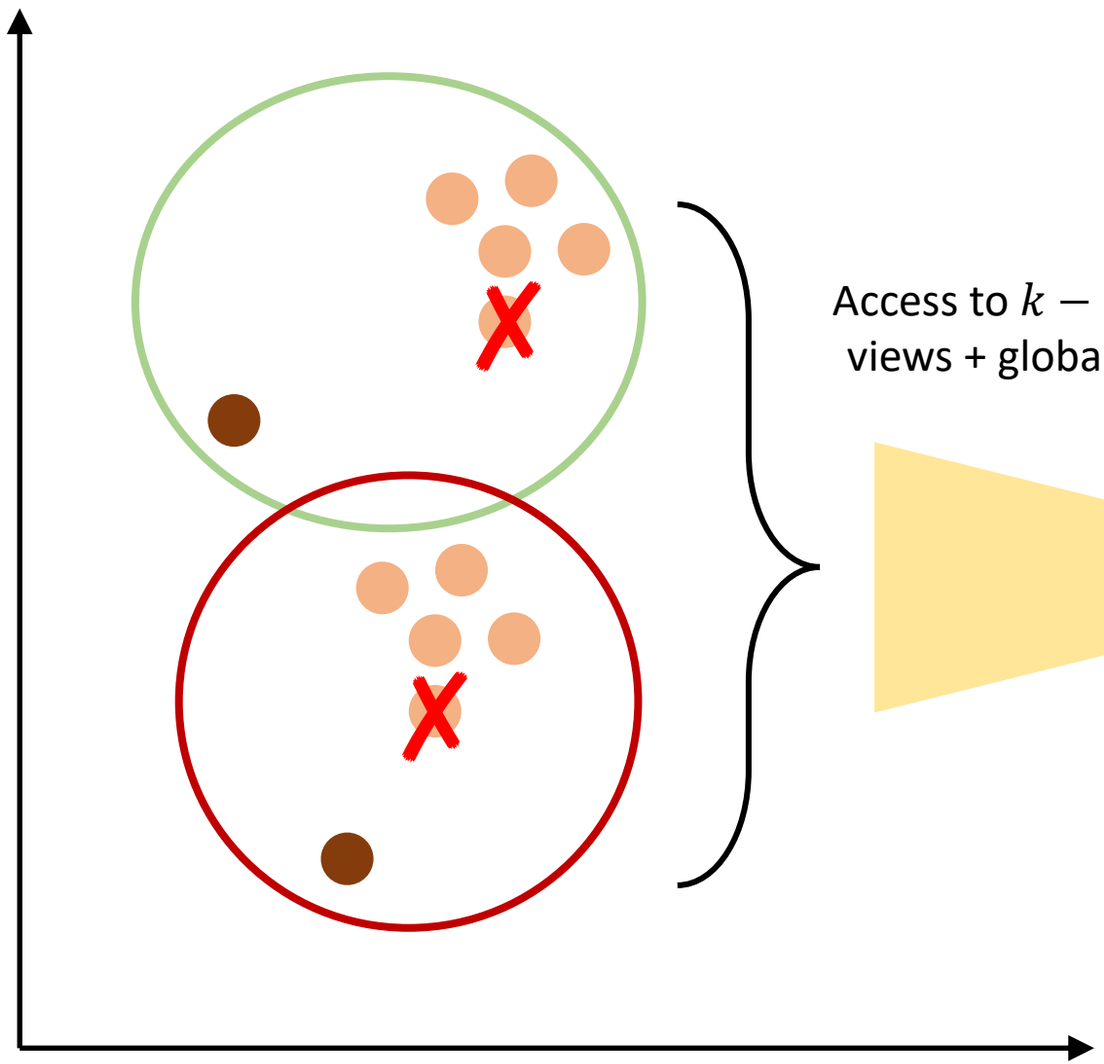


Kentish Plover

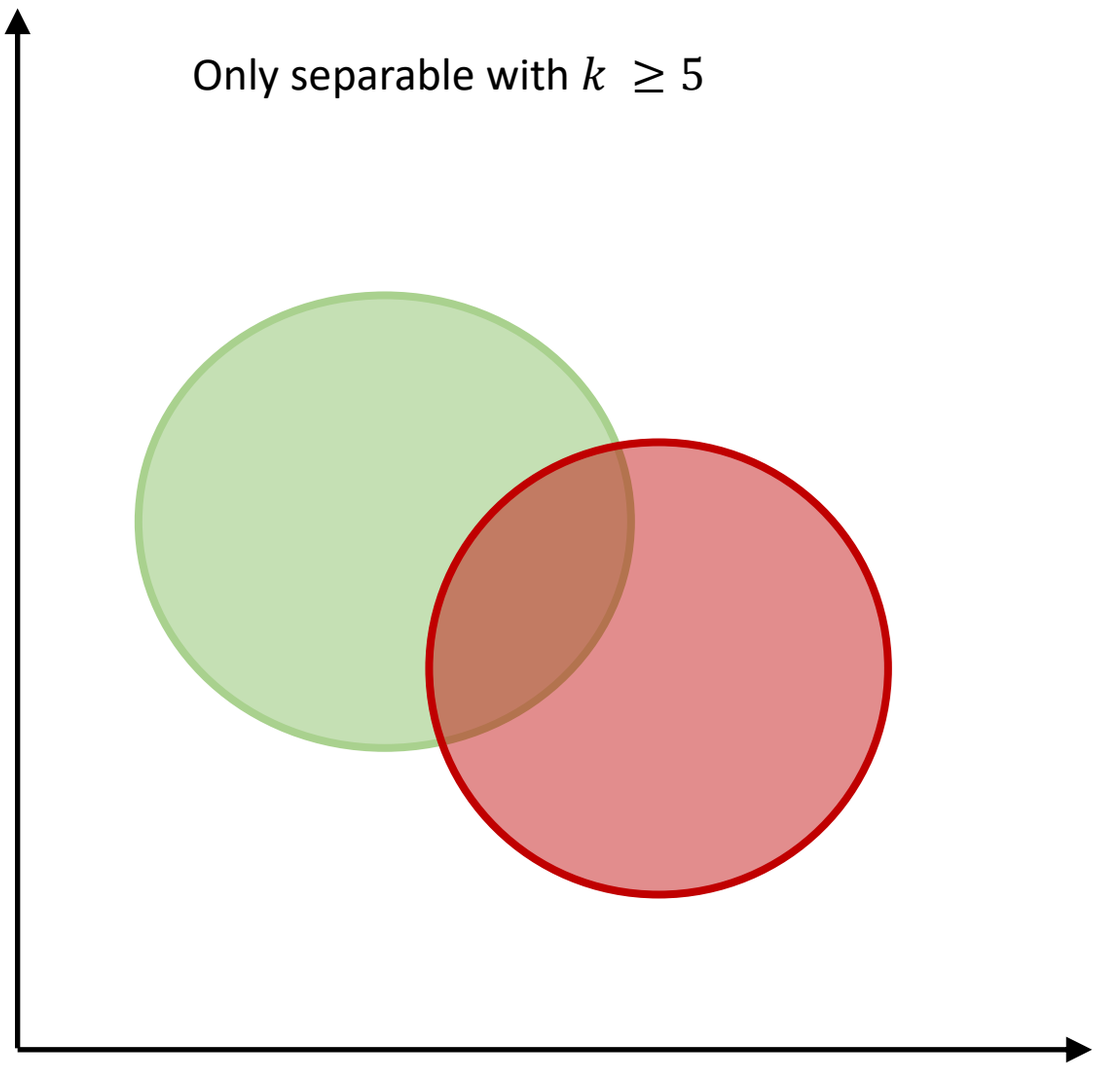
k-distinguishability



k-distinguishability



Access to $k-1$ local views + global view



Only separable with $k \geq 5$

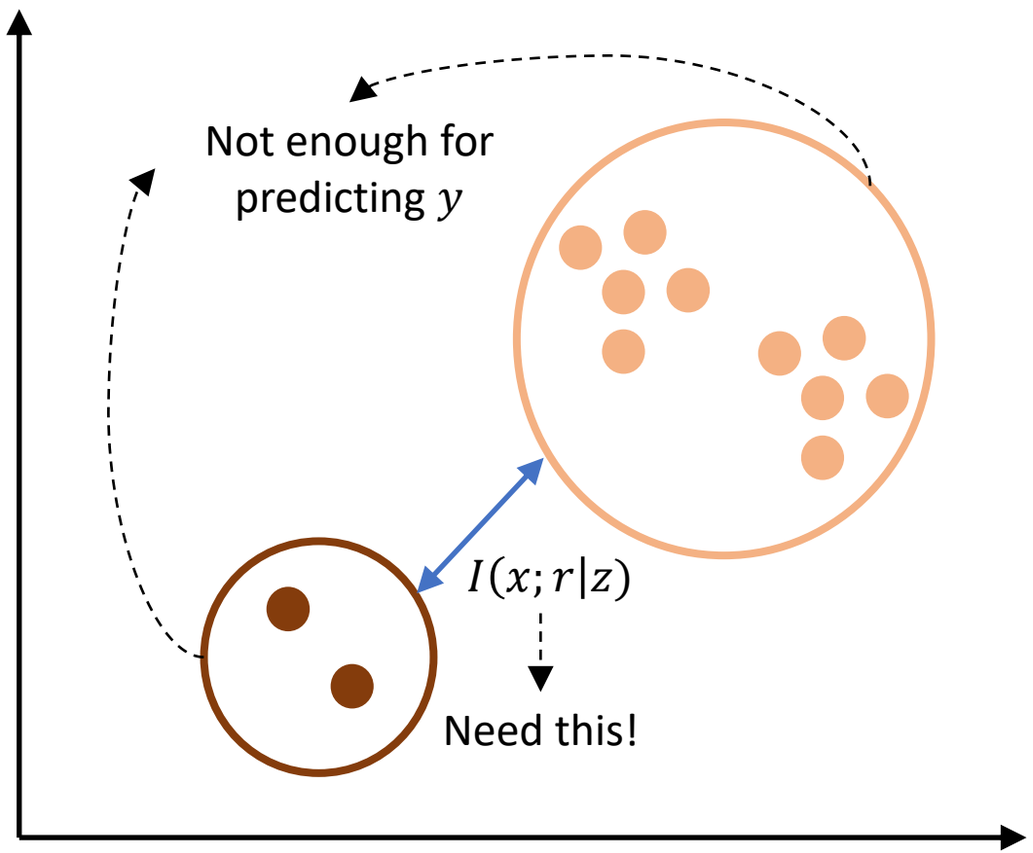
Fine-Grained Visual Categorization

$\mathcal{P}_{FGVC}$: Problems with at least one pair of k-distinguishable classes

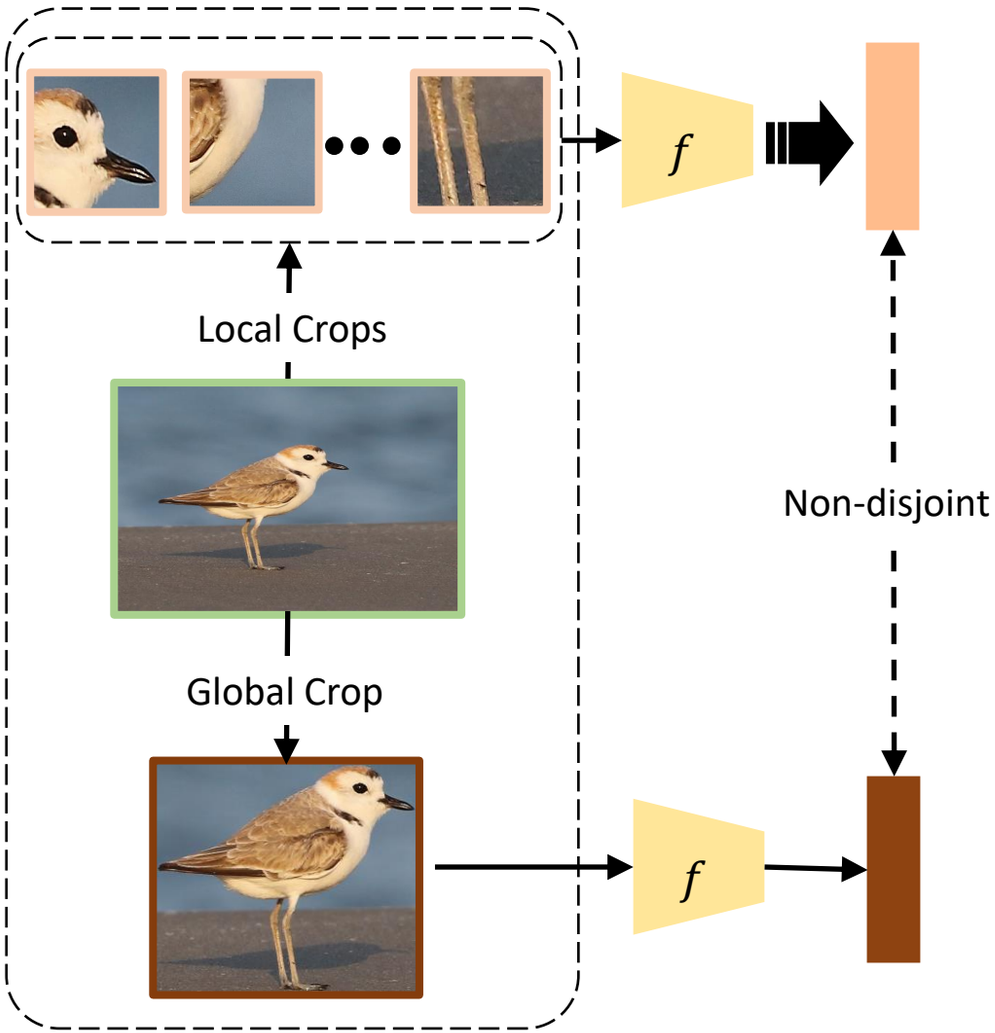
Disjoint Representations

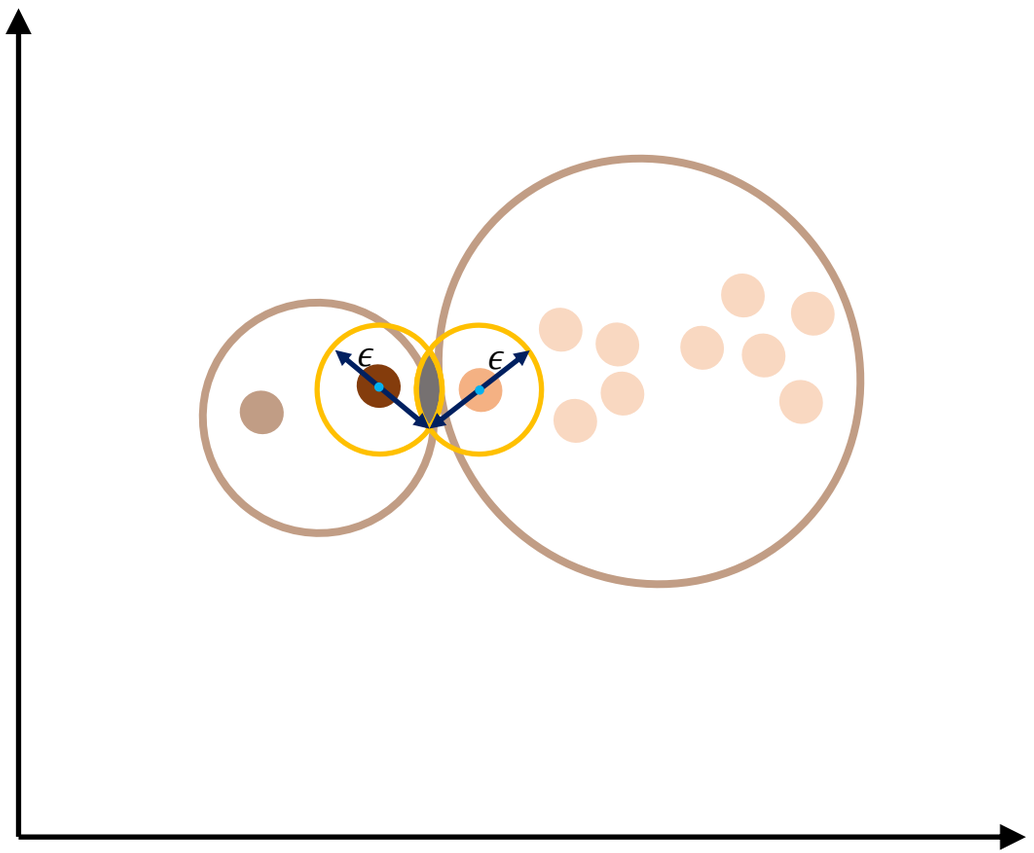
- Independently encodes global and local views
- Does not consider relational information
- $I(x;y) = I(z;y) + \underbrace{I(x;r|z)}_{\text{Information Gap}}$

Proposition 1: $I(x;y) > I(z;y)$



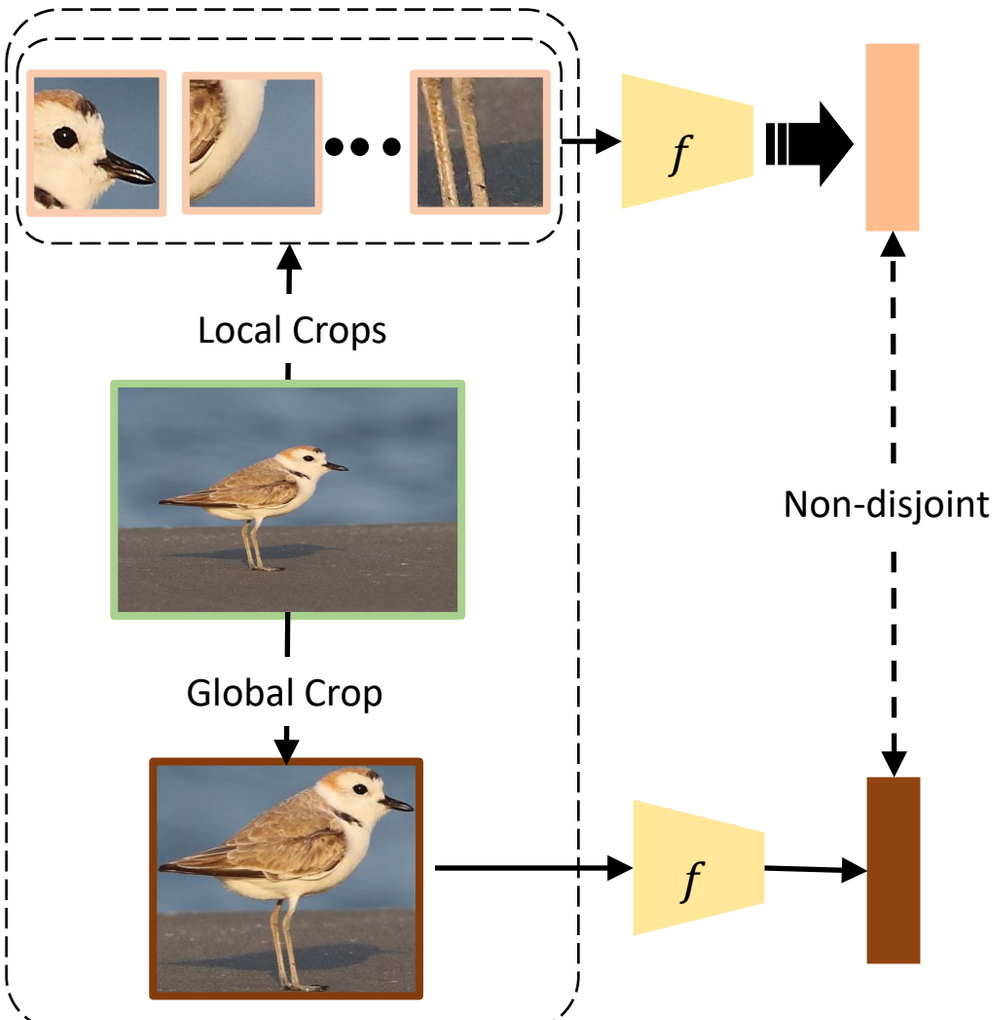
Necessity for Disjointness



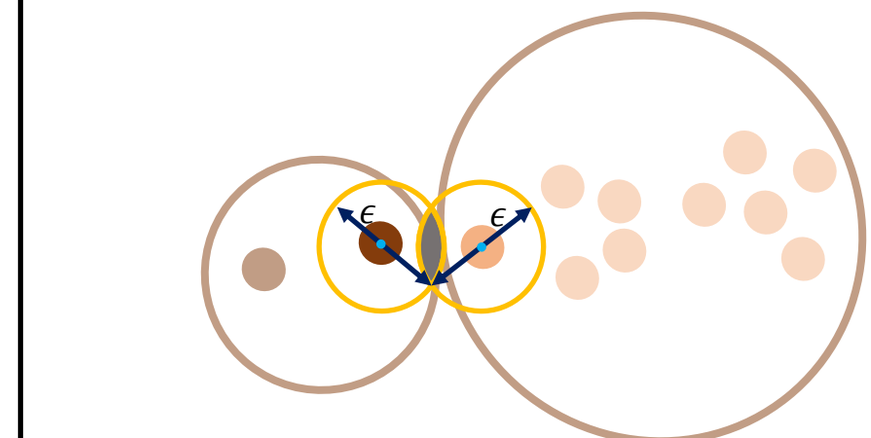


Lemma 3: Leads to misclassification in the fine-grained setting

Necessity of Separate Relational Encoder

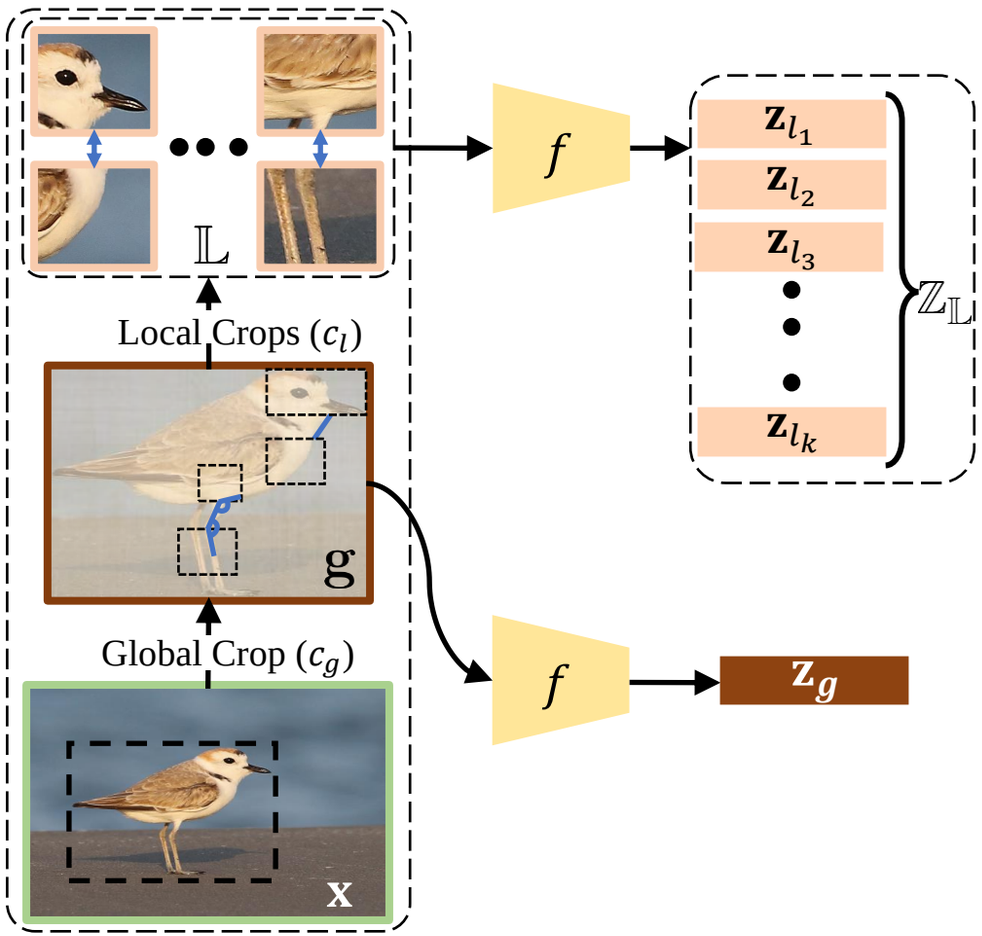


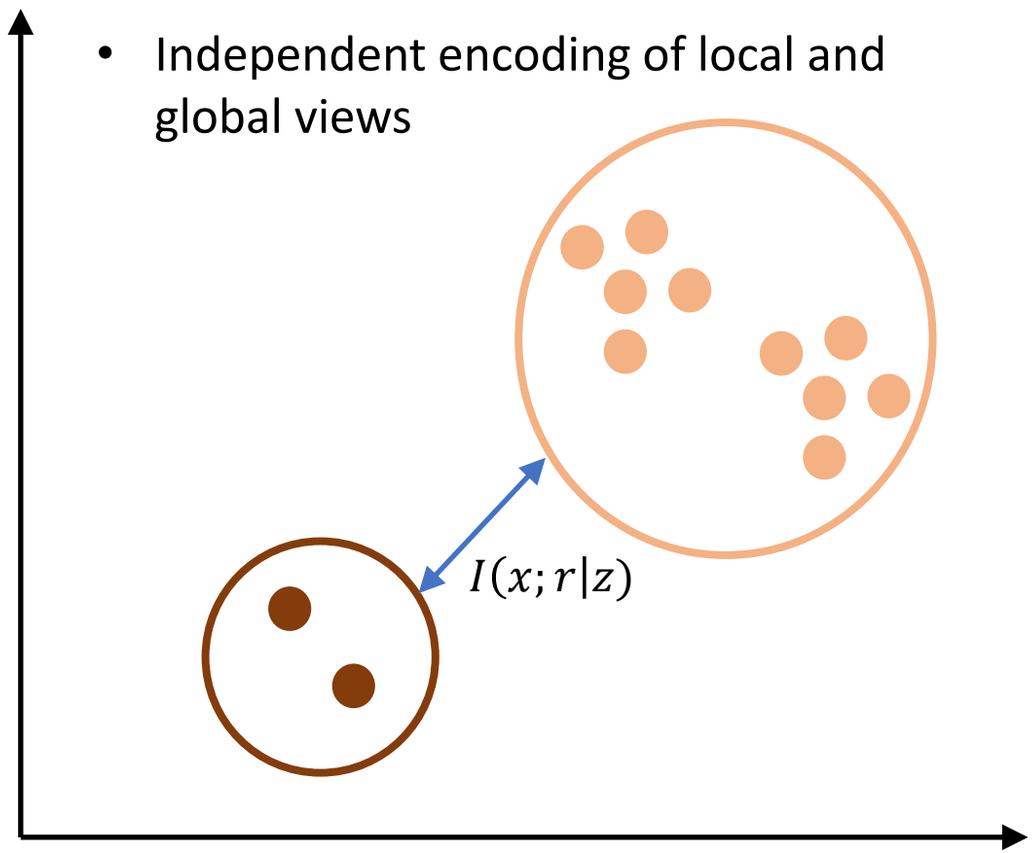
Proposition 2: f alone cannot model relationships



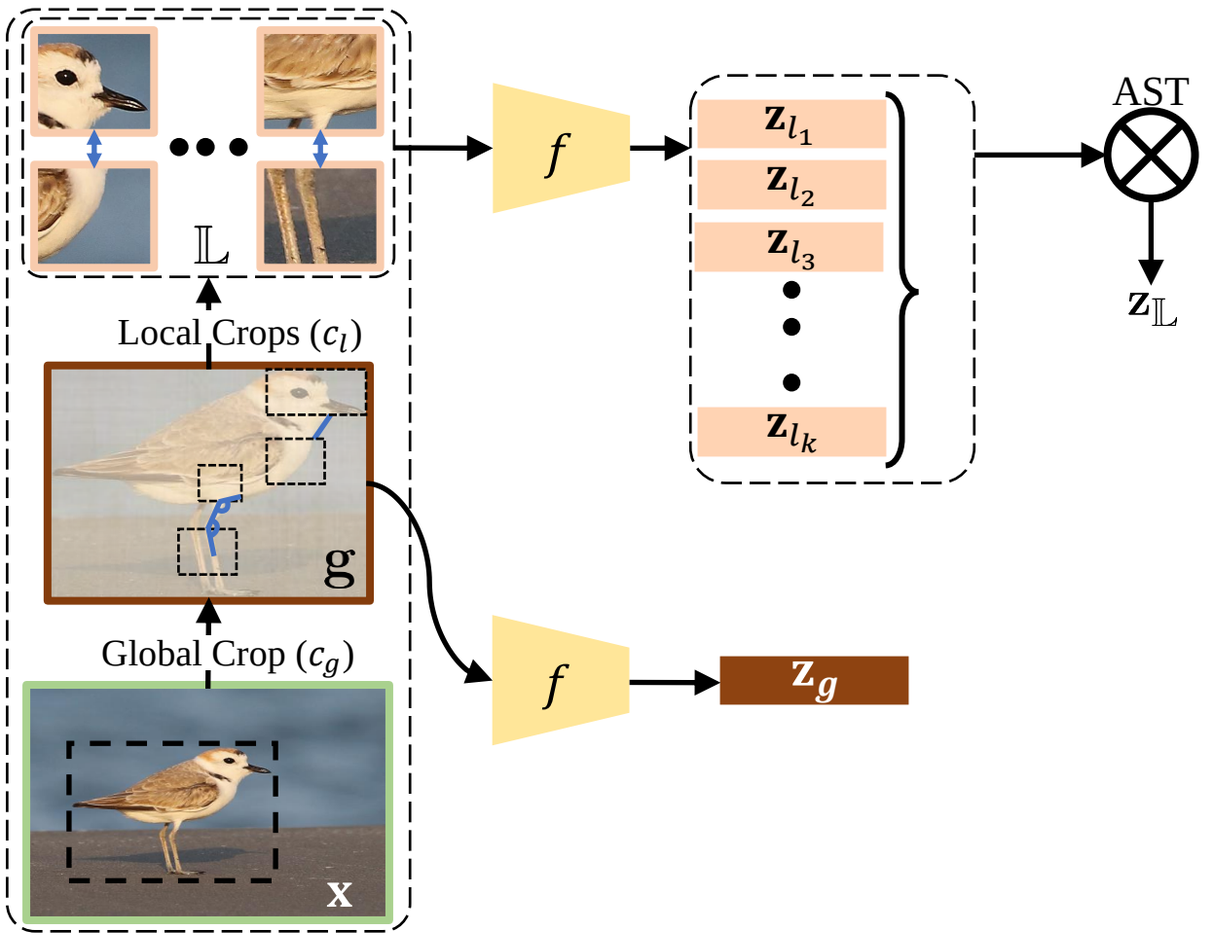
Lemma 3: Leads to misclassification in the fine-grained setting

Learning a Disjoint Representation Space



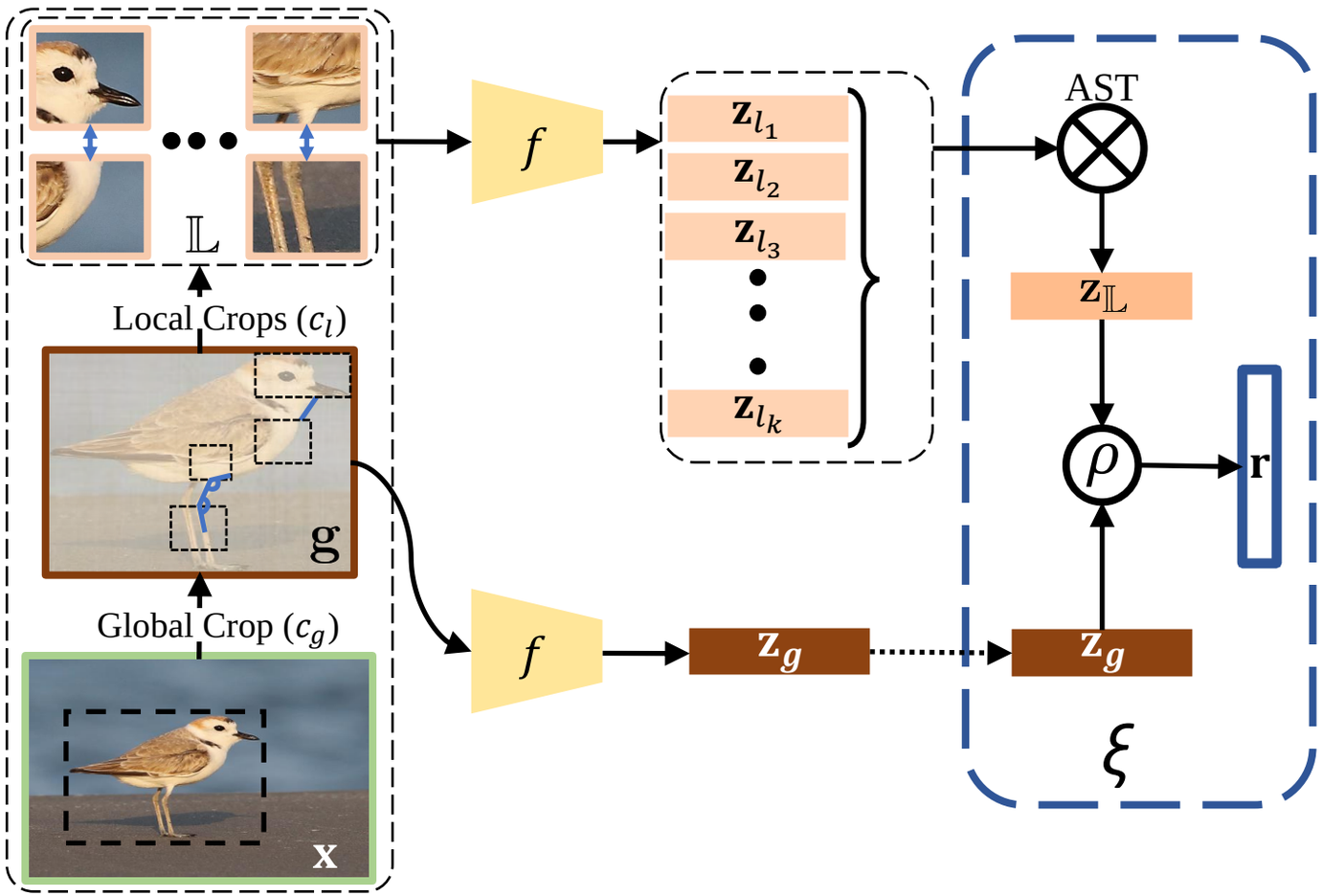


Summarizing Local Attributes



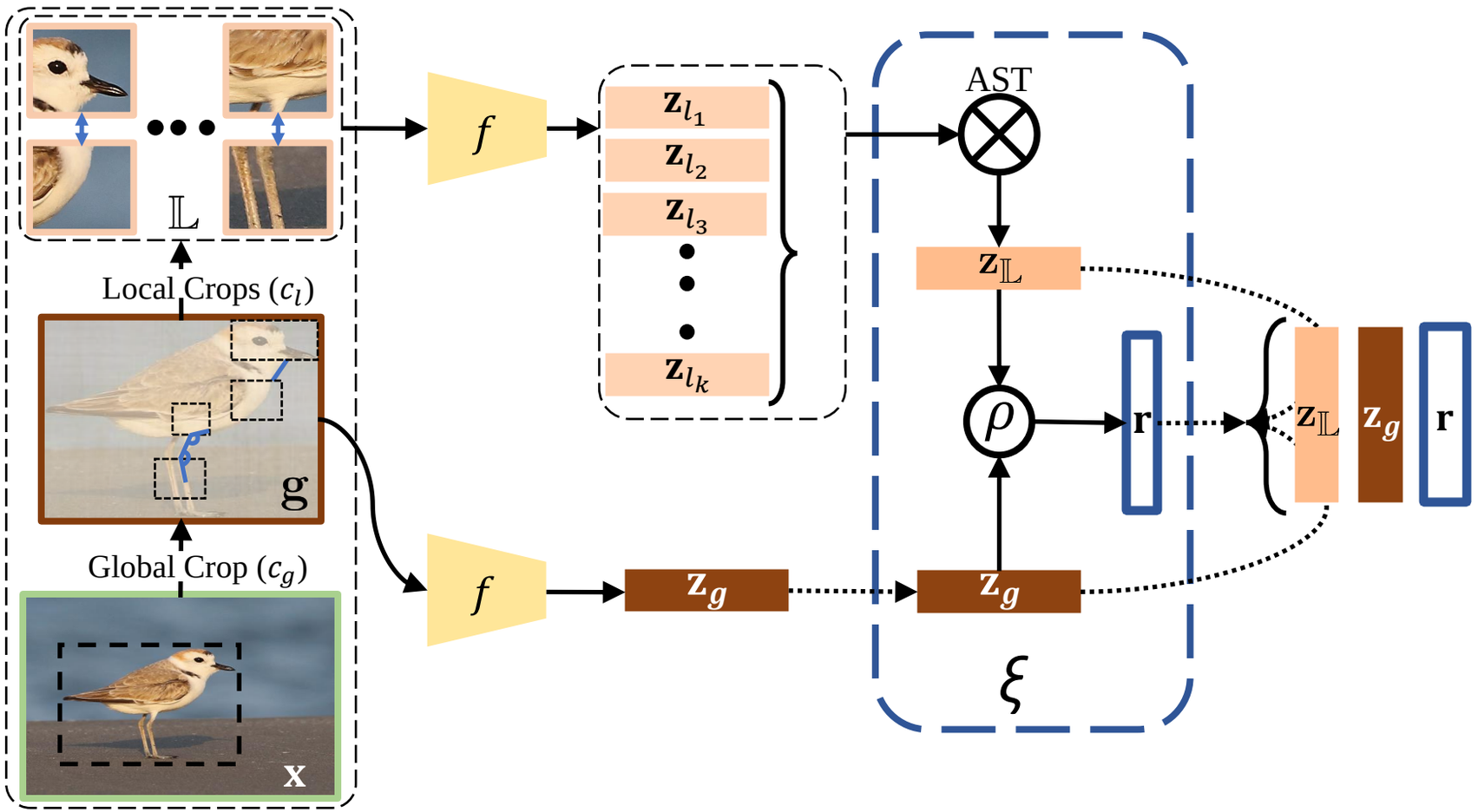
- Independent encoding of local and global views
- Right combination of local views that produce the global view.

Relationship Modelling

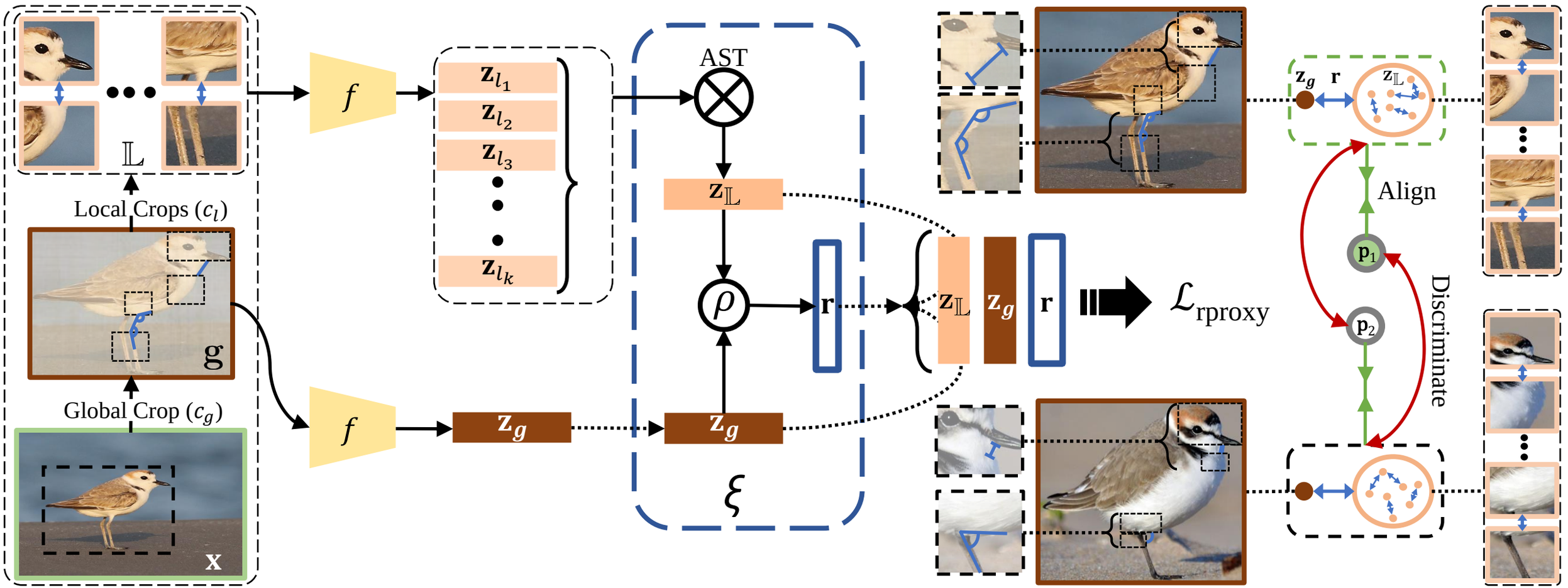


- Independent encoding of local and global views
- Right combination of local views that produce the global view.
- Learn the local-to-global emergent relationship

Sufficient Learner



Relational Proxies

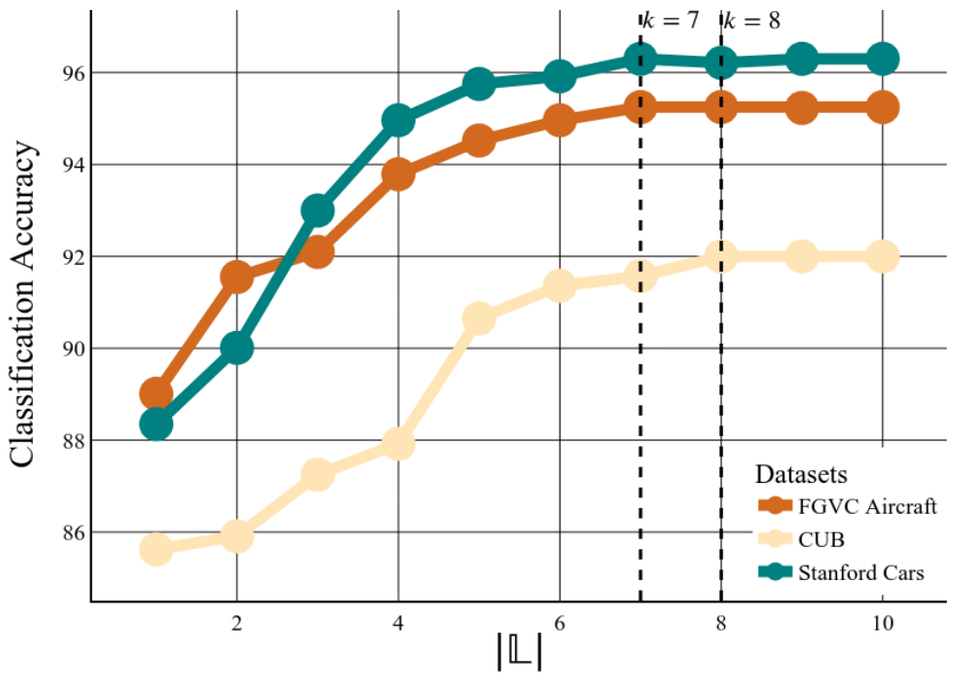


Comparison with State-of-the-Art

- We evaluate both on common **medium and large-scale benchmarks**, as well as the recently proposed, highly challenging **cultivar datasets**.
- We surpass SOTA on all of them, more significantly on the Cultivar datasets **by over 4%**.

Method	Benchmark					Cultivar	
	FGVC Aircraft	Stanford Cars	CUB	NA Birds	iNaturalist	Cotton	Soy
MaxEnt, NeurIPS'18	89.76	93.85	86.54	-	-	-	-
DBTNet, NeurIPS'19	91.60	94.50	88.10	-	-	-	-
StochNorm, NeurIPS'20	81.79	87.57	79.71	74.94	60.75	45.41	38.50
MMAL, MMM'21	94.70	95.00	89.60	87.10	69.85	65.00	47.00
FFVT, BMVC'21	79.80	91.25	91.65	89.42	70.30	57.92	44.17
CAP, AAAI'21	94.90	95.70	91.80	91.00	-	-	-
TransFG, AAAI'22	80.59	94.80	91.70	90.80	71.70	45.84	38.67
Ours (Relational Proxy)	**95.25** ± 0.02	**96.30** ± 0.04	**92.00** ± 0.01	**91.20** ± 0.02	**72.15** ± 0.03	**69.81** ± 0.04	**51.20** ± 0.02

Studies on Local-Views



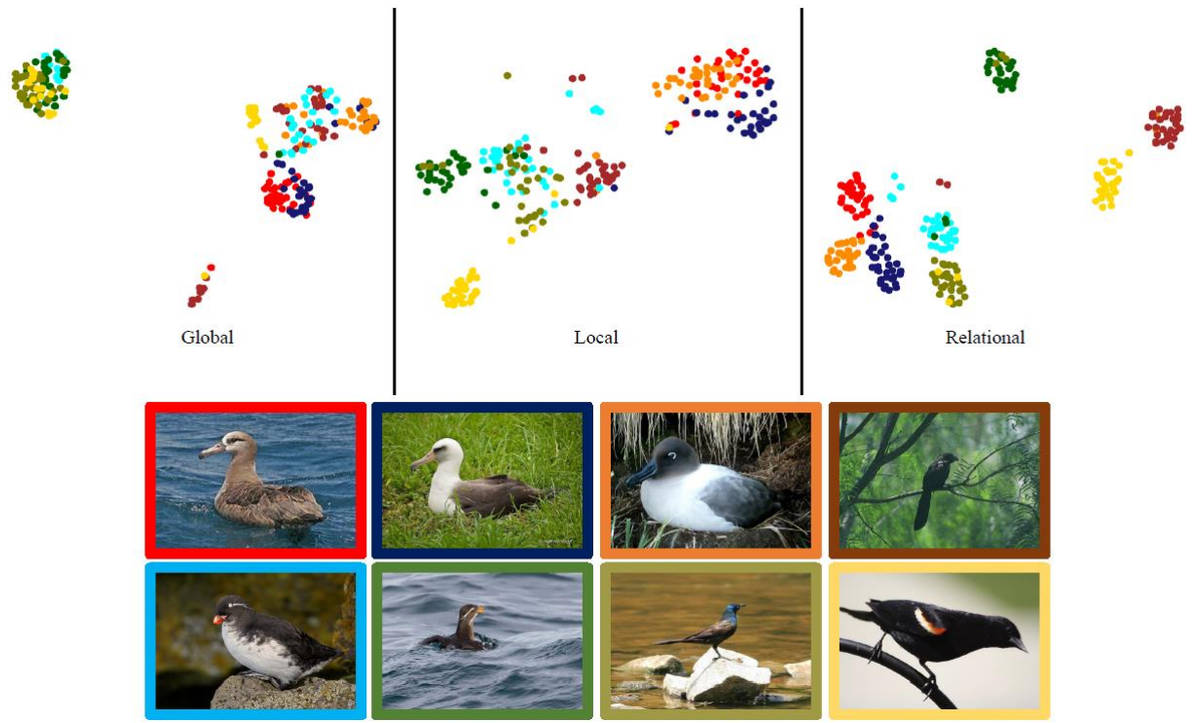
Effect of varying the number of local views

$\lvert\mathbb{L}\rvert$	N/5	N/4	N/3	N/2
7	-0.03	-0.02	0.00	-0.14
12	+0.01	+0.02	0.00	-0.11
15	+0.05	+0.03	+0.01	-0.11
18	+0.05	+0.03	+0.00	-0.10

Correlation between number of local views and size of local patches

Global *vs* Local *vs* Relational Embeddings

- Relational Embeddings provide the best class-wise separation among all.



Contributions

- Theoretical formalization of the FGVC problem and algorithms.
- Derivation of necessary and sufficient conditions for FGVC.
- An algorithm for implementing a *sufficient* learner based on local-to-global relationships.
- Experiments demonstrating the benefits of using the relational embeddings from a *sufficient* learner for FGVC.

Contributions

- Theoretical formalization of the FGVC problem and algorithms.
- Derivation of necessary and sufficient conditions for FGVC.
- An algorithm for implementing a *sufficient* learner based on local-to-global relationships.
- Experiments demonstrating the benefits of using the relational embeddings from a *sufficient* learner for FGVC.

Contributions

- Theoretical formalization of the FGVC problem and algorithms.
- Derivation of necessary and sufficient conditions for FGVC.
- An algorithm for implementing a *sufficient* learner based on local-to-global relationships.
- Experiments demonstrating the benefits of using the relational embeddings from a *sufficient* learner for FGVC.

Contributions

- Theoretical formalization of the FGVC problem and algorithms.
- Derivation of necessary and sufficient conditions for FGVC.
- An algorithm for implementing a *sufficient* learner based on local-to-global relationships.
- Experiments demonstrating the benefits of using the relational embeddings from a *sufficient* learner for FGVC.

Relational Proxies: Emergent Relationships as Fine-Grained Discriminators

Get in touch:
Abhra Chaudhuri
ac1151@exeter.ac.uk

https://arxiv.org/abs/2210.02149

https://github.com/abhrac/relational-proxies