

PDiscoFormer: Relaxing Part Discovery Constraints with Vision Transformers



Ananthu Aniraj



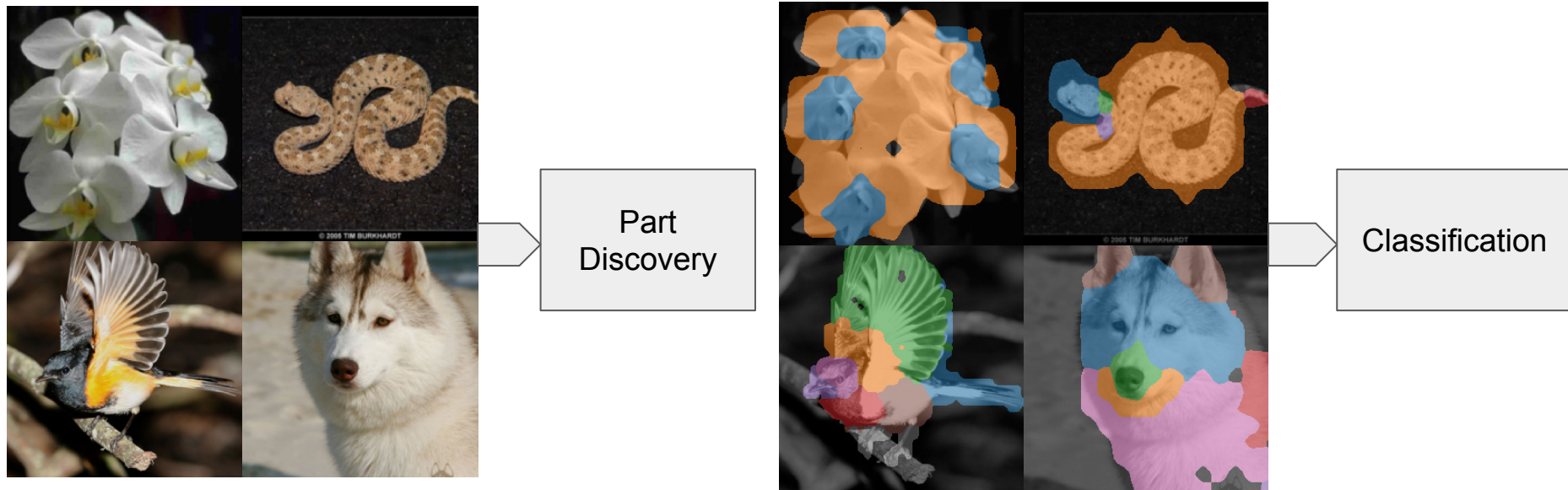
Cassio Fraga Dantas



Dino Ienco

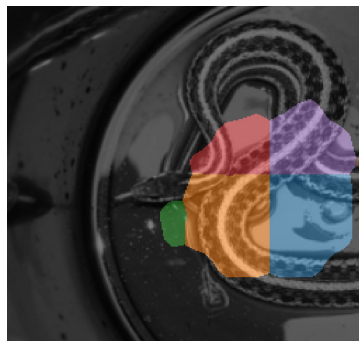
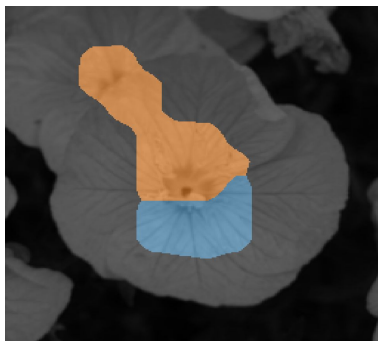


Diego Marcos



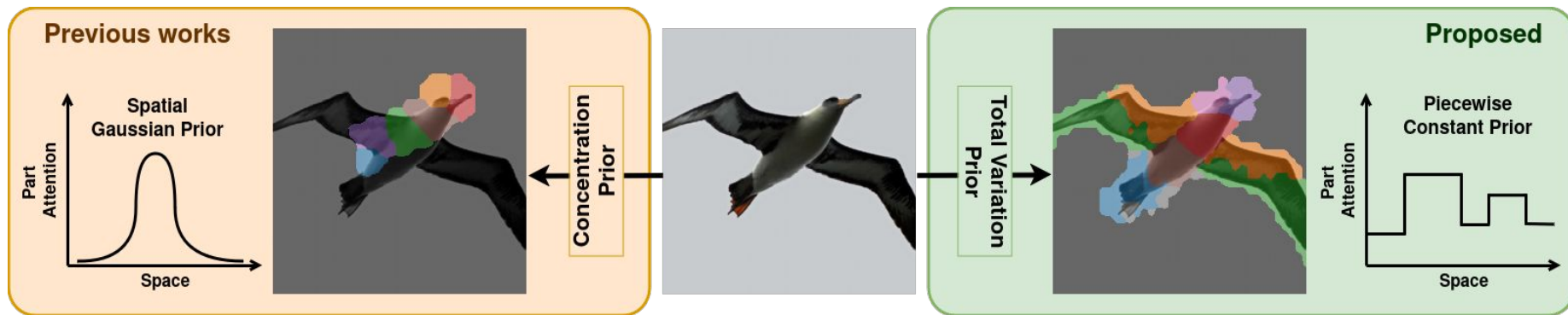
Labelled information: image class labels

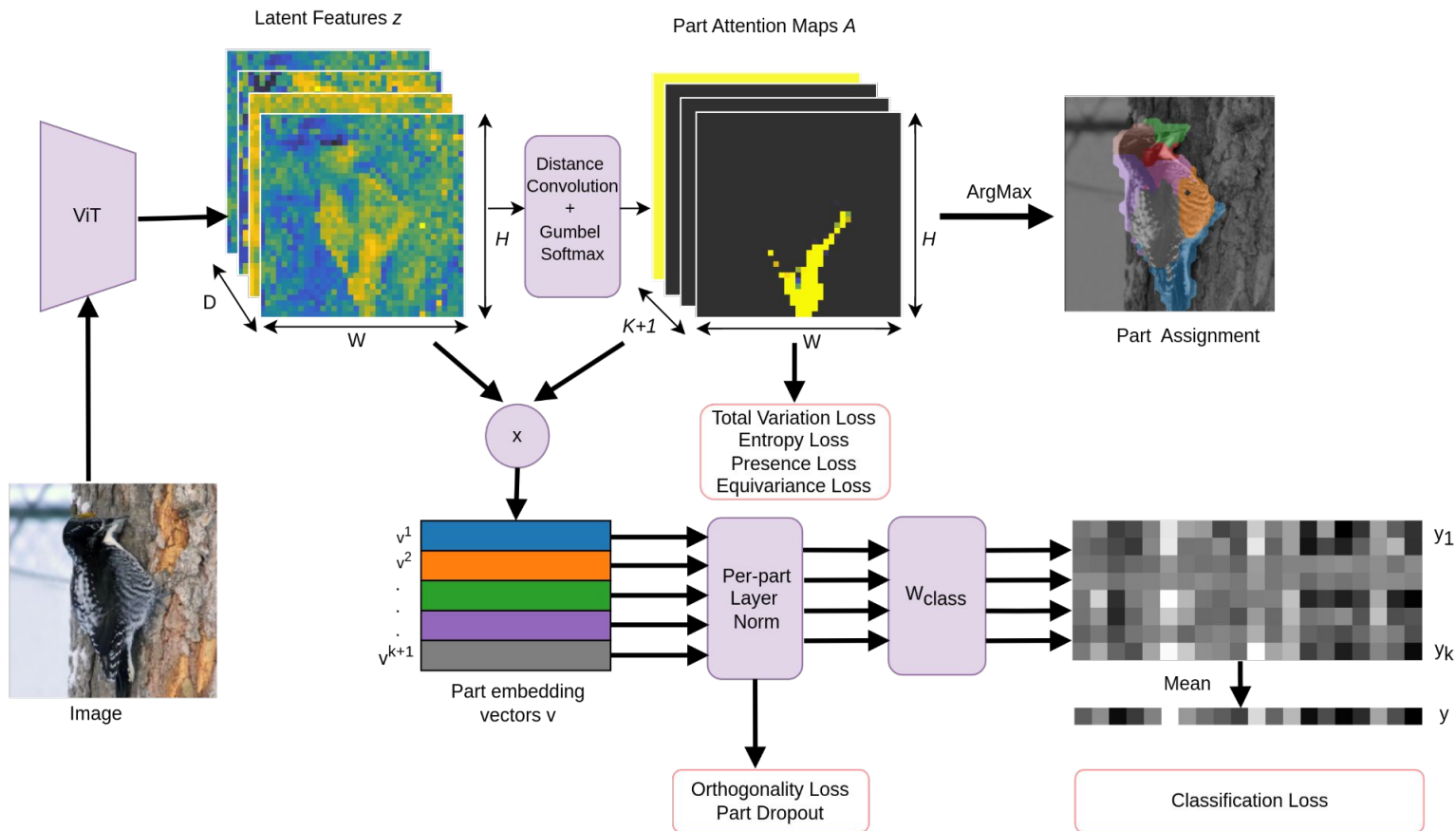
- Restrictive constraints (assumptions) for discovered parts
 - Parts occur only once in an image
 - Parts are compact
- Designed for CNNs



- Can we use self-supervised ViTs for part discovery?
- Can we perform part discovery without restrictive constraints?

Possible solution: Use a total variation prior





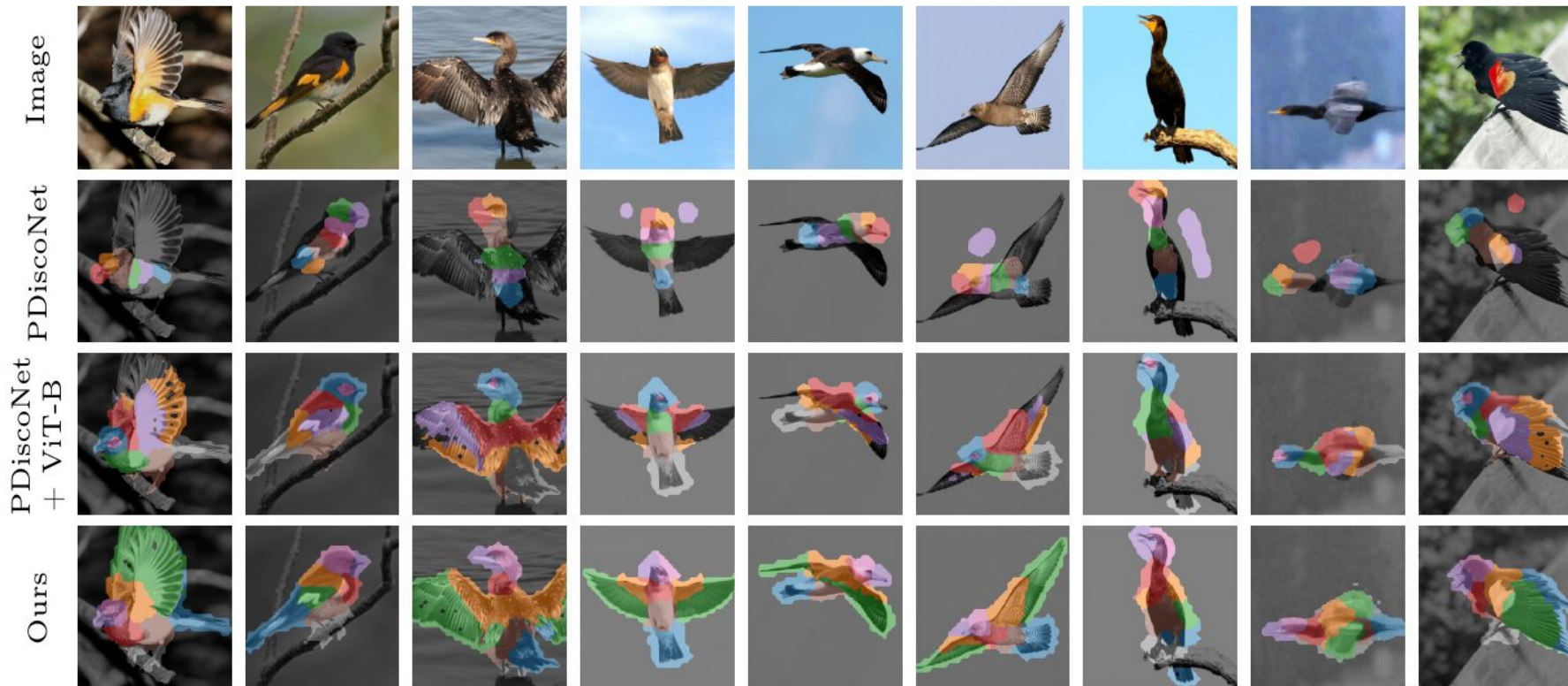
Method	CUB (%) K=16			PartImageNet OOD (%) K=50			Flowers (%) K=2	
	NMI (↑)	ARI (↑)	Top-1 Acc. (↑)	NMI (↑)	ARI (↑)	Top-1 Acc. (↑)	Fg. mIoU (↑)	Top-1 Acc. (↑)
Dino [1]	50.57	26.14	-	37.81	16.50	-	54.44	-
Huang [2]	43.92	21.10	85.93	10.19	1.05	73.20	17.26	92.86
PDiscoNet [3]	56.87	38.05	87.49	41.49	14.17	86.06	49.10	81.04
Pdisconet + ViT-B	68.63	43.41	84.04	29.48	27.80	89.69	13.18	97.40
Ours	73.38	55.83	88.72	46.29	62.21	91.01	69.59	99.64

[1] Amir, Shir, et al. "Deep vit features as dense visual descriptors." *arXiv preprint arXiv:2112.05814* 2.3 (2021): 4.

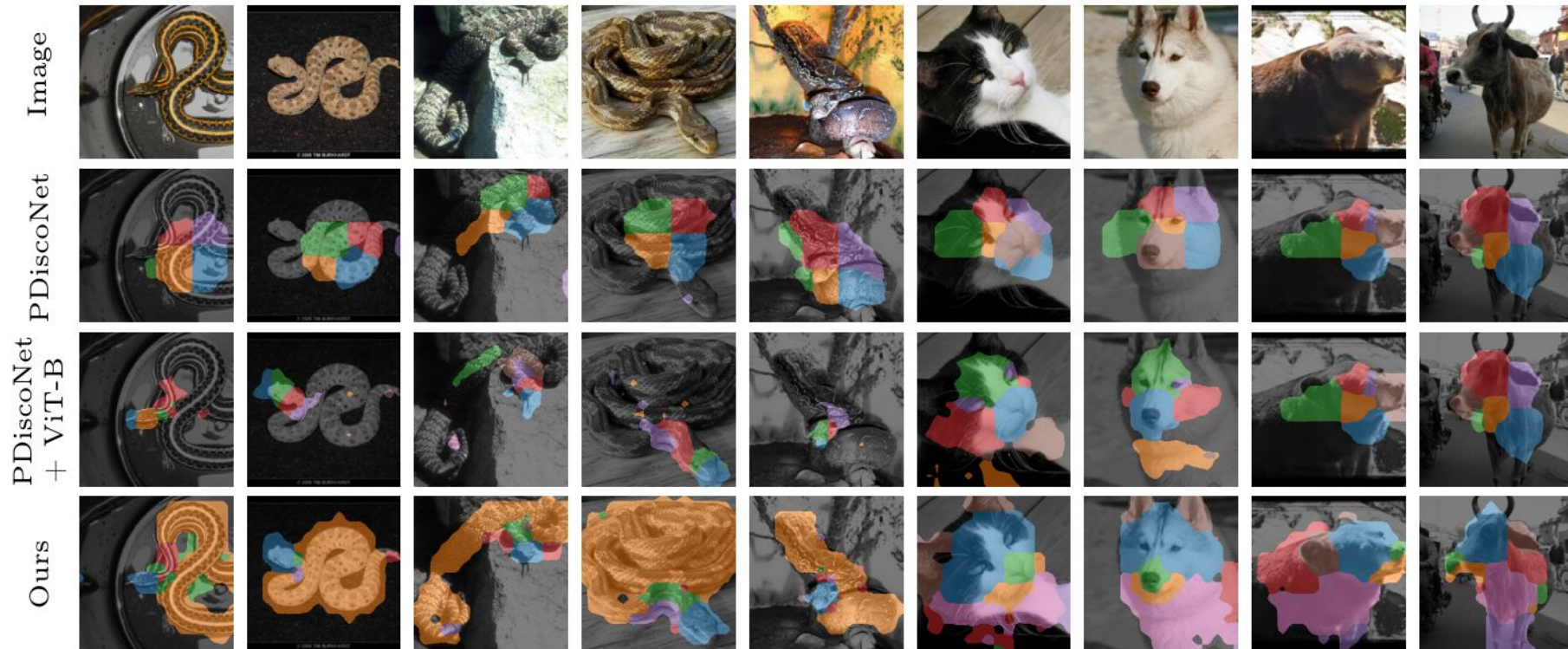
[2] Huang, Z., & Li, Y. (2020). Interpretable and accurate fine-grained recognition via region grouping. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8662–8672.

[3] van der Klis, R., Alaniz, S., Mancini, M., Dantas, C. F., Ienco, D., Akata, Z., & Marcos, D. (2023). PDiscoNet: Semantically consistent part discovery for fine-grained recognition. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1866–1876.

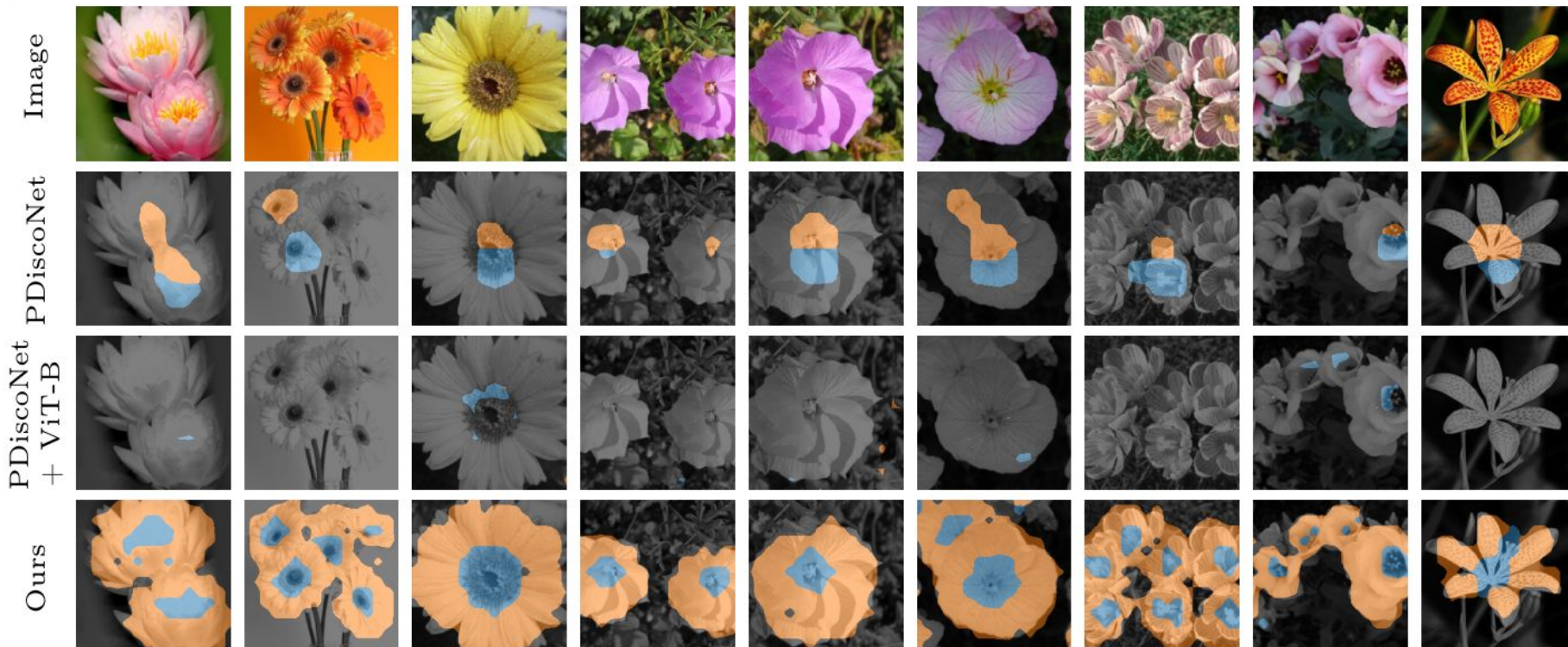
Qualitative Results (CUB) - 8 parts



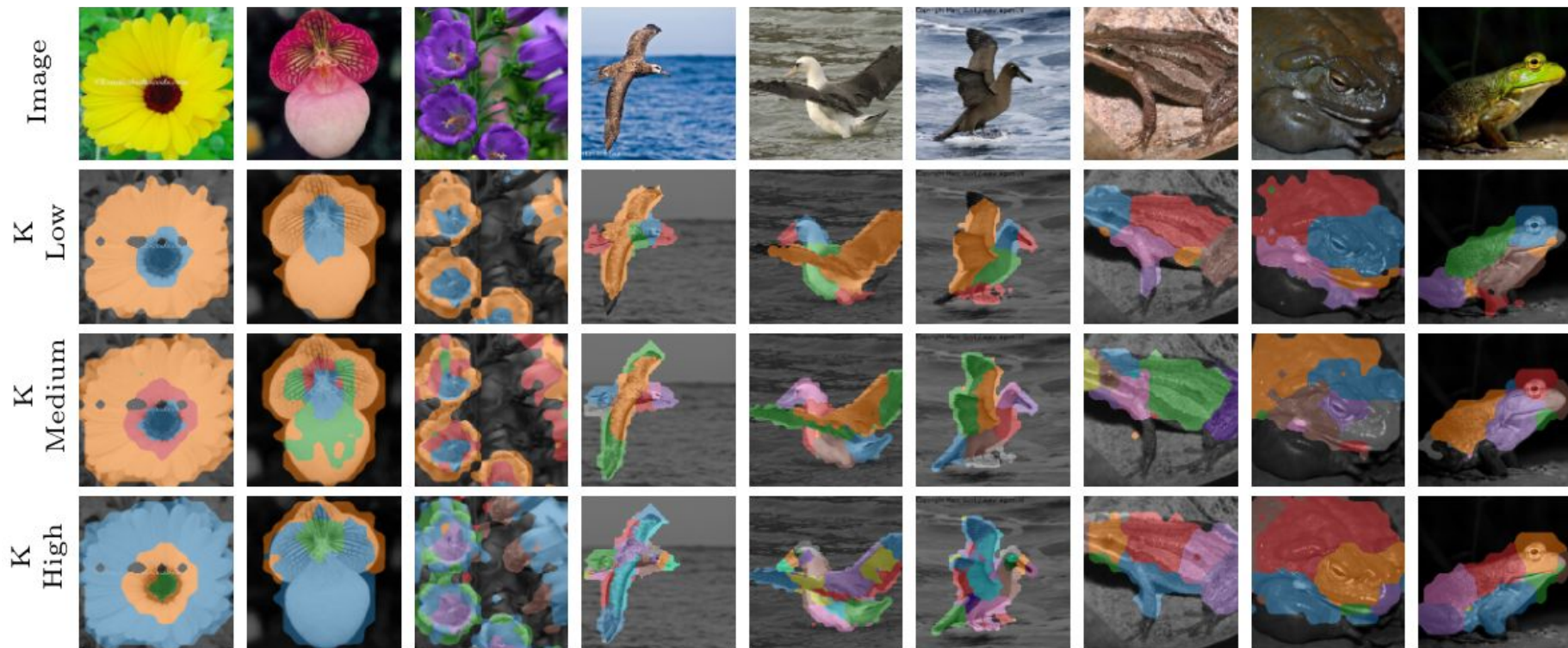
Qualitative Results (PartImageNet) - 8 parts



Qualitative Results (Flowers) - 2 parts



Effect of varying K (number of parts)



Conclusion

- Novel approach for part discovery with self-supervised ViTs
- Total variation acts as a flexible part shape prior
- SOTA quantitative results for part discovery across datasets
- Good qualitative results

Thank you!



Paper



Code