NEURAL INFORMATION PROCESSING SYSTEMS
2023

Transitivity Recovering Decompositions: Interpretable and Robust Fine-Grained Relationships

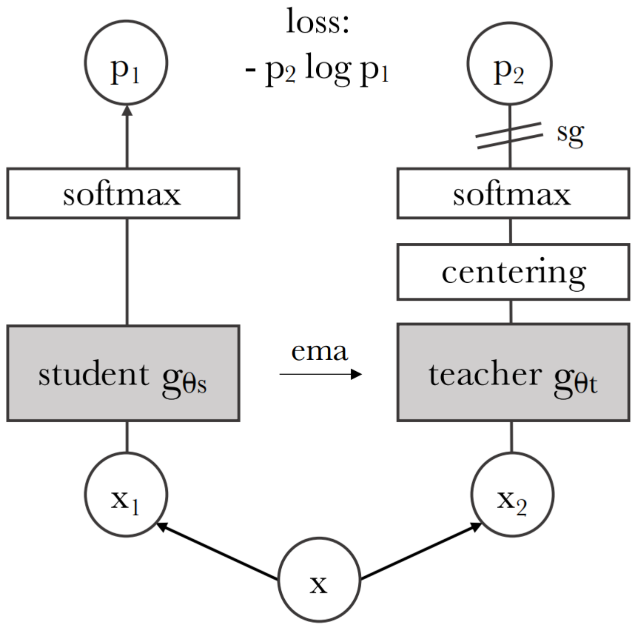



Abhra Chaudhuri | Massimiliano Mancini | Zeynep Akata | Anjan Dutta

University of Exeter | Università di Trento | Eberhard Karls Universität Tübingen | University of Surrey

Relationships Encode Fine-Grained Semantics

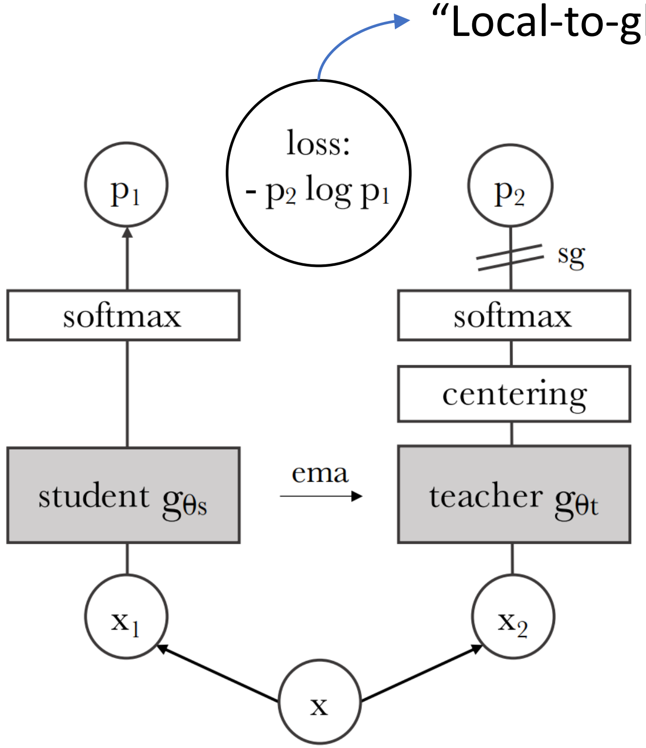
Relationships Encode Fine-Grained Semantics



Caron *et al*. "Emerging Properties in Self-Supervised Vision Transformers" (aka, DINO), ICCV 2021.

Relationships Encode Fine-Grained Semantics

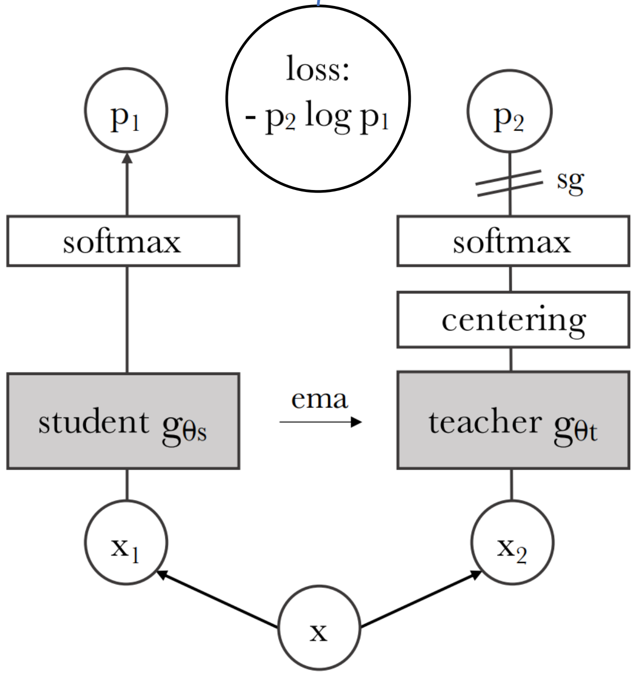
“Local-to-global” (emergent) correspondences



Caron *et al*. “Emerging Properties in Self-Supervised Vision Transformers” (aka, DINO), ICCV 2021.

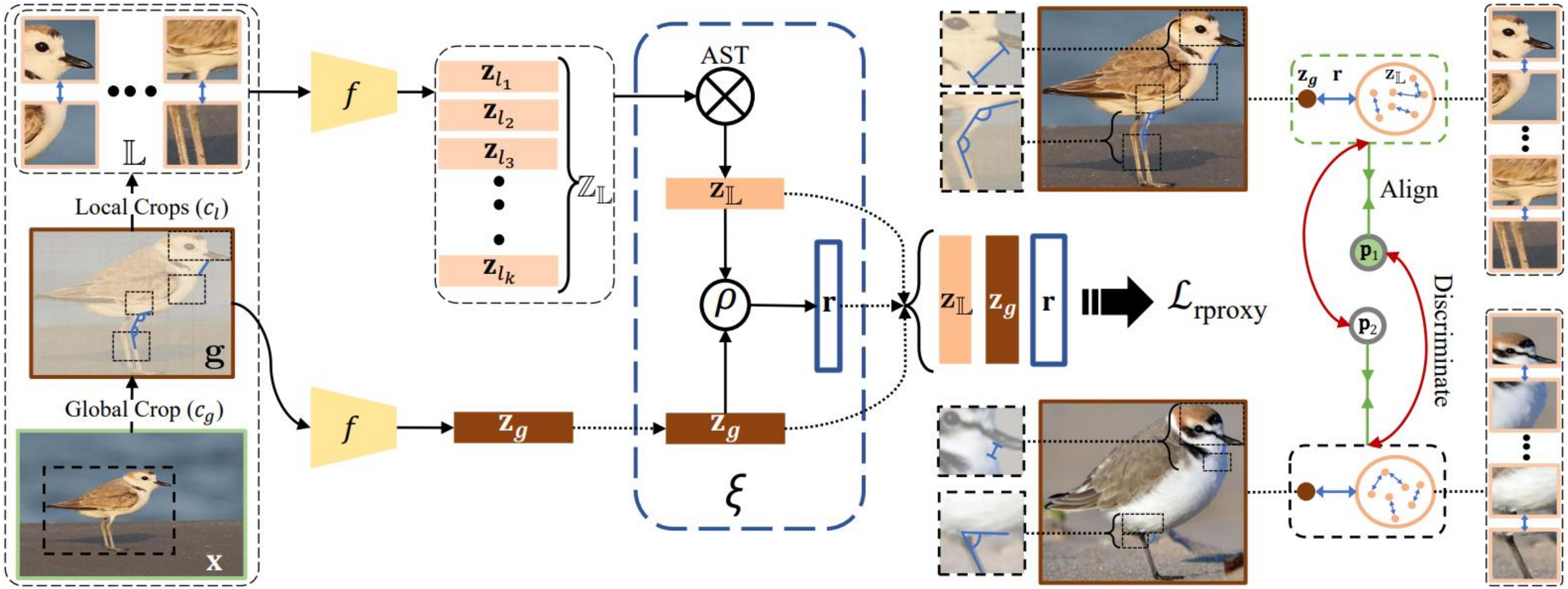
Relationships Encode Fine-Grained Semantics

"Local-to-global" (emergent) correspondences



Caron *et al*. "Emerging Properties in Self-Supervised Vision Transformers" (aka, DINO), ICCV 2021.

Relationships Encode Fine-Grained Semantics

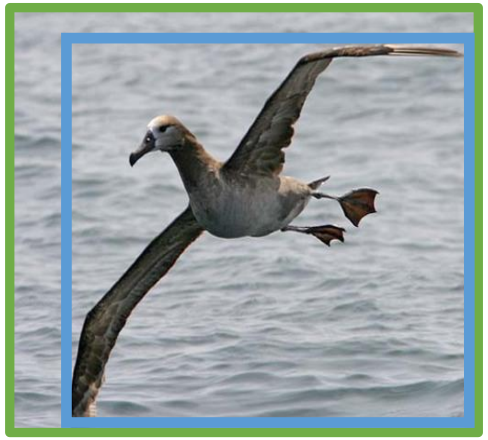


Chaudhuri *et al*. “Relational Proxies: Emergent Relationships as Fine-Grained Discriminators”, NeurIPS 2022.

Global and Local Views

Global and Local Views

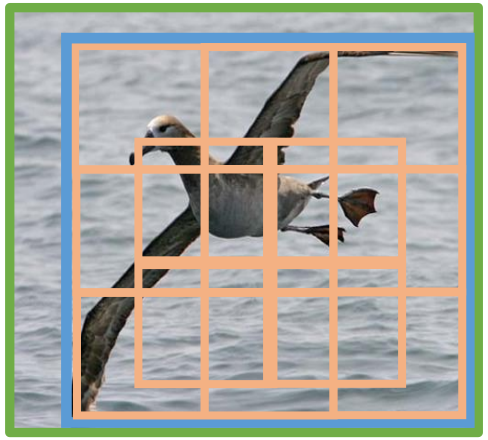
- Global view – Most salient region in the image based on ResNet50 feature maps.



Global View

Global and Local Views

- Local views – Random crops within the global view.

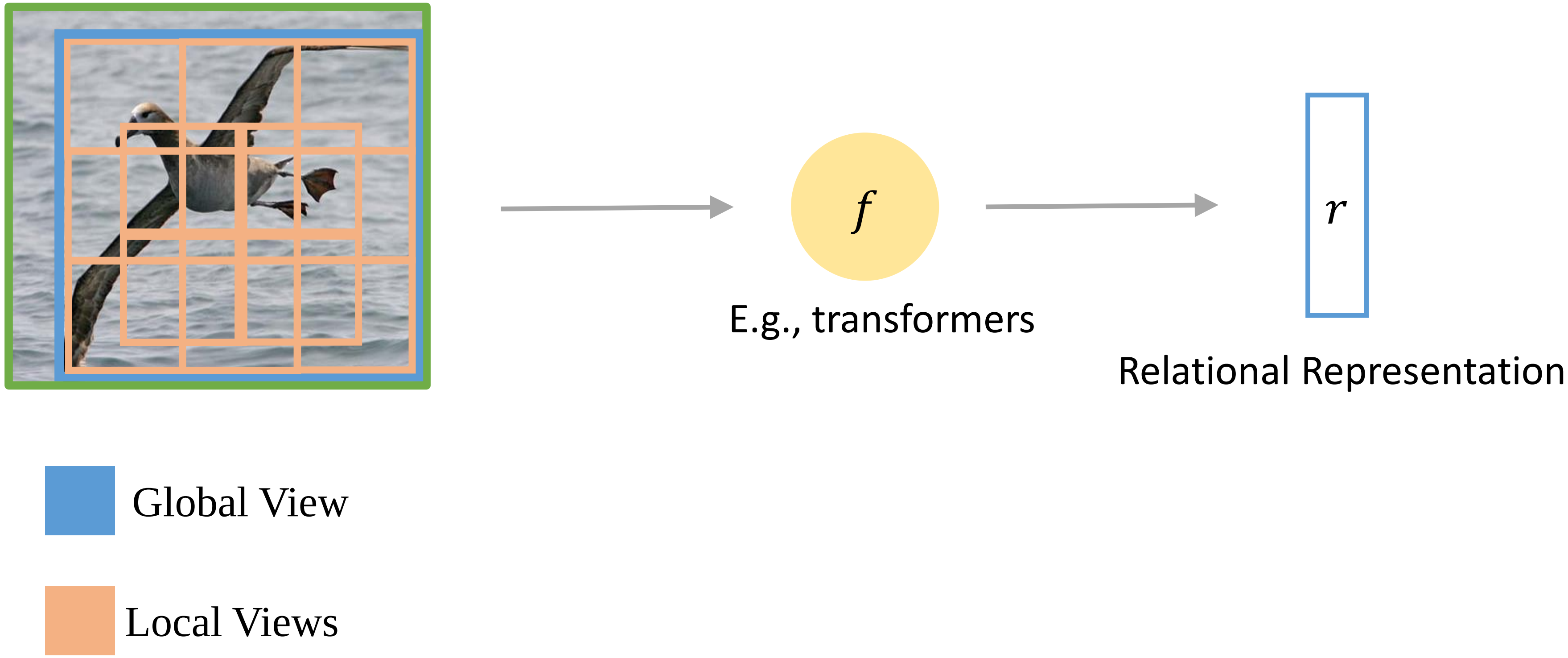


Global View

Local Views

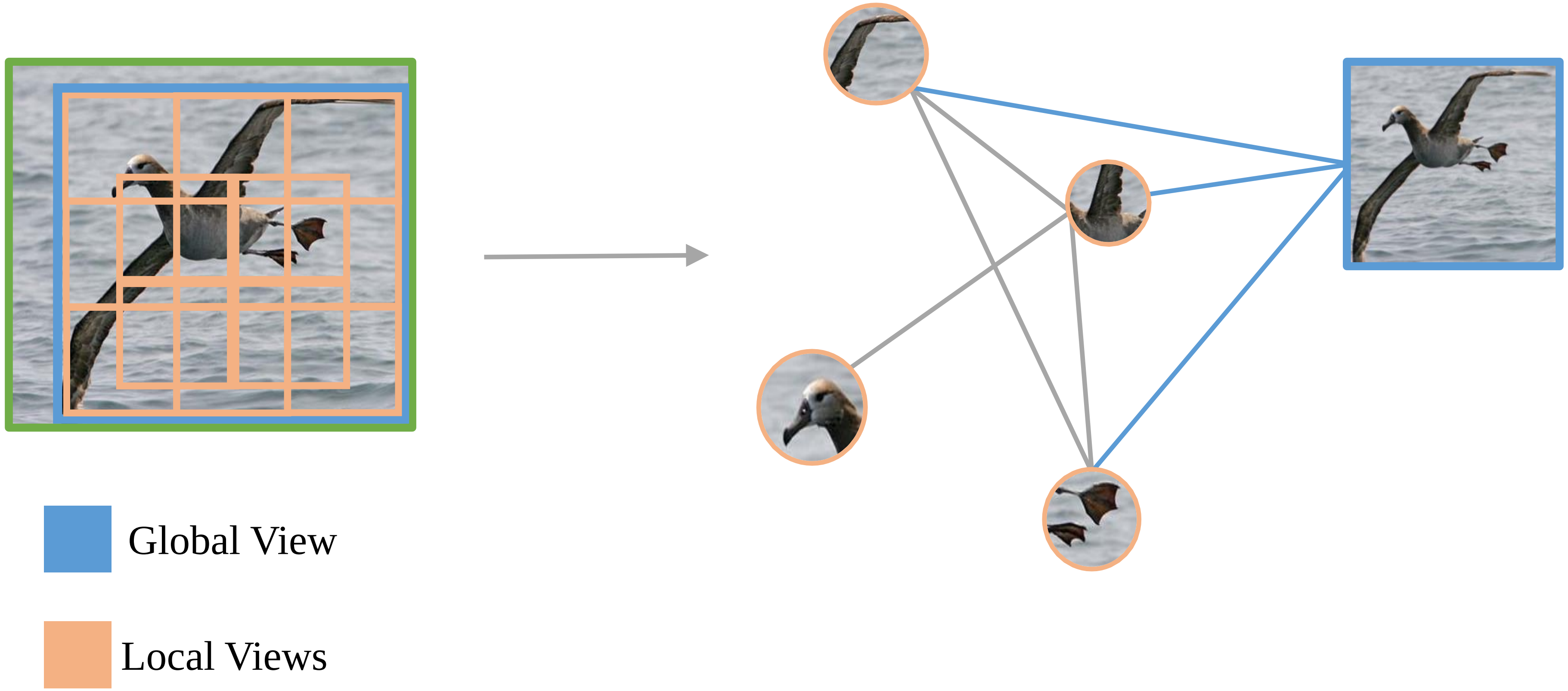
Existing Works – Abstract Aggregation

- Existing works abstractly summarize emergent (local-to-global relationships) into a single n-dimensional vector.



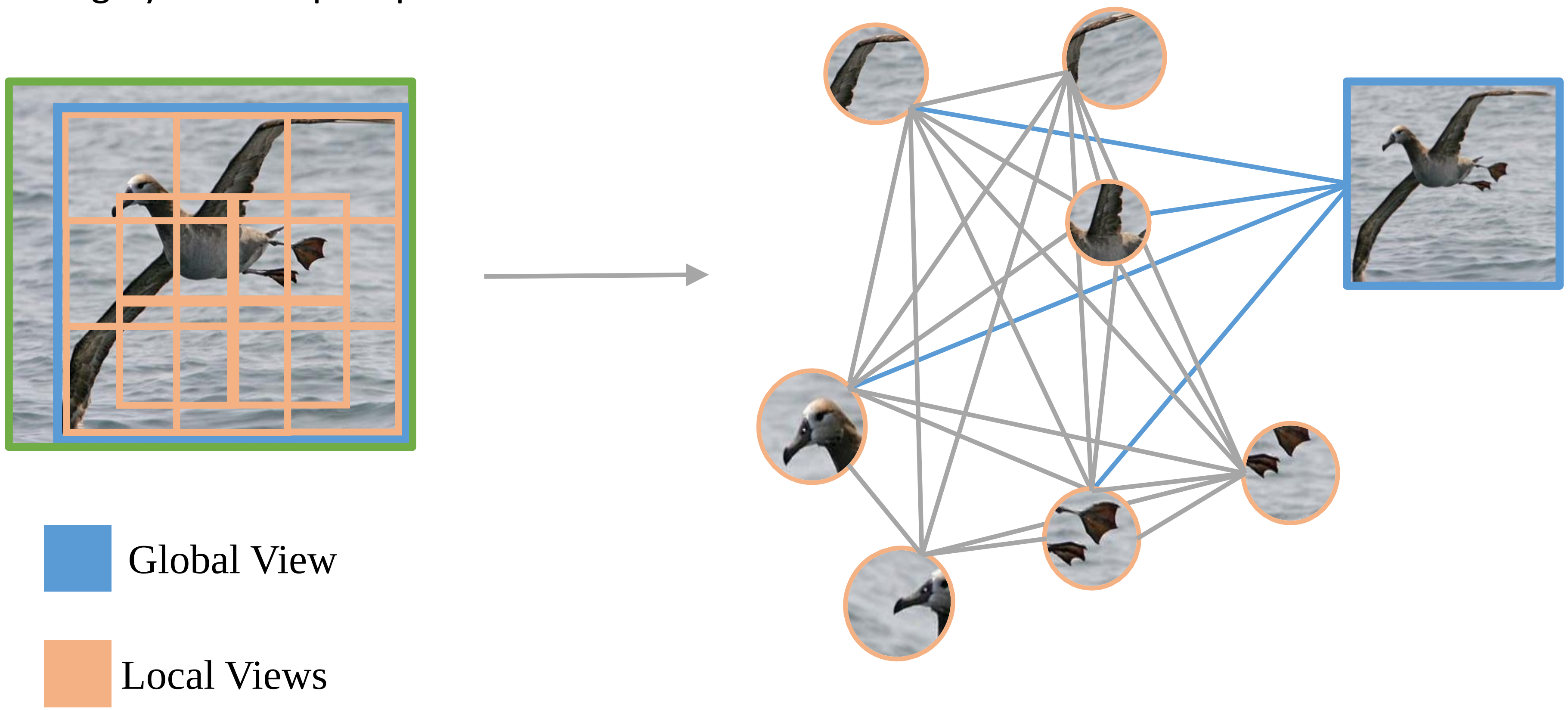
Motivation – Getting Rid of the Abstraction

- Produce graphs as interpretable alternatives to such abstract relational representations.



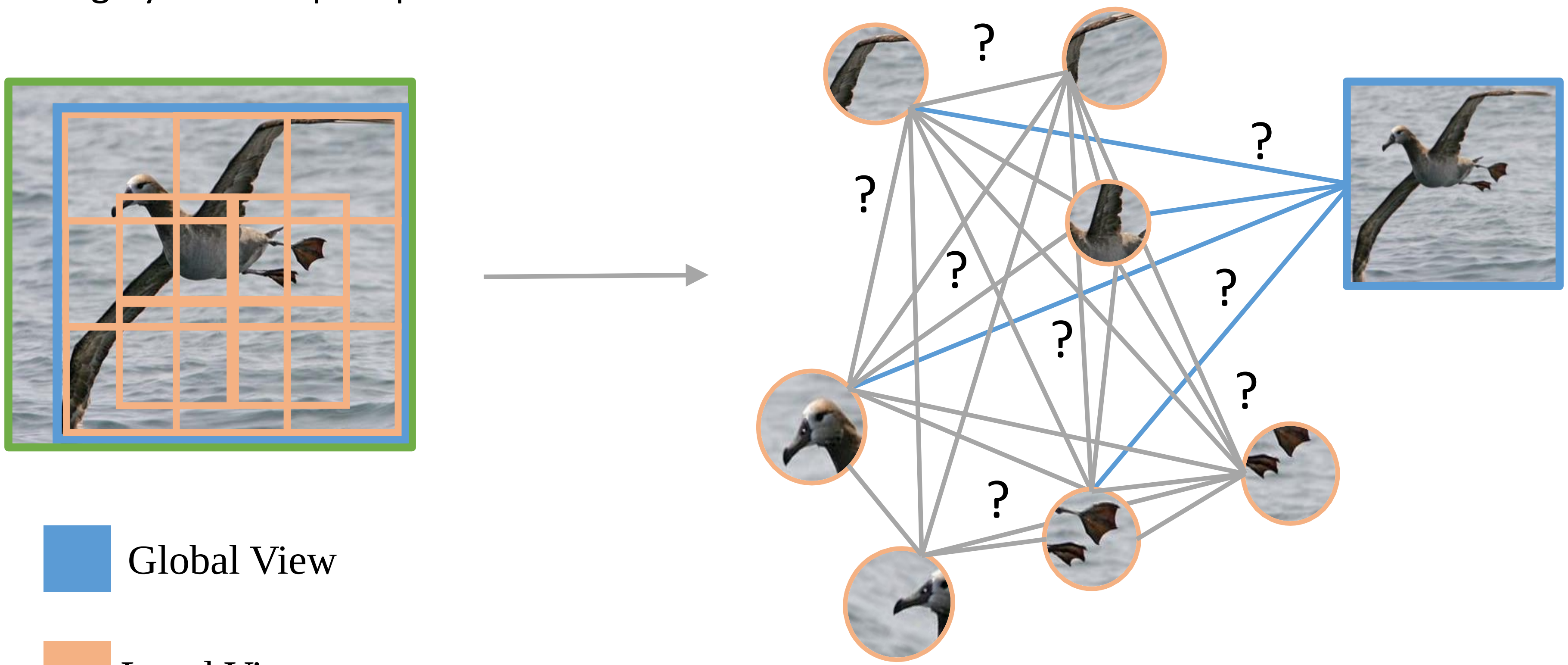
No Inductive Bias – Complete Graph

- Highly dense input space.



No Inductive Bias – Complete Graph

- Highly dense input space.



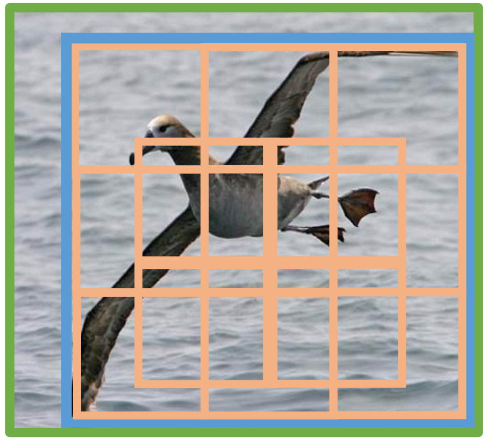
Global View

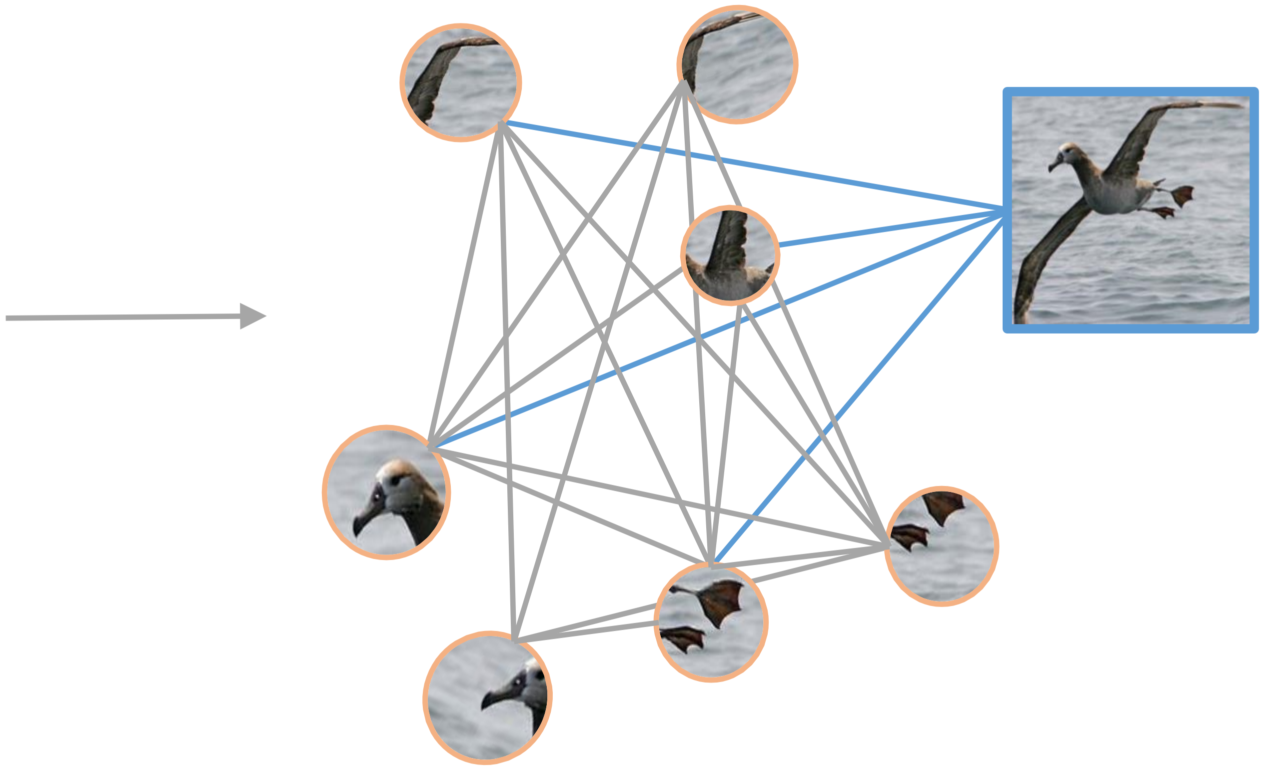
Local Views

Which subgraph is encoded in the abstract representation r?

Inductive Bias – Complementarity Graph

- Complementarity leads to sparser input graphs.



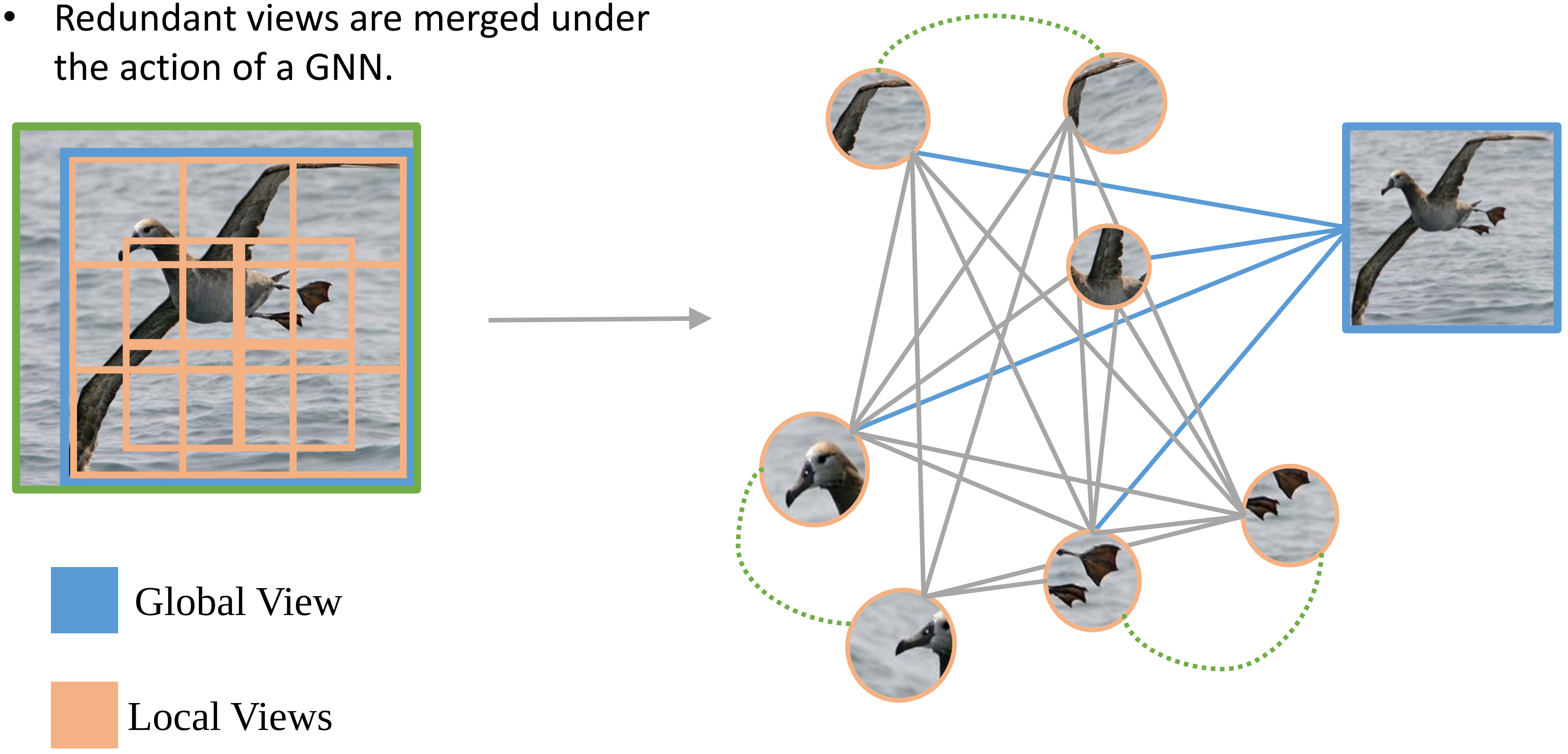


Global View

Local Views

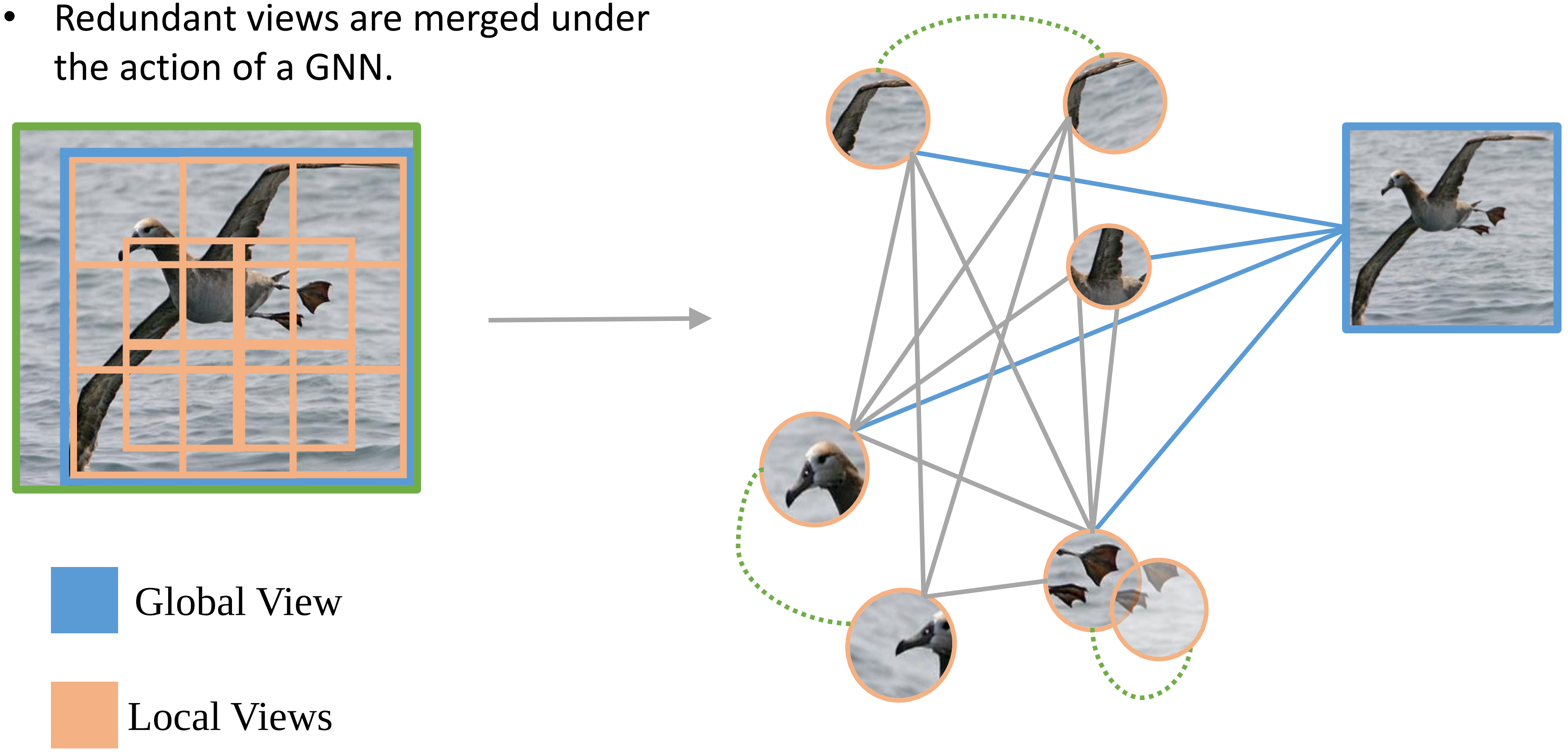
Inductive Bias – Complementarity Graph

- Redundant views are merged under the action of a GNN.



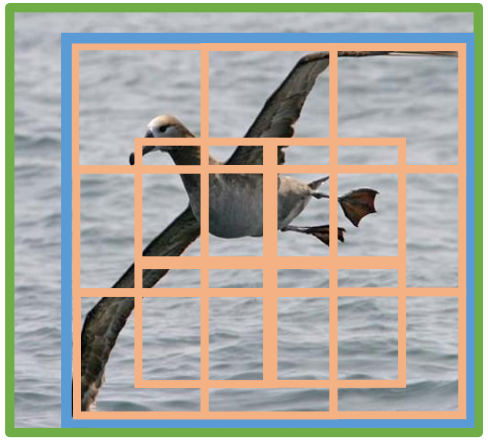
Inductive Bias – Complementarity Graph

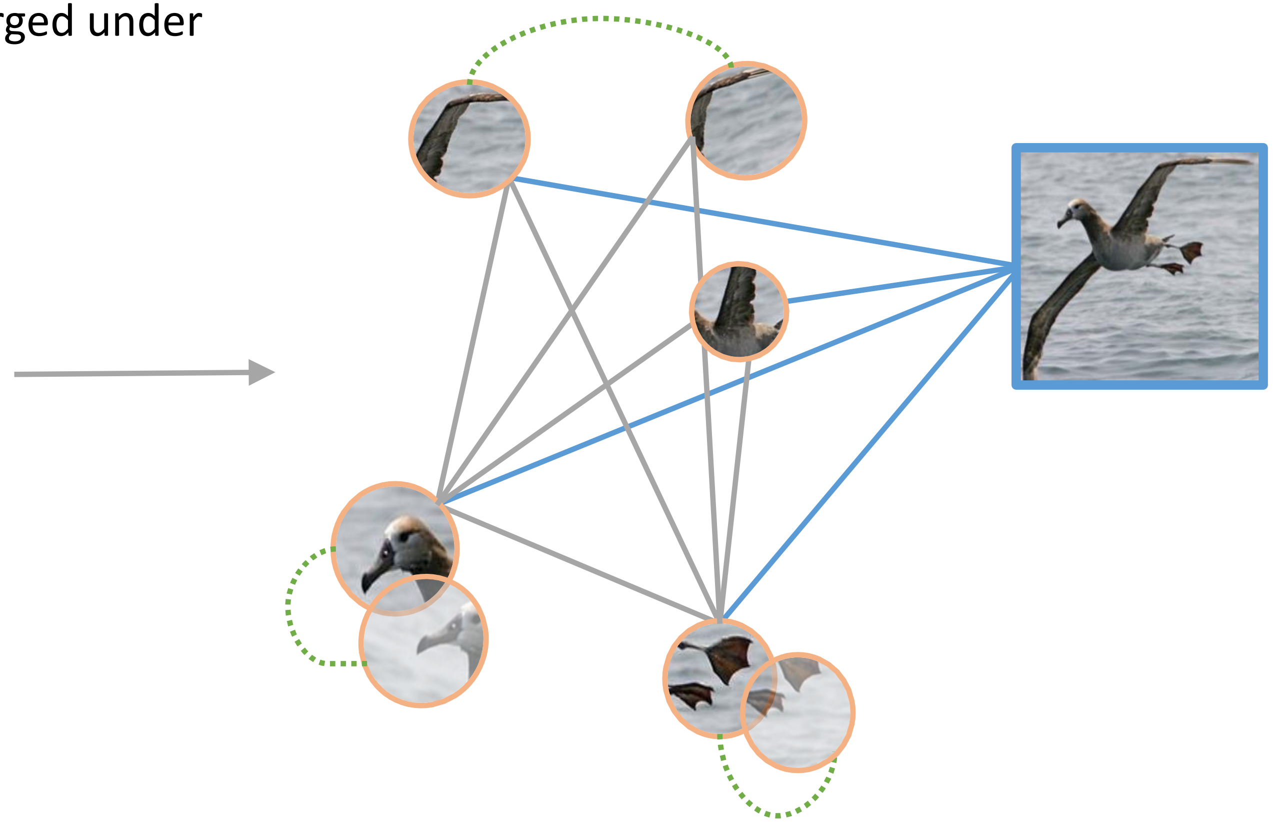
- Redundant views are merged under the action of a GNN.



Inductive Bias – Complementarity Graph

- Redundant views are merged under the action of a GNN.



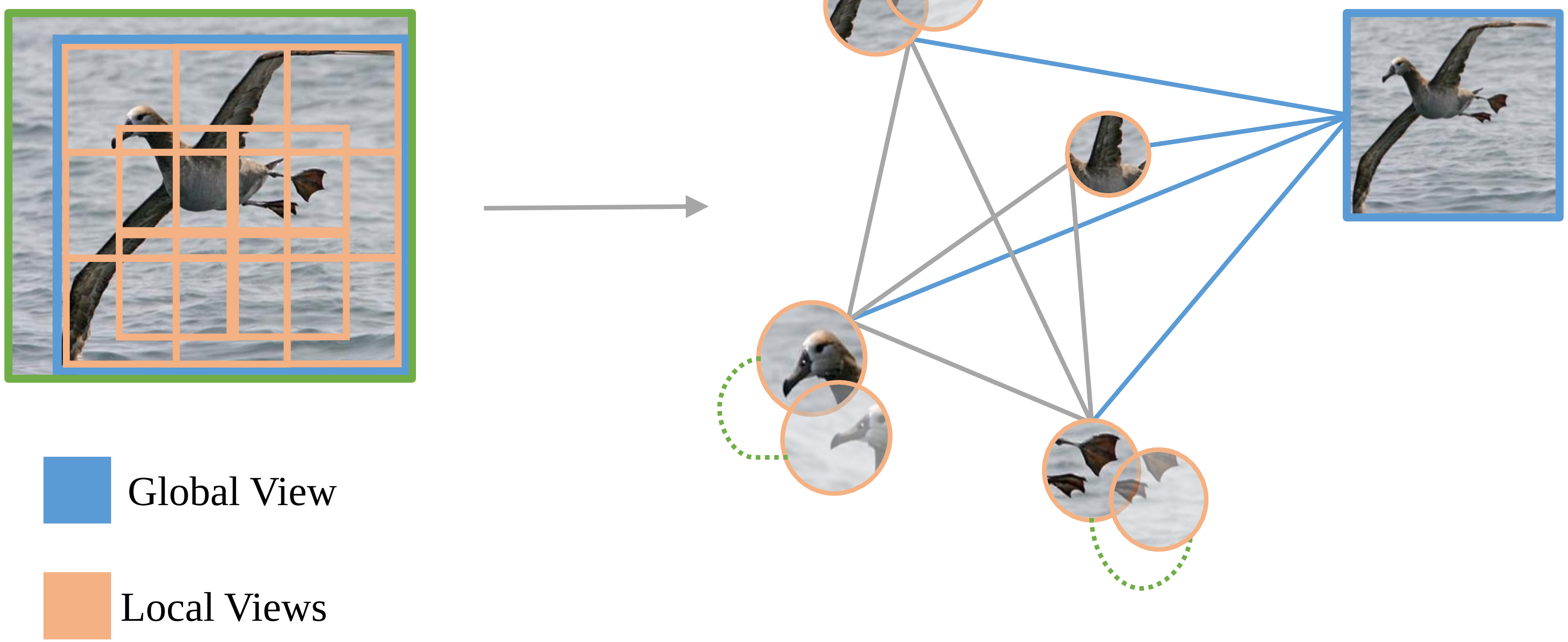


Global View

Local Views

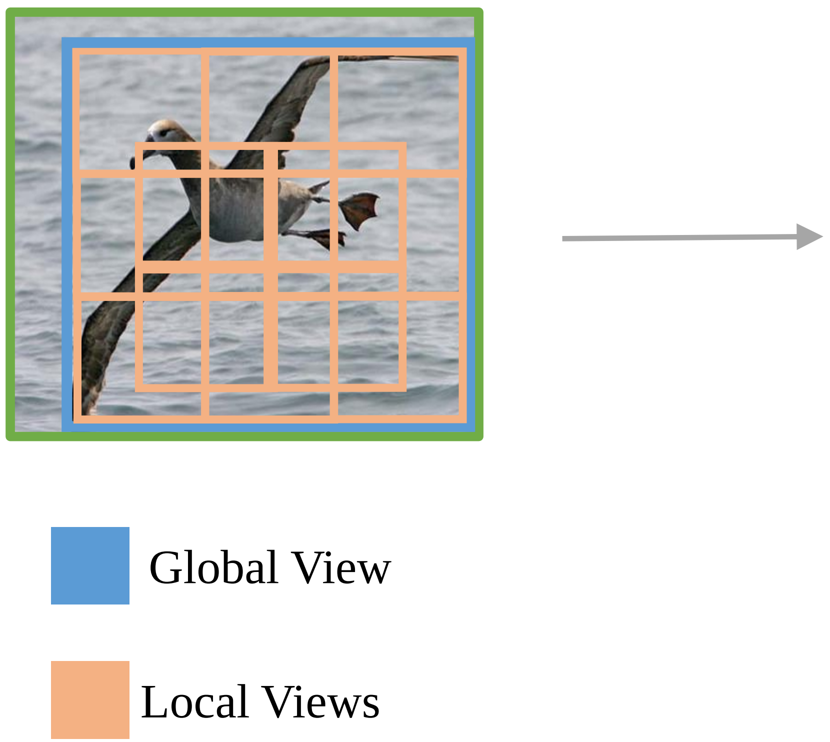
Inductive Bias – Complementarity Graph

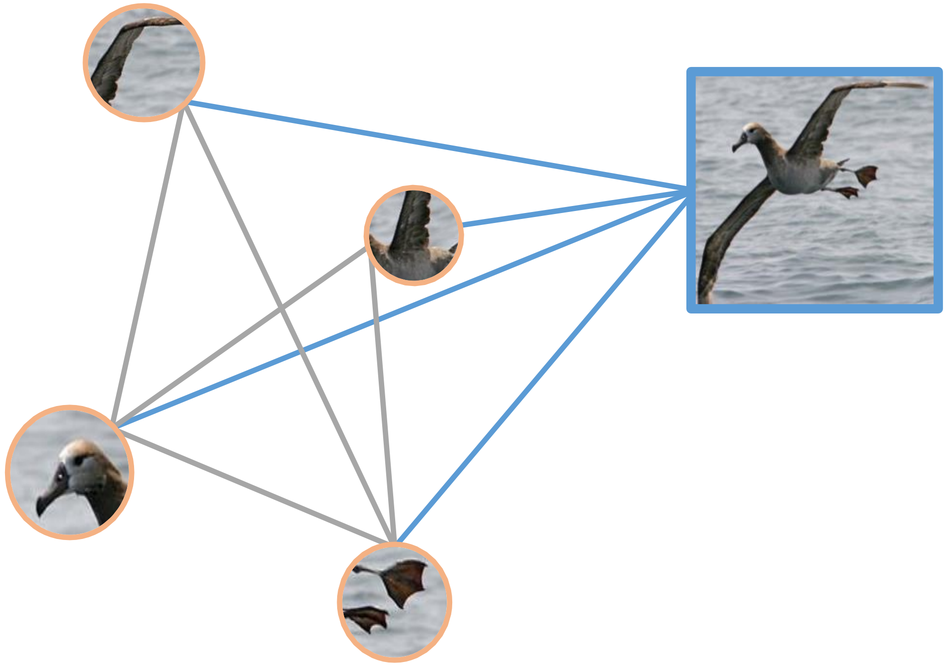
- Redundant views are merged under the action of a GNN.



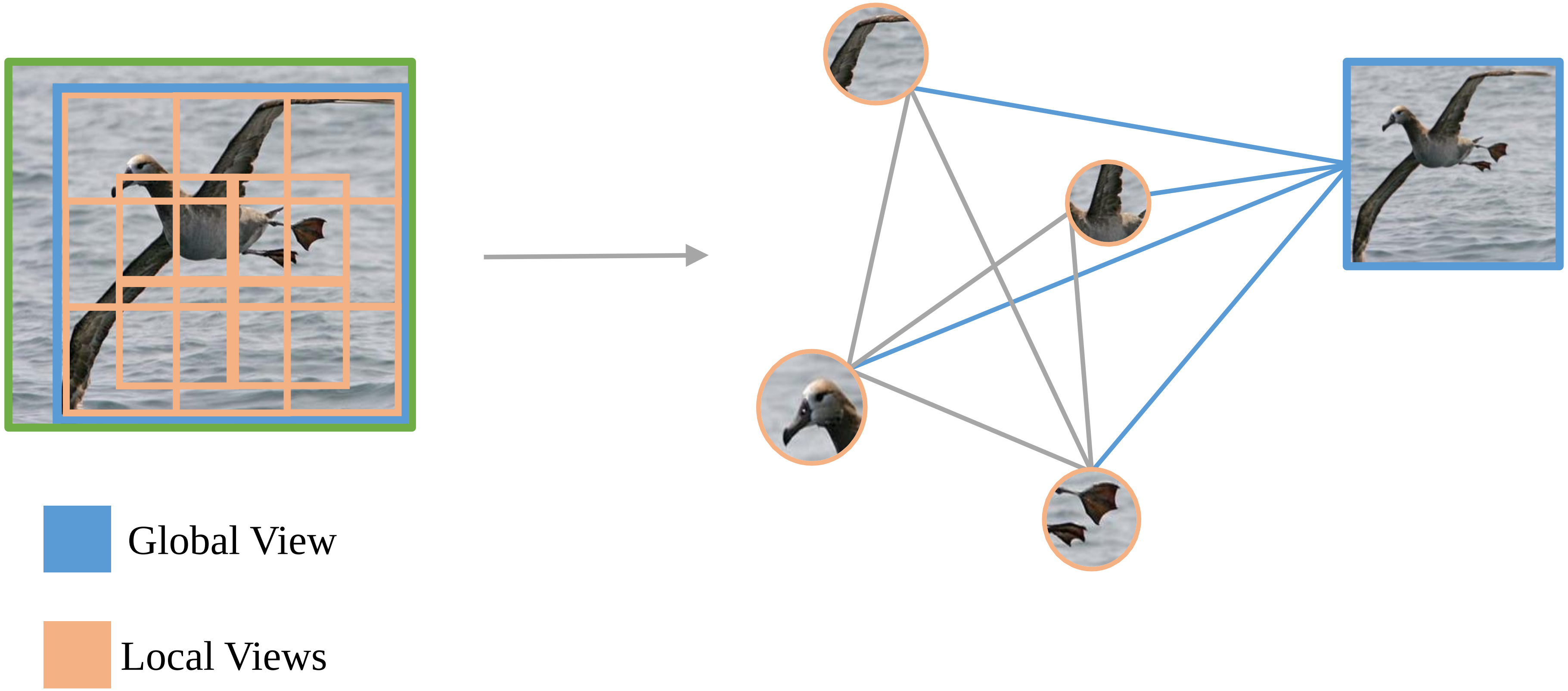
Inductive Bias – Complementarity Graph

- Redundant views are merged under the action of a GNN.

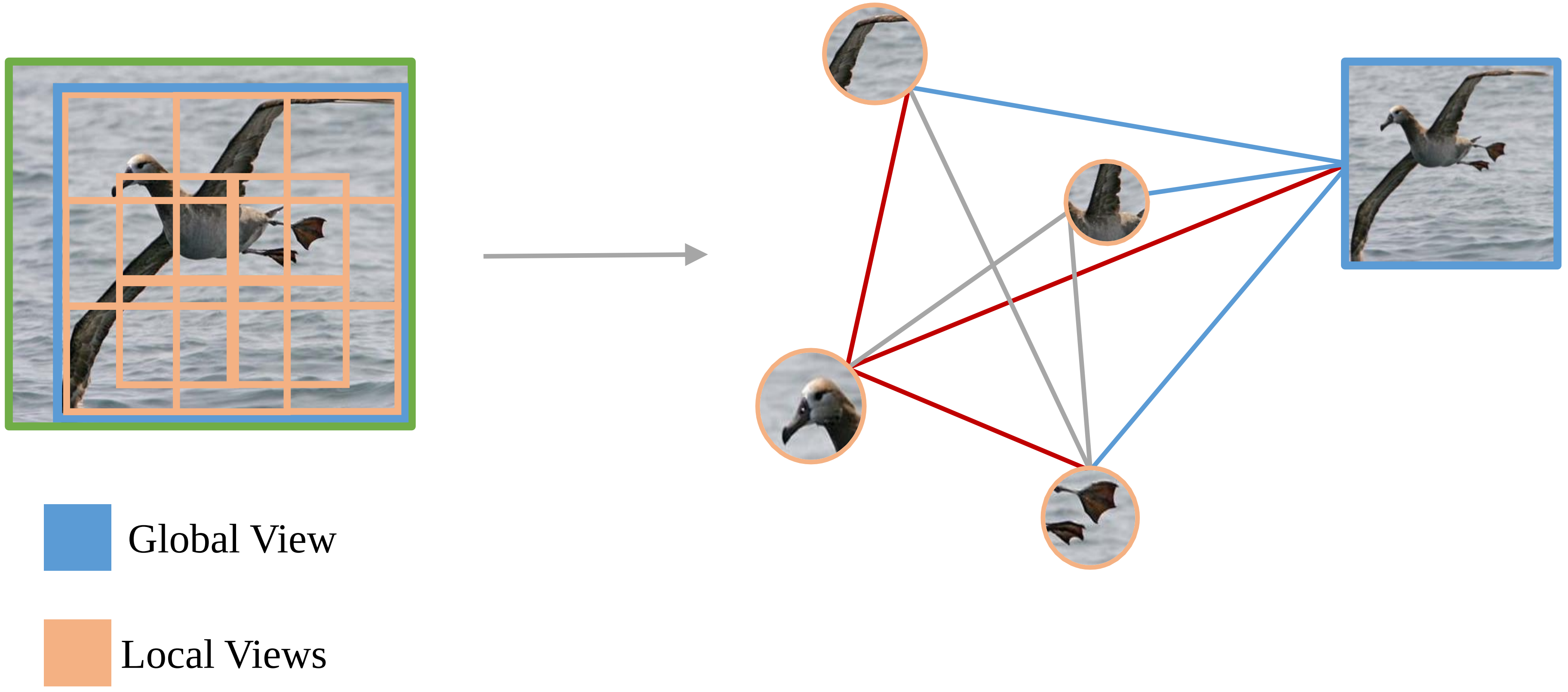




What Else Remains?

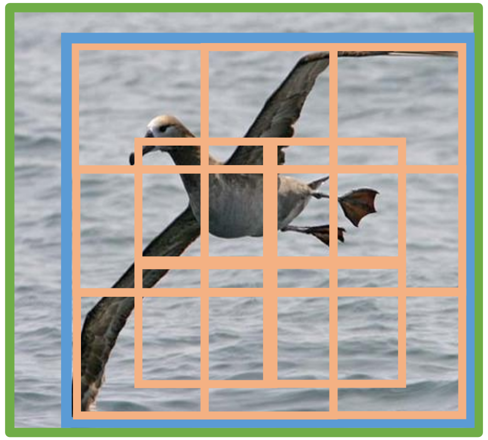


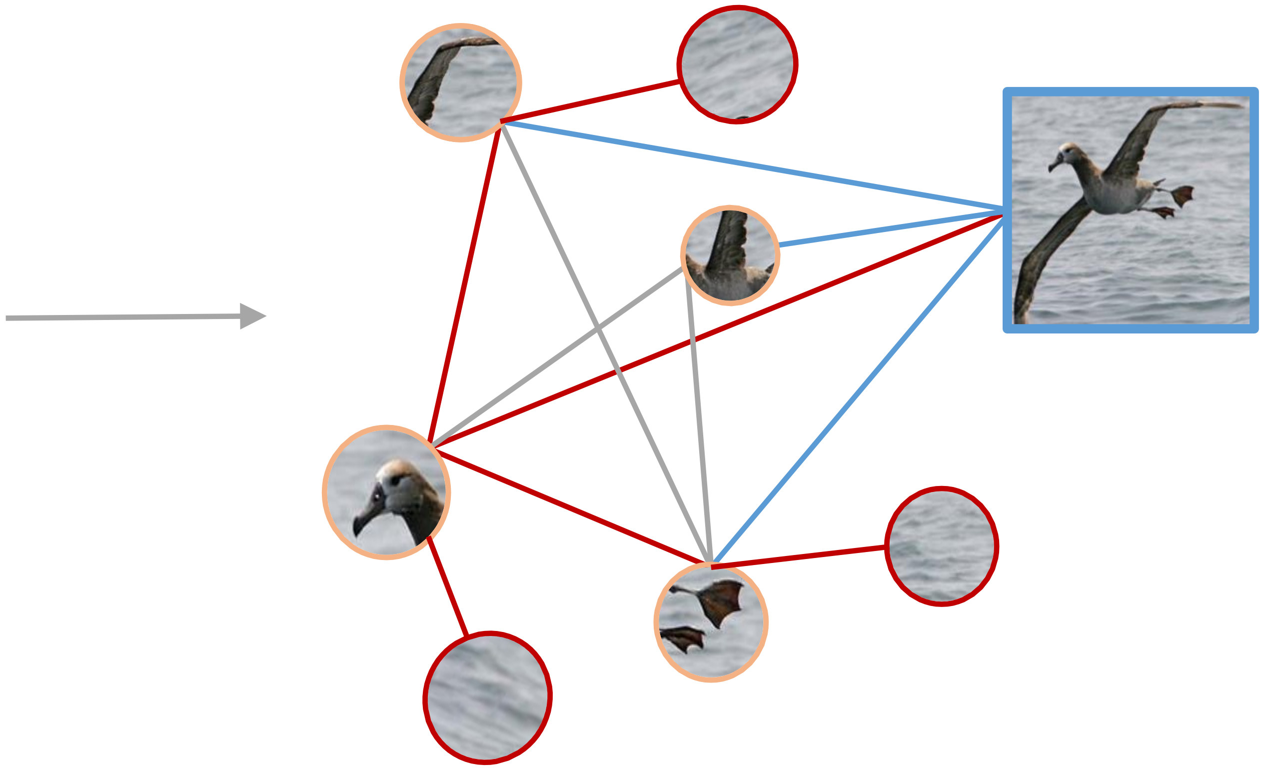
What Else Remains?



- Unnecessary edges.

What Else Remains?





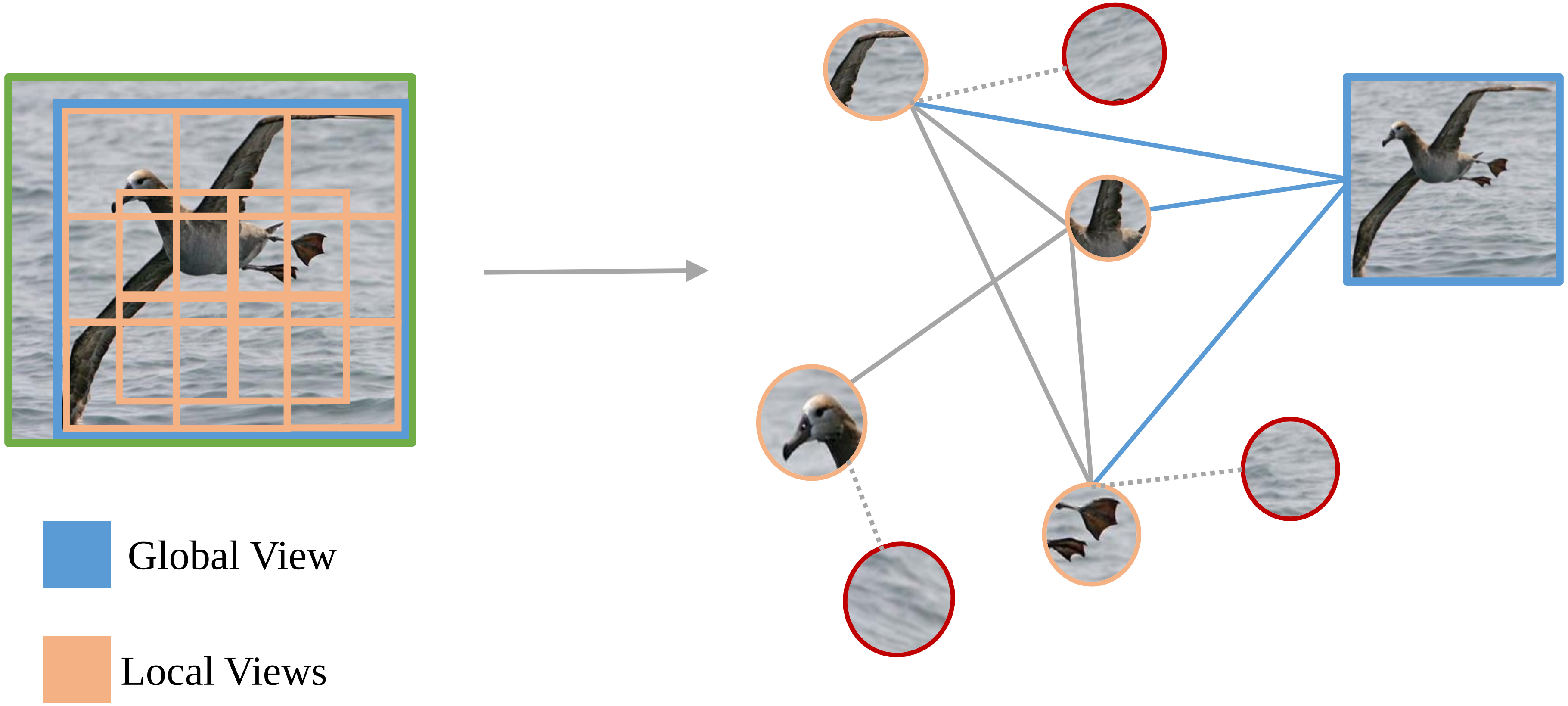
Global View

Local Views

- Unnecessary edges.
- Noisy views.

Transitivity Recovery Achieves Both

- Recovering ***transitive relationships*** that co-occur at instance and class-levels can effectively remove both.

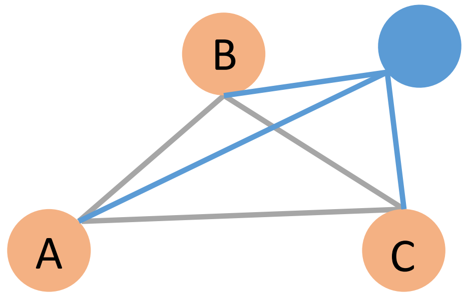


✓ Unnecessary edges.
✓ Noisy views.

- Sufficiency
- Transparency
- Robustness

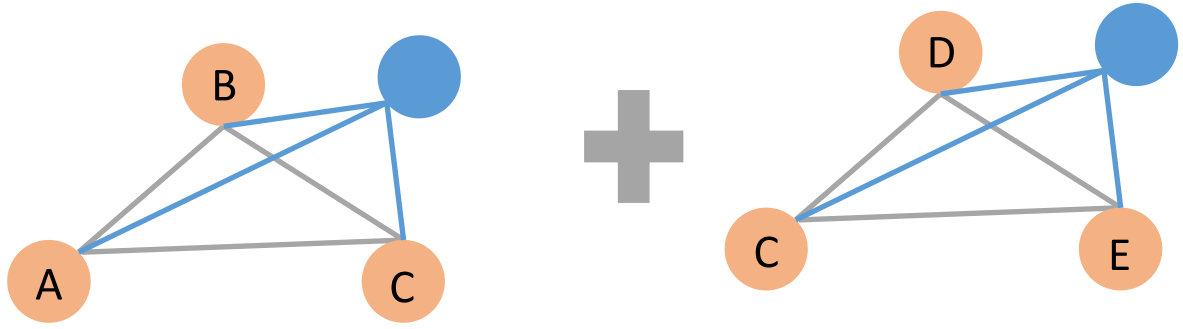
Transitive Relationships

Instance



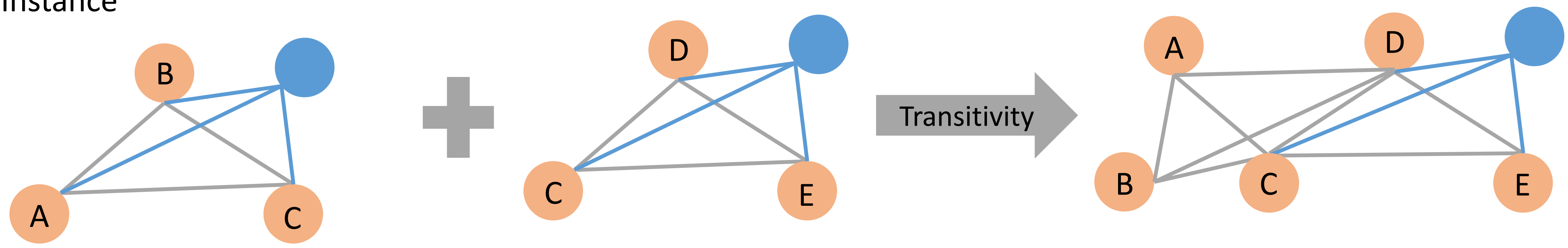
Transitive Relationships

Instance



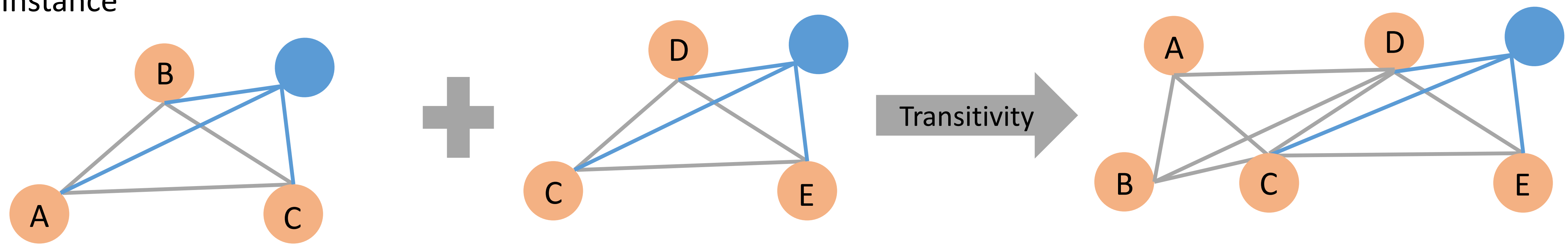
Transitive Relationships

Instance

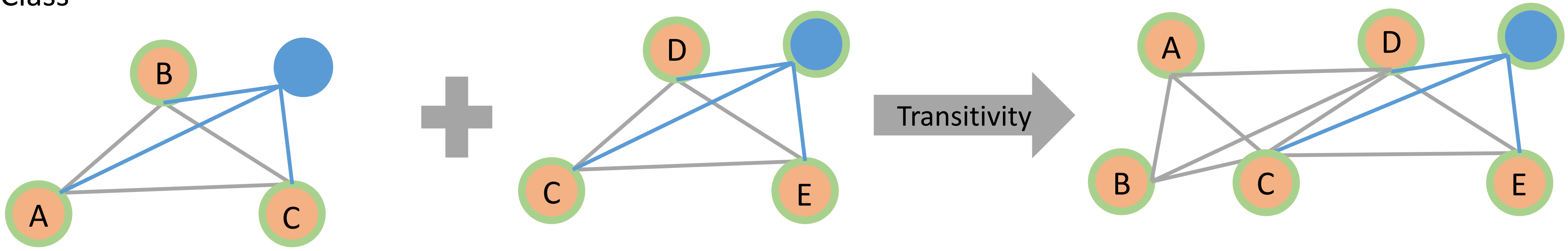


Transitive Relationships

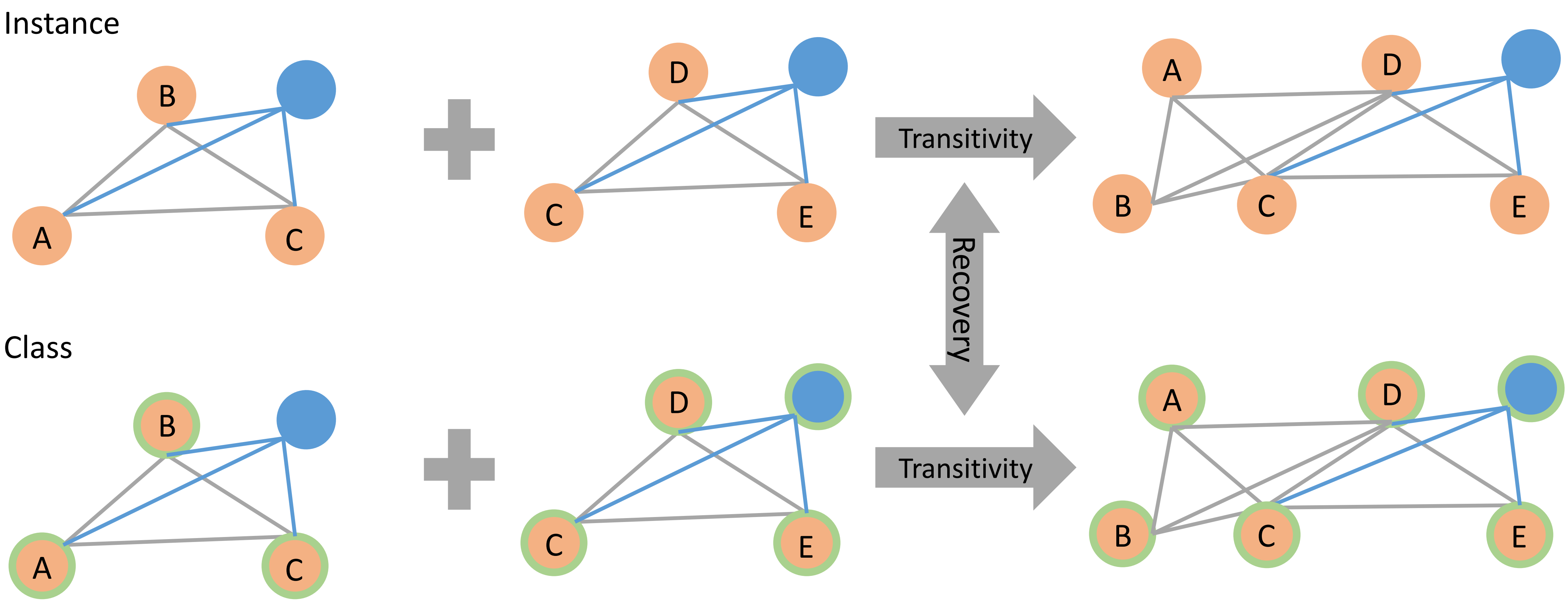
Instance



Class



Recovering Transitive Relationships



Practically Implementing Transitivity Recovery

Requires:

- Representing classes as graphs

Practically Implementing Transitivity Recovery

Requires:

- Representing classes as graphs – online clustering of local view and learnable edge embeddings.

Practically Implementing Transitivity Recovery

Requires:

- Representing classes as graphs – online clustering of local view and learnable edge embeddings.
- Realtime matching of instance and class graphs

Practically Implementing Transitivity Recovery

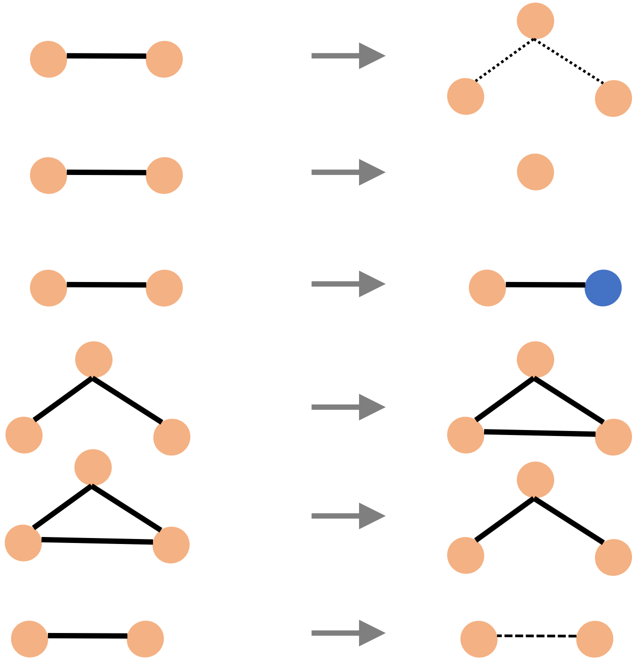
Requires:

- Representing classes as graphs – online clustering of local view and learnable edge embeddings.
- Realtime matching of instance and class graphs – minimizing the Hausdorff Edit Distance between instance and the class proxy graphs.

Graph Kernel: Hausdorff Distance

Graph Edit Distance – Expressed in terms of:

- Node Insertion
- Node Deletion
- **Node Substitution**
- Edge Insertion
- Edge Deletion
- Edge Substitution



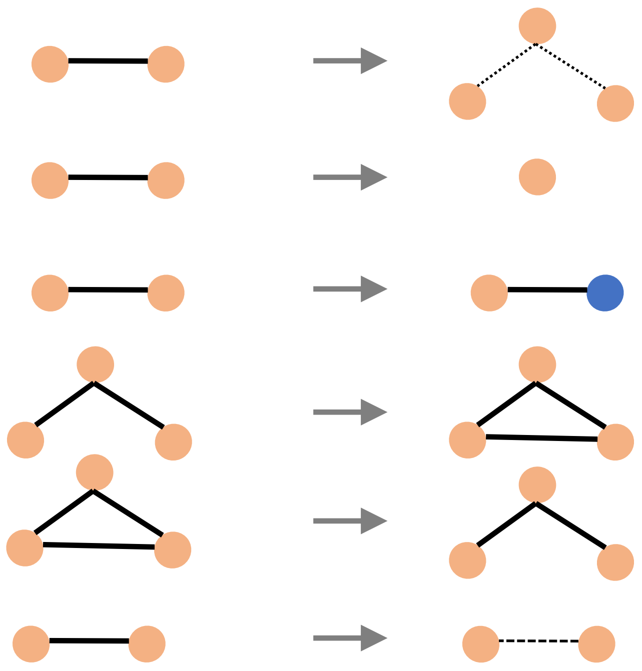
Lower bound approximation of Graph Edit Distance

Fischer *et al.,* Approximation of graph edit distance based on Hausdorff matching. Pattern Recognition, 2015.

Graph Kernel: Hausdorff Distance

Graph Edit Distance – Expressed in terms of:

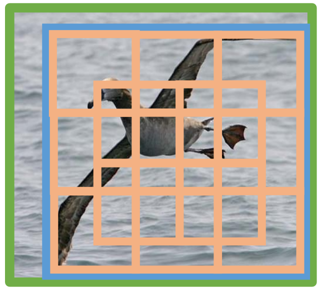
- Node Insertion
- Node Deletion
- Node Substitution
- Edge Insertion
- Edge Deletion
- Edge Substitution



More graph edit operations relative to vanilla Hausdorff

Riba *et al.*, Learning graph edit distance by graph neural networks. Pattern Recognition, 2021.

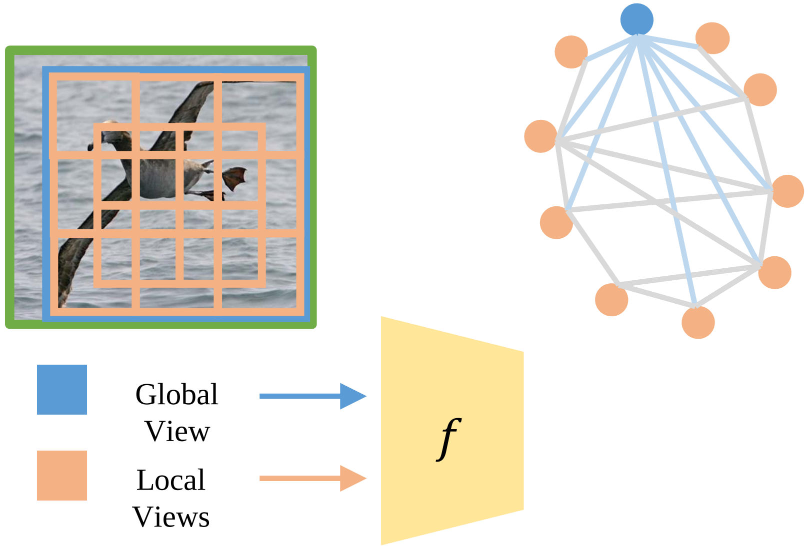
Transitivity Recovering Decompositions



Global View

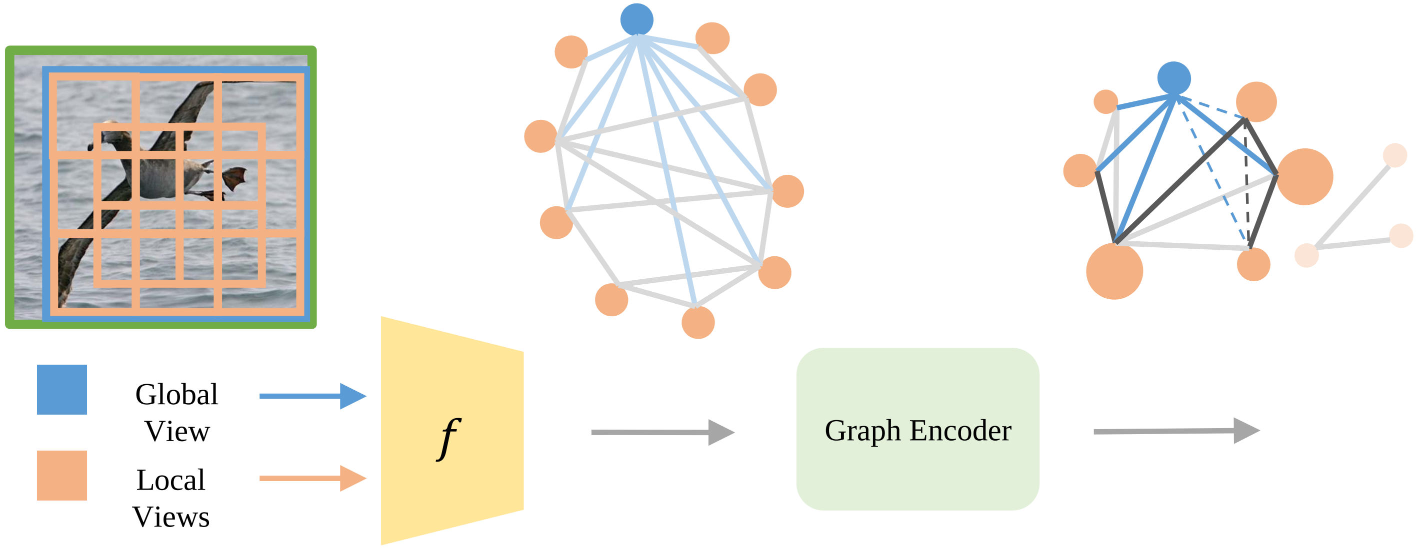
Local Views

Transitivity Recovering Decompositions



Complementarity Graph

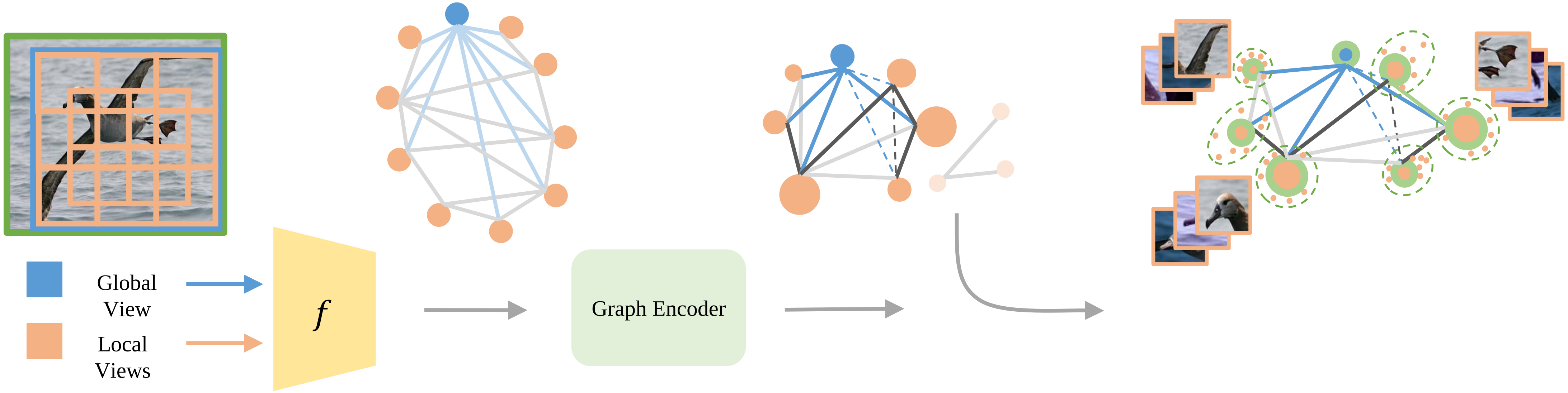
Transitivity Recovering Decompositions



Complementarity Graph

Semantic Relevance Graph

Transitivity Recovering Decompositions

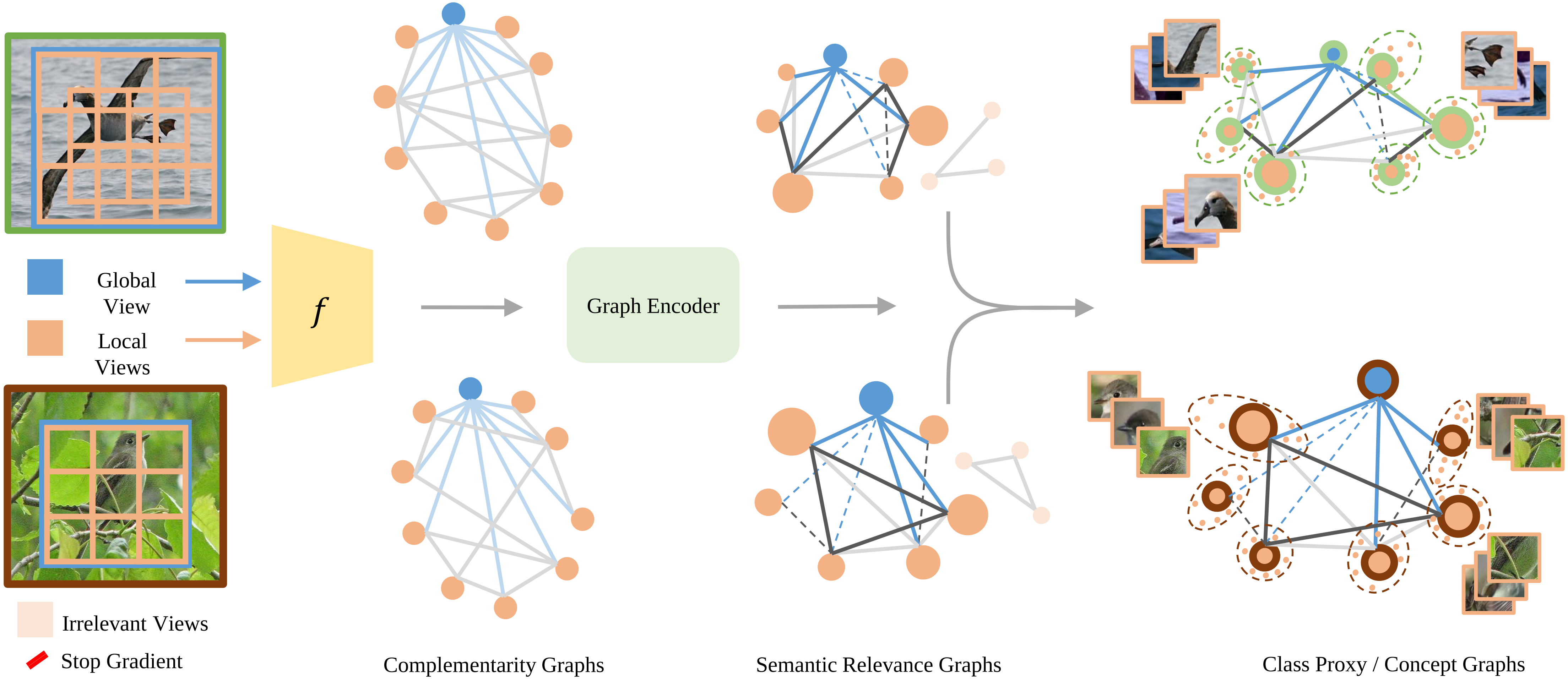


Complementarity Graph

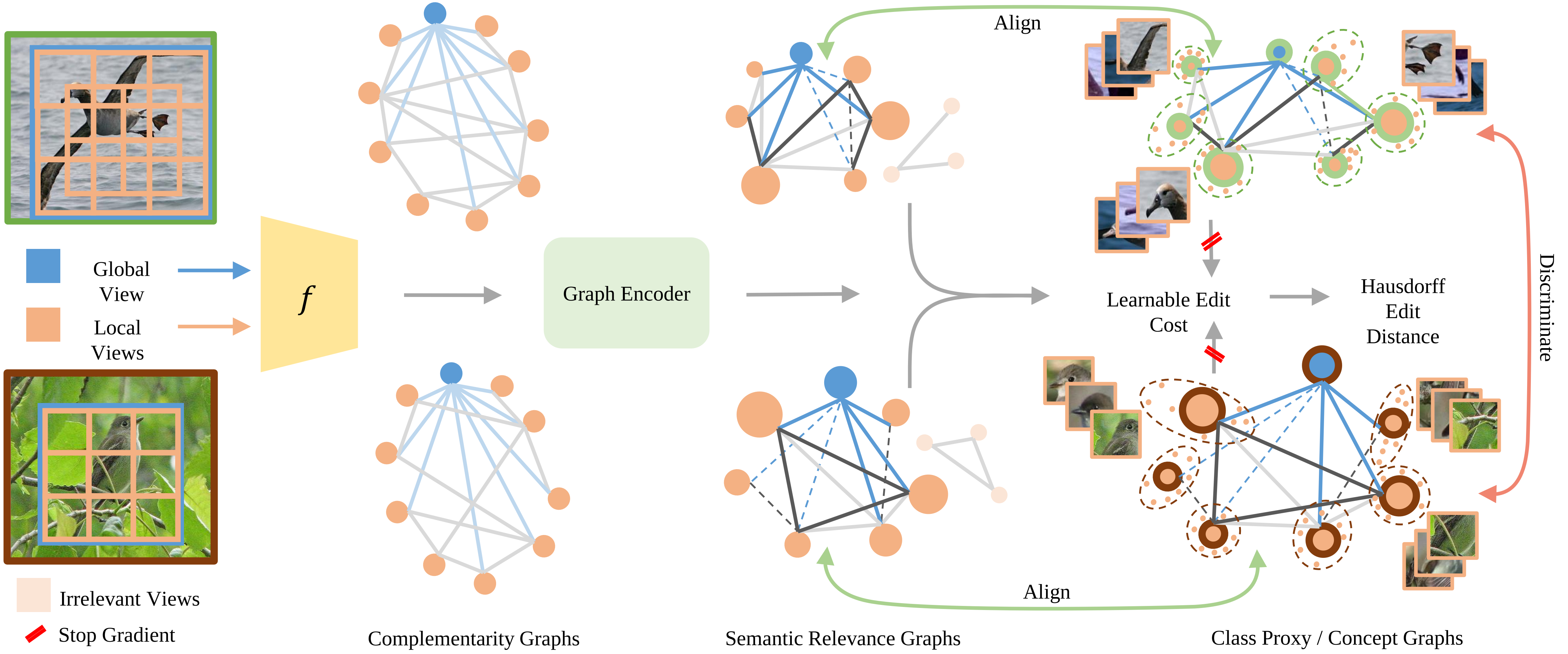
Semantic Relevance Graph

Class Proxy / Concept Graph

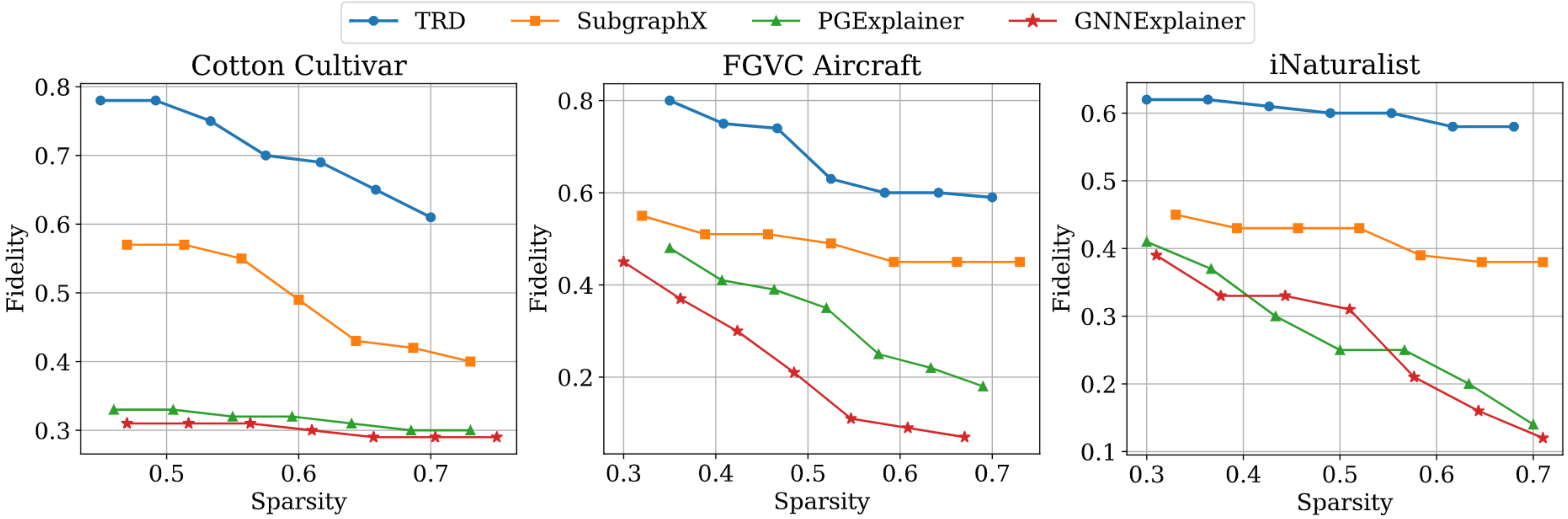
Transitivity Recovering Decompositions



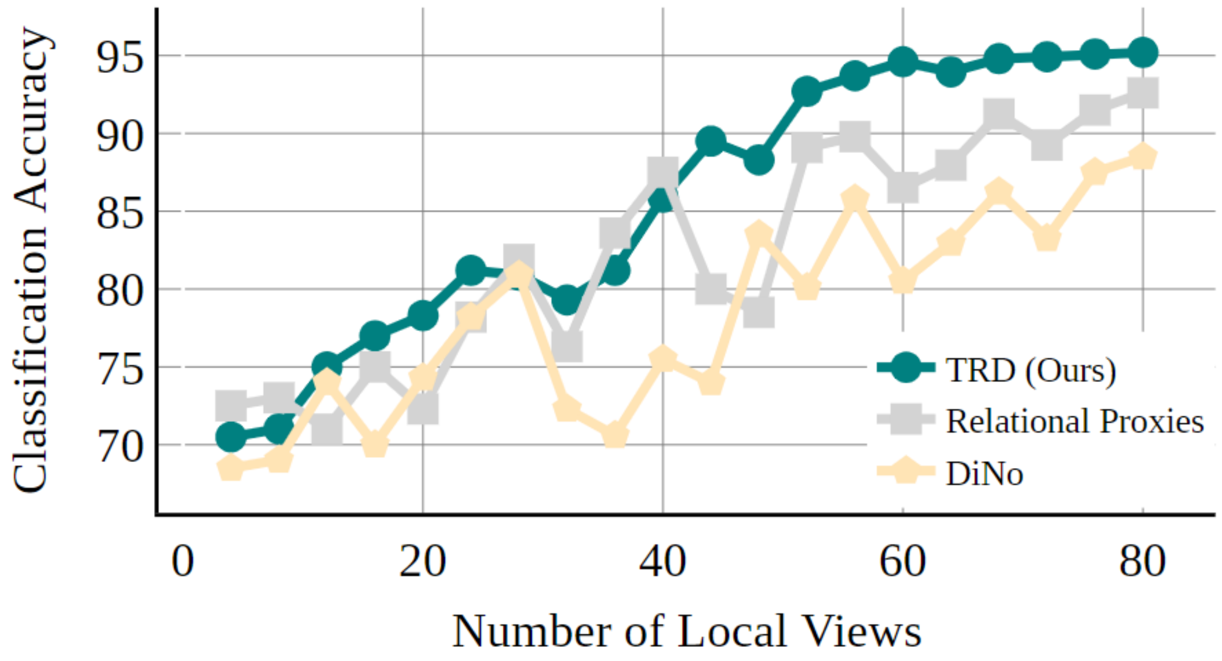
Transitivity Recovering Decompositions

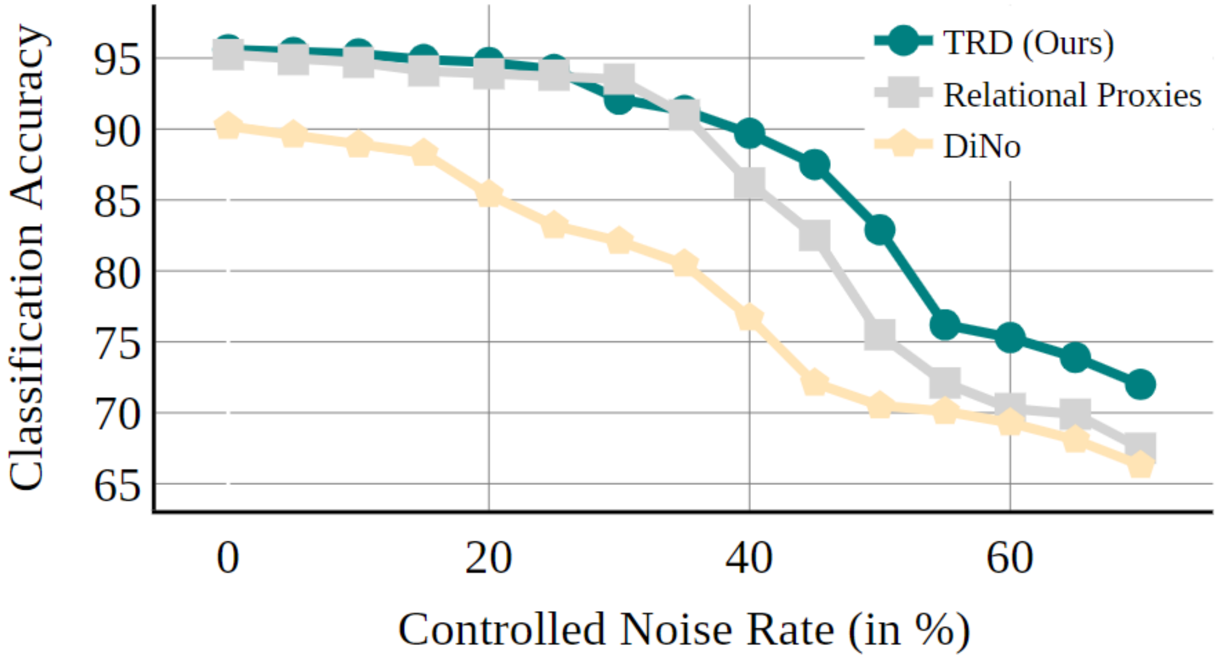


Performance – Interpretability



Performance – Robustness

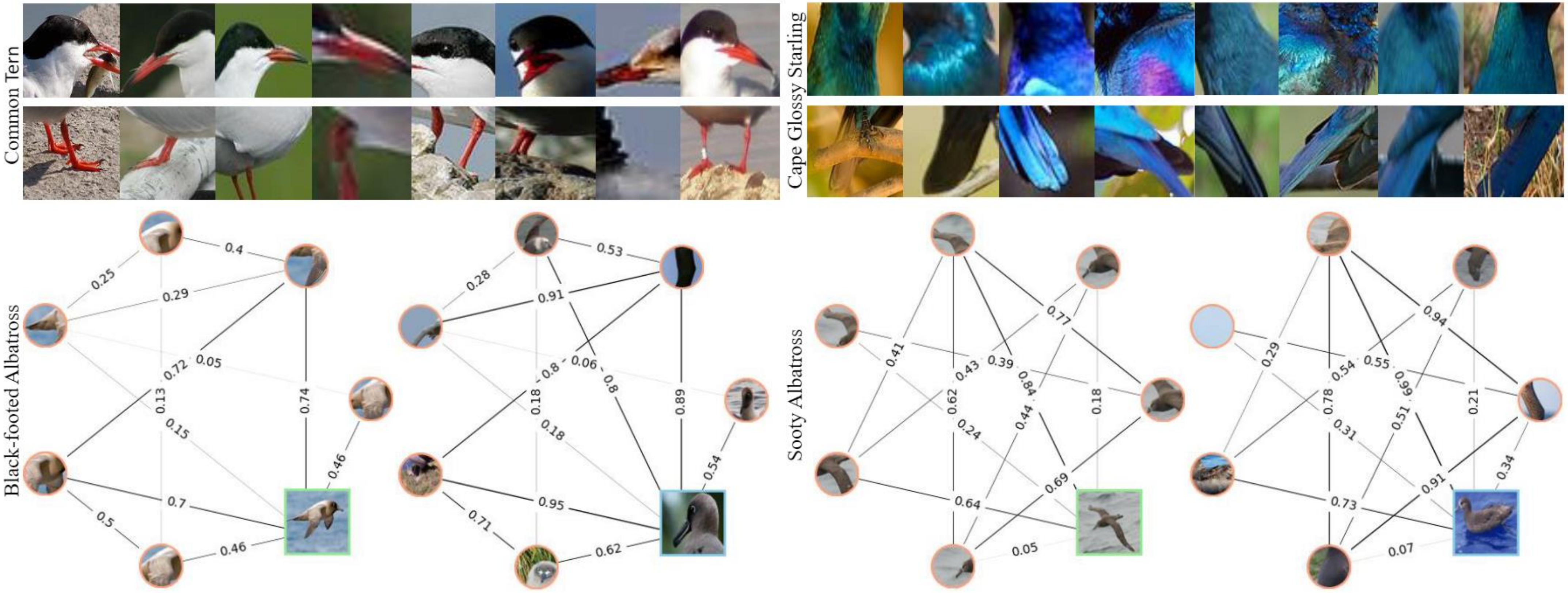




Performance – FGVC

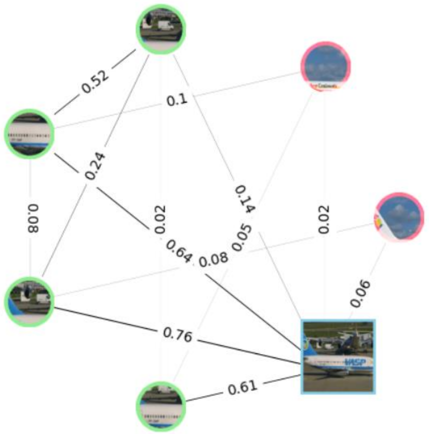
Method	Small		Medium				Large
	Cotton	Soy	FGVC Aircraft	Stanford Cars	CUB	NA Birds	iNaturalist
MaxEnt, NeurIPS'18	-	-	89.76	93.85	86.54	-	-
DBTNet, NeurIPS'19	-	-	91.60	94.50	88.10	-	-
StochNorm, NeurIPS'20	45.41	38.50	81.79	87.57	79.71	74.94	60.75
MMAL, MMM'21	65.00	47.00	94.70	95.00	89.60	87.10	69.85
FFVT, BMVC'21	57.92	44.17	79.80	91.25	91.65	89.42	70.30
CAP, AAAI'21	-	-	94.90	95.70	91.80	91.00	-
GaRD, CVPR'21	64.80	47.35	94.30	95.10	89.60	88.00	69.90
TransFG, AAAI'22	45.84	38.67	80.59	94.80	91.70	90.80	71.70
Relational Proxies, NeurIPS'22	69.81	51.20	95.25	96.30	92.00	91.20	72.15
TRD (Ours)	**70.90** ± 0.22	**52.15** ± 0.12	**95.60** ± 0.08	**96.35** ± 0.03	**92.10** ± 0.04	**91.45** ± 0.12	**72.27** ± 0.05

Qualitative Results

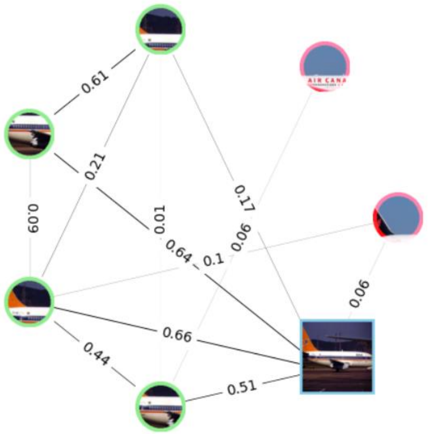


Instance Graph Class-Proxy Graph Instance Graph Class-Proxy Graph

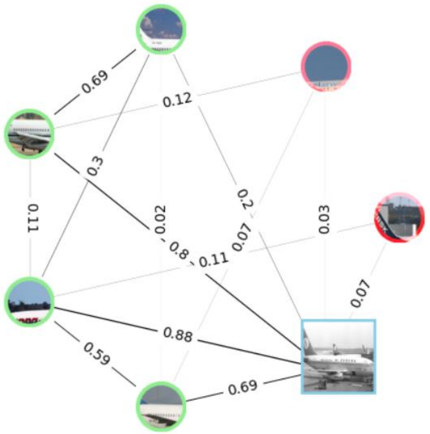
Causal Interventions



Instance Graph



Instance Graph



Class-Proxy Graph

η	10	20	30	40	50
Relational Proxies	93.22	87.12	79.35	70.99	63.60
TRD (Ours)	**94.90**	**91.54**	**82.80**	**76.35**	**70.55**

η: Percentage of local-views from a different class

Conclusion

- Abstract emergent relationships can be expressed in terms of graphs.

Conclusion

- Abstract emergent relationships can be expressed in terms of graphs.
- Transitivity Recovering Decompositions (TRD) is a provably efficient approach to achieve the same.

Conclusion

- Abstract emergent relationships can be expressed in terms of graphs.
- Transitivity Recovering Decompositions (TRD) is a provably efficient approach to achieve the same.
- TRD encodes the complete relational semantics while being interpretable.

Conclusion

- Abstract emergent relationships can be expressed in terms of graphs.
- Transitivity Recovering Decompositions (TRD) is a provably efficient approach to achieve the same.
- TRD encodes the complete relational semantics while being interpretable.
- Recovering transitive relationships inherently filters out noisy views.

Transitivity Recovering Decompositions

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

https://arxiv.org/abs/2310.15999

https://github.com/abhrac/trd