Unlocking Compositional Generalization in Continual Few-Shot Learning

Phu-Quy Nguyen-Lam[1,2,†] Phu-Hoa Pham[1,2,†] Duy-Minh Dao-Sy[1,2] Chi-Nguyen Tran[1,2] Trung-Kiet Huynh[1,2] Long Tran-Thanh[3,*]

[1]Faculty of Information Technology, University of Science, Ho Chi Minh City, Vietnam [2]Vietnam National University, Ho Chi Minh City, Vietnam
[3]Department of Computer Science, University of Warwick, Coventry, United Kingdom [†]Equal contribution [*]Corresponding author

NEURAL INFORMATION PROCESSING SYSTEMS

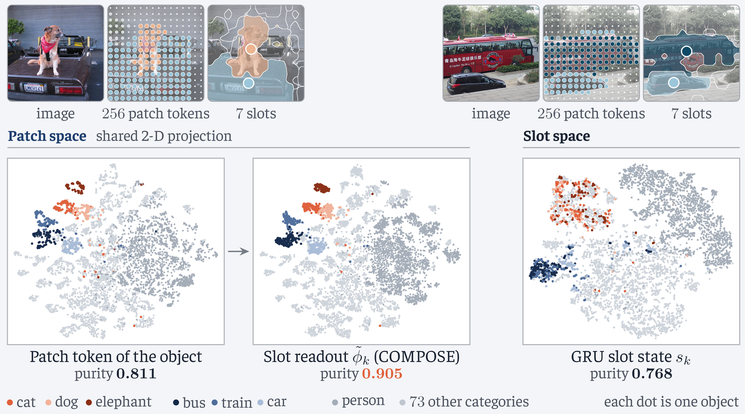
Code

Motivation

Object-centric representations should let a model recognize a novel concept from known parts. Continual learners lose this capacity through two coupled flaws.

Flaw 1 · Representation Pollution

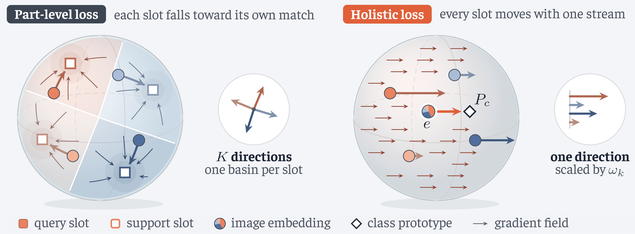
The GRU hidden state s_k entangles routing dynamics with visual content.
Remedy: an attention-weighted readout of the patch tokens, $\tilde{\phi}_k$.



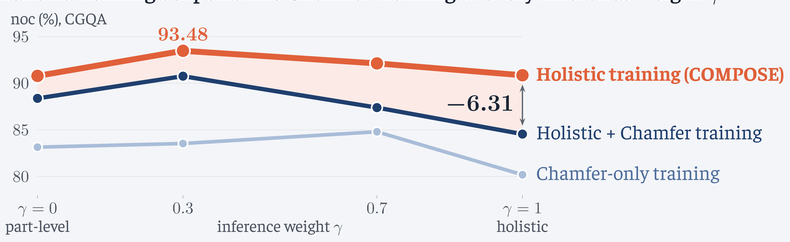
Flaw 2 · The Compositional Optimization Trap

A part-level matching objective pulls each slot toward its own nearest support slot.

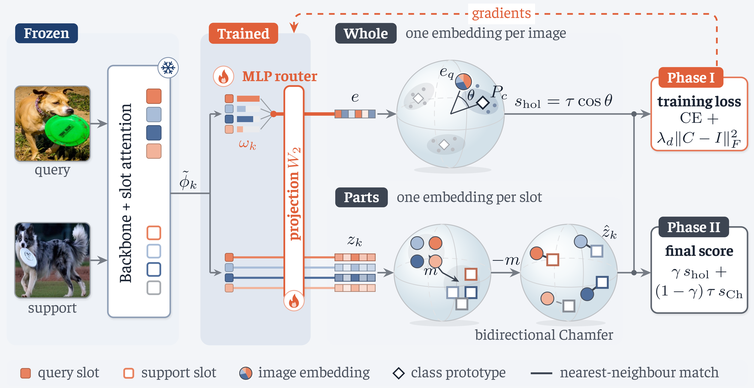
Train holistically, infer compositionally.



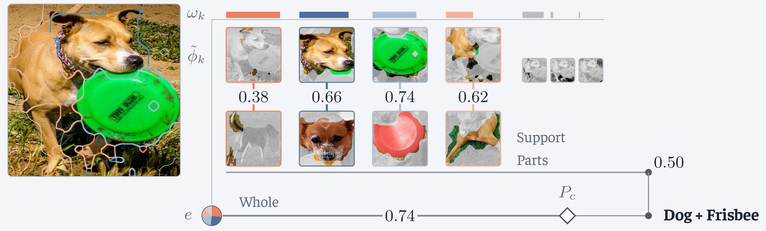
Holistic training outperforms Chamfer training at every inference weight γ



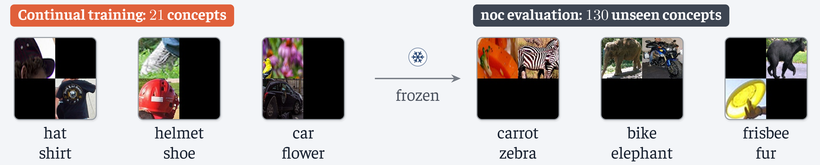
The COMPOSE Framework



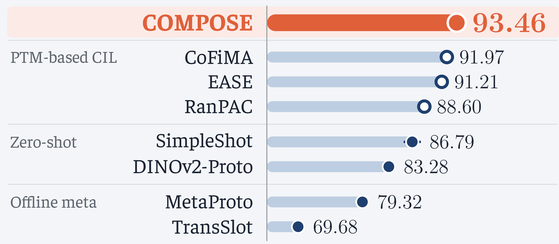
Compositional Inference on an Unseen Concept



Highest Unseen-Concept Accuracy on CGQA



93.46%
noc accuracy,
10-way 10-shot



Conclusion

1. Attention-weighted slot readouts preserve the semantic geometry of the patch tokens.
2. Training holistically and inferring compositionally escapes the compositional trap.
3. A cross-correlation penalty prevents redundancy in the slot embedding.

Experiments

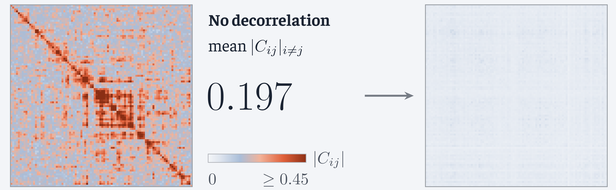
The Gain Stems from Holistic Training

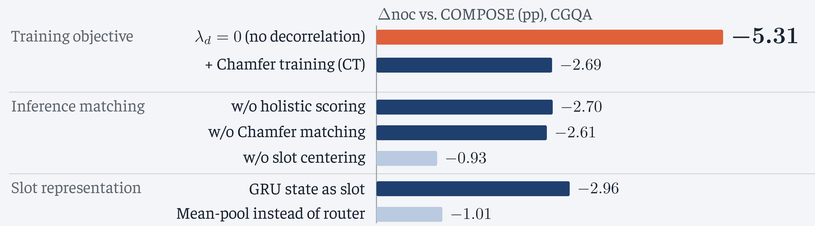
Training objective	Inference matcher	noc on CGQA (%), 3 seeds			
		1-shot	5-shot	10-shot	20-shot
Holistic	Chamfer	74.67	91.22	93.46	94.83
Holistic	Hungarian	73.11	90.66	92.97	94.44
Holistic + Chamfer	Chamfer	69.69	87.88	90.77	92.41

Adding any part-level matching loss reduces noc and increases sys

CGQA, 3 sessions	Holistic	Holistic + Chamfer: hard, or soft with temperature β					
		Hard	$\beta = 50$	$\beta = 20$	$\beta = 10$	$\beta = 5$	$\beta = 1$
noc	92.02	87.85	87.98	88.25	88.07	86.76	82.57
sys	86.81	88.89	88.86	88.94	89.01	89.03	87.04

Ablation: Decorrelation Is the Primary Driver





Standard Continual Learning Benchmarks

Class-incremental learning: highest final accuracy on ImageNet-A

Method	CIFAR-100		ImageNet-R		ImageNet-A	
	Accuracy	Forgetting	Accuracy	Forgetting	Accuracy	Forgetting
COMPOSE	86.52	7.70	82.30	5.97	70.31	4.56
RanPAC	**86.61**	**7.30**	81.60	8.18	63.99	2.28
FeCAM	86.09	9.20	**83.12**	**4.40**	66.95	2.28
F-OAL	86.44	7.40	81.70	6.45	67.22	**0.65**

Few-shot class-incremental learning: far less forgetting than Comp-FSCIL

