南京航空航天大學

Nanjing University of Aeronautics and Astronautics

ParNeC 模式识别与神经计算研究组

PAttern Recognition and NEural Computing

Sparse Mixture-of-Experts are Domain Generalizable Learners

Bo Li, Yifei Shen, Jingkang Yang, Yezhen Wang, Jiawei Ren, Tong Che, Jun Zhang, Ziwei Liu

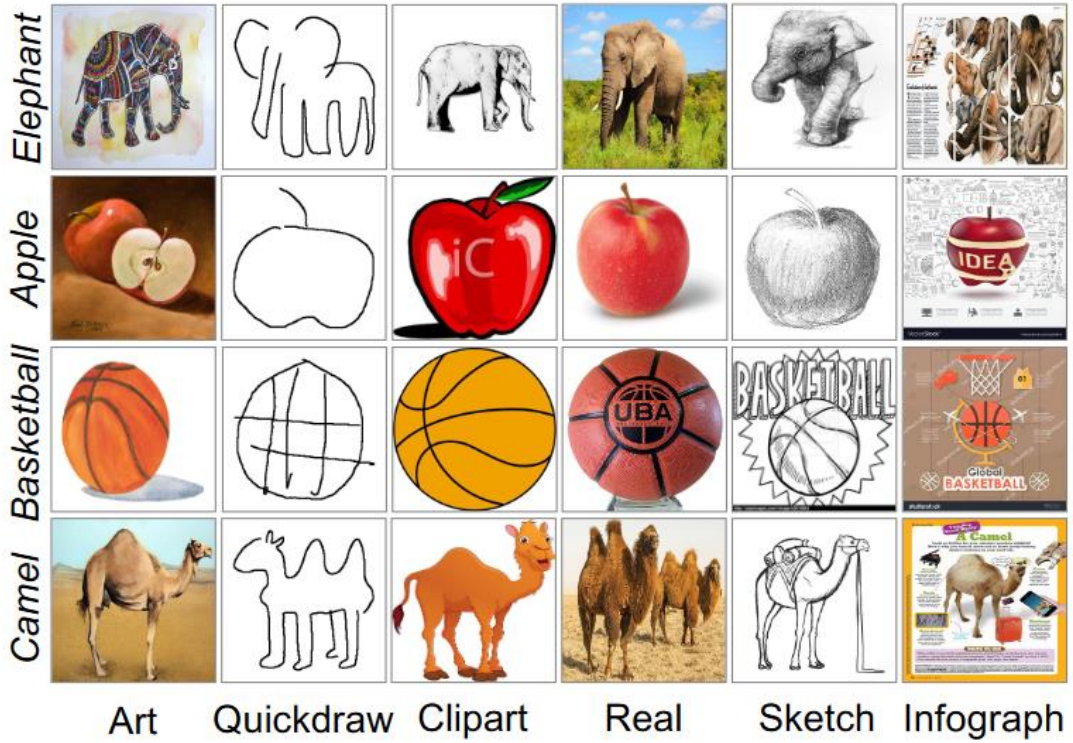
S-Lab, NTU; HKUST; Mila-Quebec AI Institute; Nvidia Research

ICLR 2023

Domain Generalization

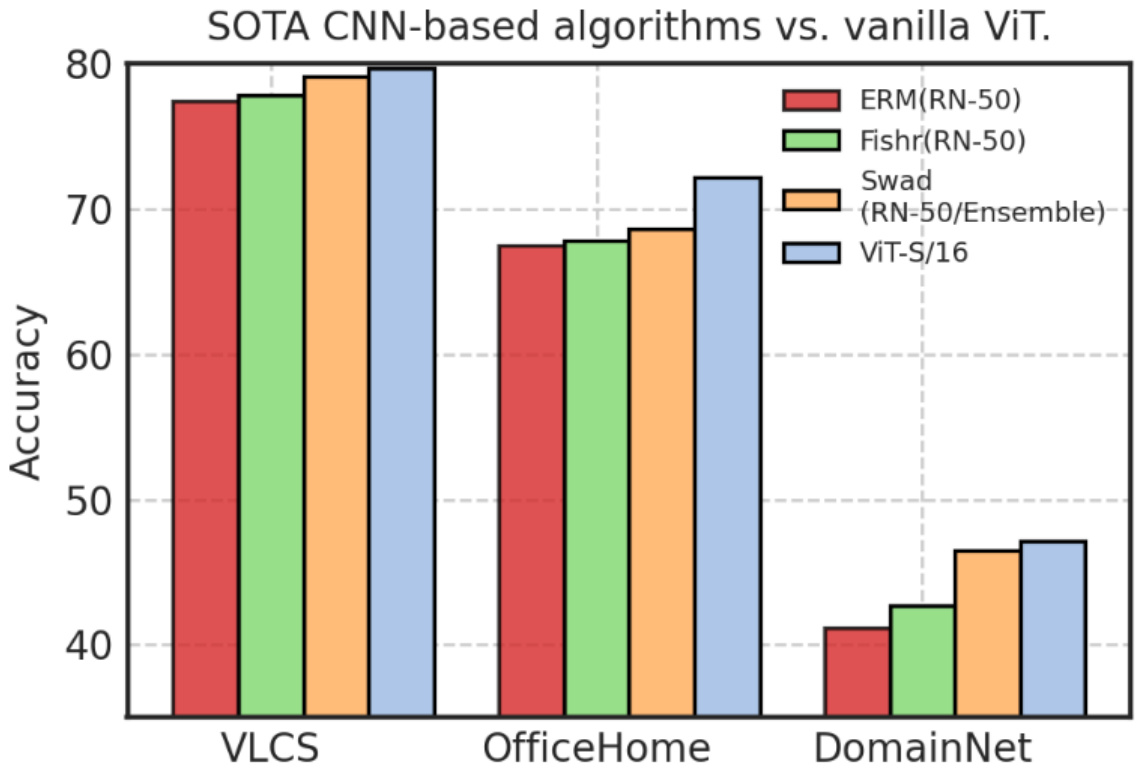
Domain Generalization:
CANNOT see the target domain when training

Domain Adaptation:
CAN see the target domain when training



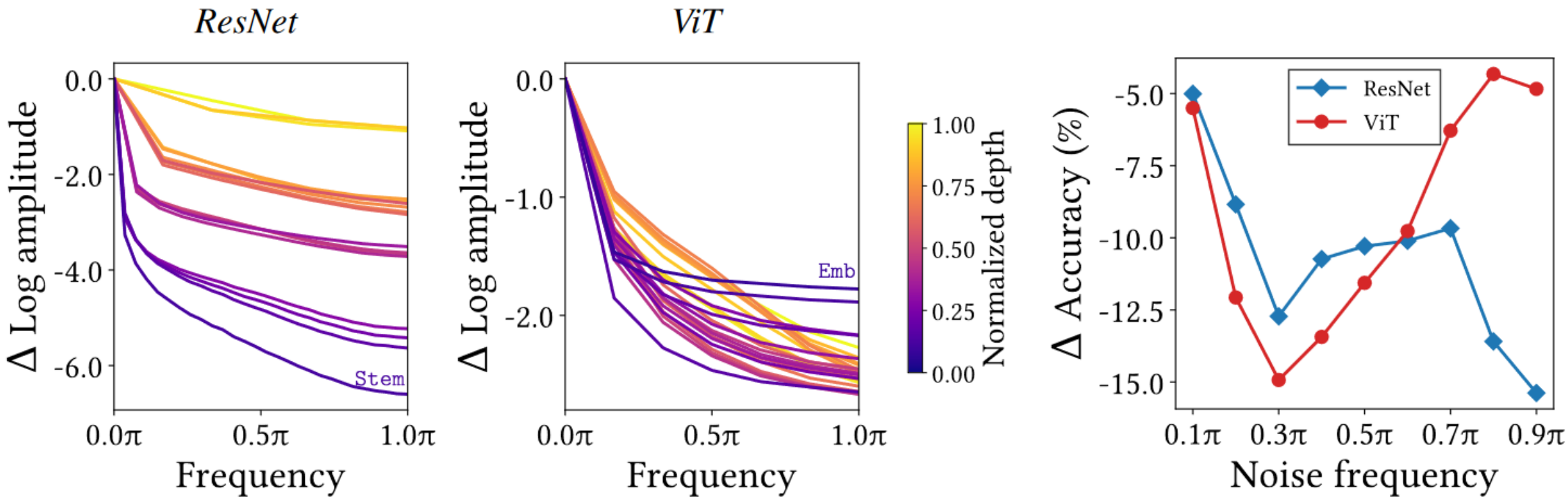
ResNet-50 vs ViT-S/16

Performance comparison of ViT-S/16 (w/**21.8M** trainable parameters) with ERM and the ResNet-50 (w/**25.6M**) backbone with SOTA DG algorithms. ERM(RN-50), Fishr(RN-50), Swad(RN50/Ensemble) denotes the ResNet-50 trained with ERM, Fish, Fishr, and Swad respectively.



An explanation of this phenomenon

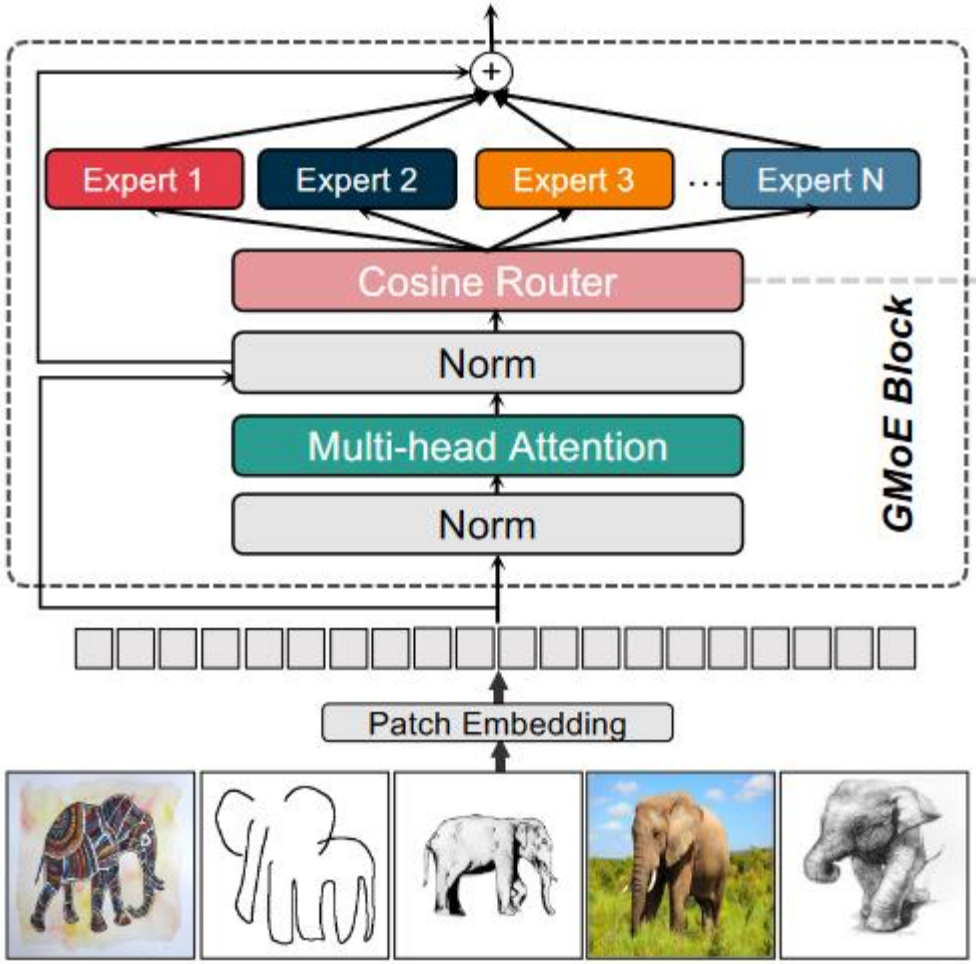
How Do Vision Transformers Work? ICLR'22

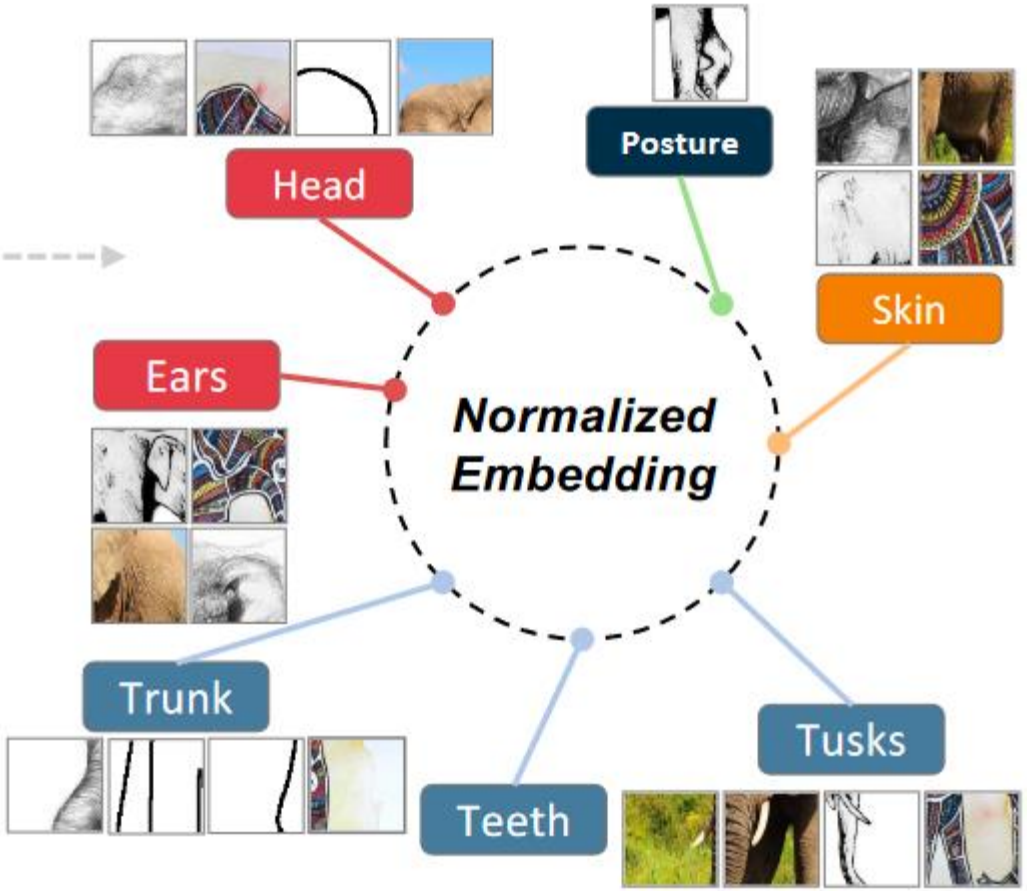


(a) Relative log amplitudes of Fourier transformed feature maps.

(b) Robustness for noise frequency

Overview architecture of GMoE





Sparse MoE Layer

Sparse MoE layer: $f_{\mathrm{MoE}}(\boldsymbol{x}) = \sum_{i=1}^{N} G(\boldsymbol{x})_i \cdot E_i(\boldsymbol{x}))$

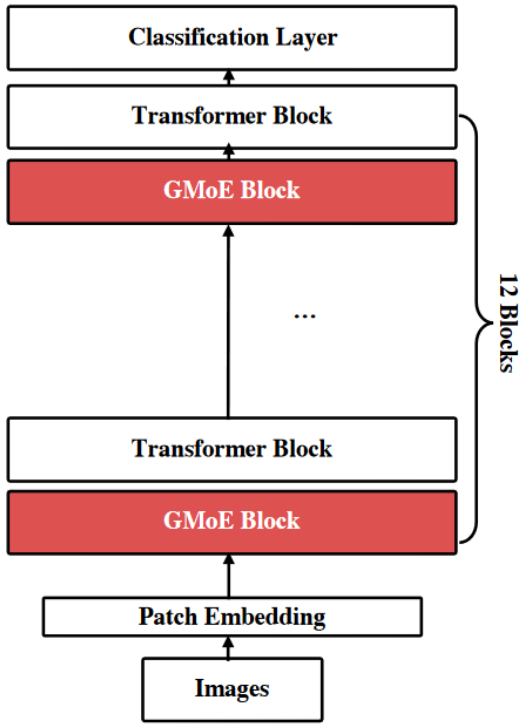
Cosine Router:

$$G(\boldsymbol{x}) = \mathrm{TOP}_k\left(\mathrm{Softmax}\left(\frac{\boldsymbol{E}^T \boldsymbol{W}\boldsymbol{x}}{\tau \|\boldsymbol{W}\boldsymbol{x}\| \|\boldsymbol{E}\|}\right)\right)$$

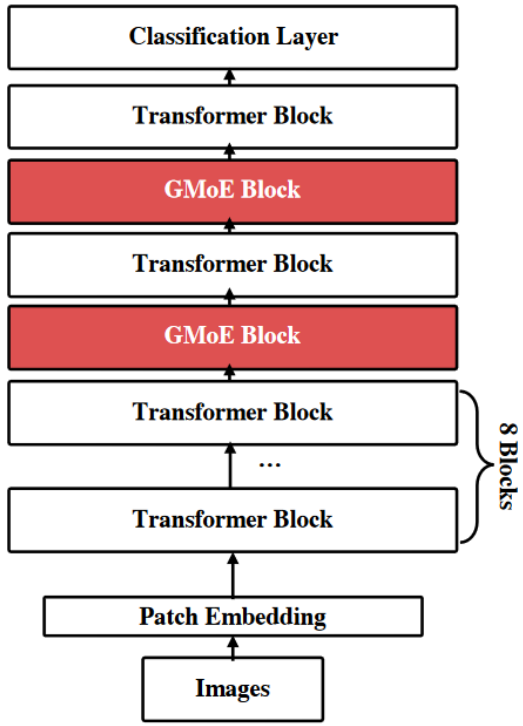
Linear Router:

$$G(\boldsymbol{x}) = \mathrm{TOP}_k(\mathrm{Softmax}(\boldsymbol{W}\boldsymbol{x}))$$

Number of MoE layers: Every-two vs Last-two



(a) *Every 2* GMoE.



(b) *Last 2* GMoE.

Algorithm	Config.	Router	H	D	PACS	VLCS	OfficeHome	TerraInc	DomainNet
GMoE-S/16	*Every 2*	Linear	6	384	81.8 $\pm$ 0.2	75.0 $\pm$ 0.1	64.0 $\pm$ 0.4	32.5 $\pm$ 0.7	46.3 $\pm$ 0.3
GMoE-S/16	*Every 2*	Cosine	6	384	81.4 $\pm$ 0.1	74.8 $\pm$ 0.2	62.2 $\pm$ 0.4	40.9 $\pm$ 0.3	46.4 $\pm$ 0.2
GMoE-S/16	*Last 2*	Linear	6	384	87.8 $\pm$ 0.2	80.0 $\pm$ 0.0	72.7 $\pm$ 0.2	46.7 $\pm$ 0.2	48.3 $\pm$ 0.1
GMoE-S/16	*Last 2*	Cosine	6	384	88.1 $\pm$ 0.1	80.2 $\pm$ 0.2	74.2 $\pm$ 0.4	48.5 $\pm$ 0.4	48.7 $\pm$ 0.2
GMoE-B/16	*Last 2*	Cosine	12	768	89.4 $\pm$ 0.1	81.2 $\pm$ 0.1	77.2 $\pm$ 0.4	49.3 $\pm$ 0.3	51.3 $\pm$ 0.1

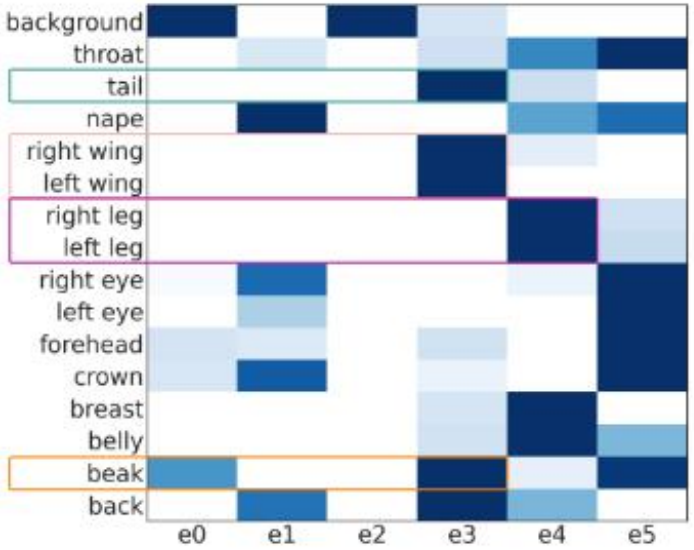
Experiments

Algorithm	PACS	VLCS	OfficeHome	TerraInc	DomainNet
ERM (ResNet50) (Vapnik, 1991)	85.7 ± 0.5	77.4 ± 0.3	67.5 ± 0.5	47.2 ± 0.4	41.2 ± 0.2
IRM [ArXiv 20] (Arjovsky et al., 2019)	83.5 ± 0.8	78.5 ± 0.5	64.3 ± 2.2	47.6 ± 0.8	33.9 ± 2.8
DANN [JMLR 16] (Ganin et al., 2016)	84.6 ± 1.1	78.7 ± 0.3	68.6 ± 0.4	46.4 ± 0.8	41.8 ± 0.2
CORAL [ECCV 16] (Sun & Saenko, 2016)	86.0 ± 0.2	77.7 ± 0.5	68.6 ± 0.4	46.4 ± 0.8	41.8 ± 0.2
MMD [CVPR 18] (Li et al., 2018b)	85.0 ± 0.2	76.7 ± 0.9	67.7 ± 0.1	42.2 ± 1.4	39.4 ± 0.8
FISH [ICLR 22] (Shi et al., 2021)	85.5 ± 0.3	77.8 ± 0.3	68.6 ± 0.4	45.1 ± 1.3	42.7 ± 0.2
SWAD [NeurIPS 21] (Cha et al., 2021a)	88.1 ± 0.1	79.1 ± 0.1	70.6 ± 0.2	50.0 ± 0.3	46.5 ± 0.1
Fishr [ICML 22] (Rame et al., 2021)	85.5 ± 0.2	77.8 ± 0.2	68.6 ± 0.2	47.4 ± 1.6	41.7 ± 0.0
MIRO [ECCV 22] (Cha et al., 2022)	85.4 ± 0.4	79.0 ± 0.0	70.5 ± 0.4	**50.4 ± 1.1**	44.3 ± 0.2
ERM (ViT-S/16) [ICLR 21] (Dosovitskiy et al., 2021)	86.2 ± 0.1	79.7 ± 0.0	72.2 ± 0.4	42.0 ± 0.8	47.3 ± 0.2
GMoE-S/16 (Ours)	**88.1 ± 0.1**	**80.2 ± 0.2**	**74.2 ± 0.4**	48.5 ± 0.4	**48.7 ± 0.2**

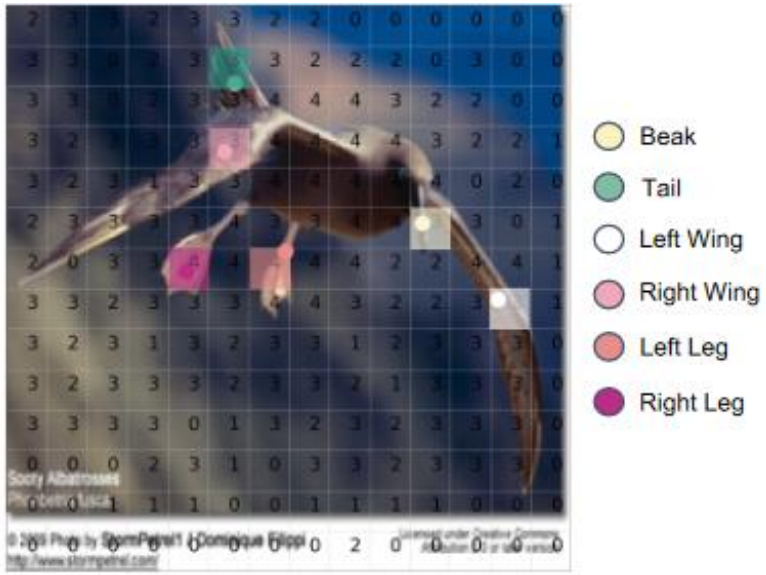
Algorithms	SVIRO	Wilds-Camelyon	Wilds-FMOW
ERM (ResNet50) (Vapnik, 1991)	85.7 ± 0.1	93.1 ± 0.2	40.6 ± 0.4
ERM (ViT-S/16) [ICLR 21] (Dosovitskiy et al., 2021)	89.6 ± 0.0	91.1 ± 0.1	44.8 ± 0.2
GMoE-S/16 (Ours)	**90.3 ± 0.1**	**93.7 ± 0.2**	**46.6 ± 0.4**

MACs	Paint	Clipart	Info	Paint	Quick	Real	Sketch	IID Imp.	OOD Imp.
4.1G	ResNet50	37.1	12.9	62.7	2.2	49.3	33.3	-	-
7.9G	ResNet101	40.5	13.1	63.4	3.1	51.2	35.4	1.1%	12.4%
4.6G	ViT-S/16	42.7	15.9	69.0	5.0	**56.4**	37.0	10.0%	38.6%
4.8G	GMoE-S/16	**43.5**	**16.1**	**69.3**	**5.3**	**56.4**	**38.0**	**10.5%**	**42.3%**

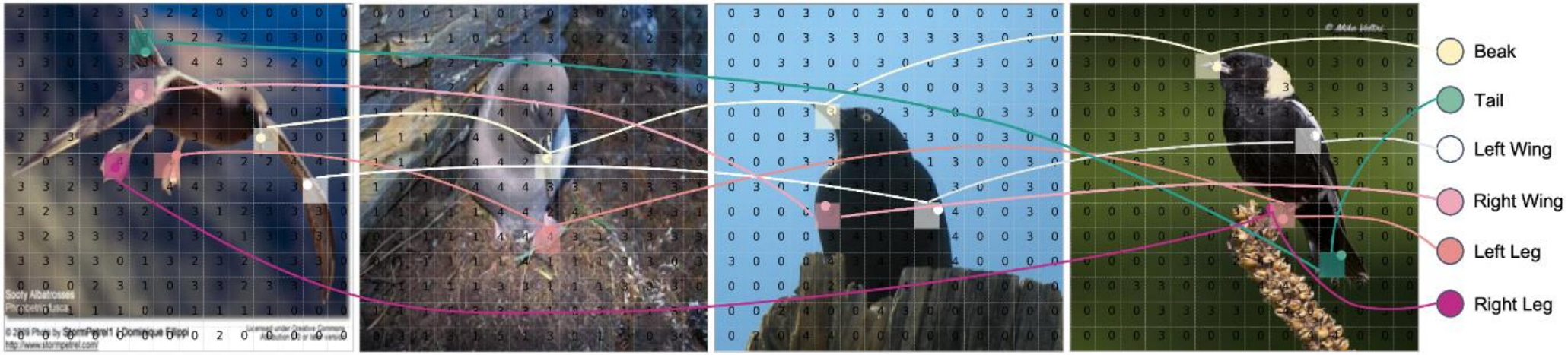
Experiments: Expert Selection



(b) Visual attributes.



(c) Expert selection.



A Stronger ResNet50

Table 14: Comparison of different training recipes on ResNet-50 V2 and ViT-S/16.

Recipe	LR Opt.	TrivialAug.	Ep.	Rand Er.	Label Sm.	FixRes Mt.	WDT	IRT	IN1K
ResNet-50 V2	✔	✔	600	✔	✔	✔	✔	✔	80.8
ViT-S/16	✔	✗	300	✔	✗	✗	✗	✔	79.9

Table 15: Train-validation selection performance comparison for ViT, GMoE, and other DG algorithms with ResNet-50 V2 as backbone model. On DomainNet, results are reported with 15K iterations.

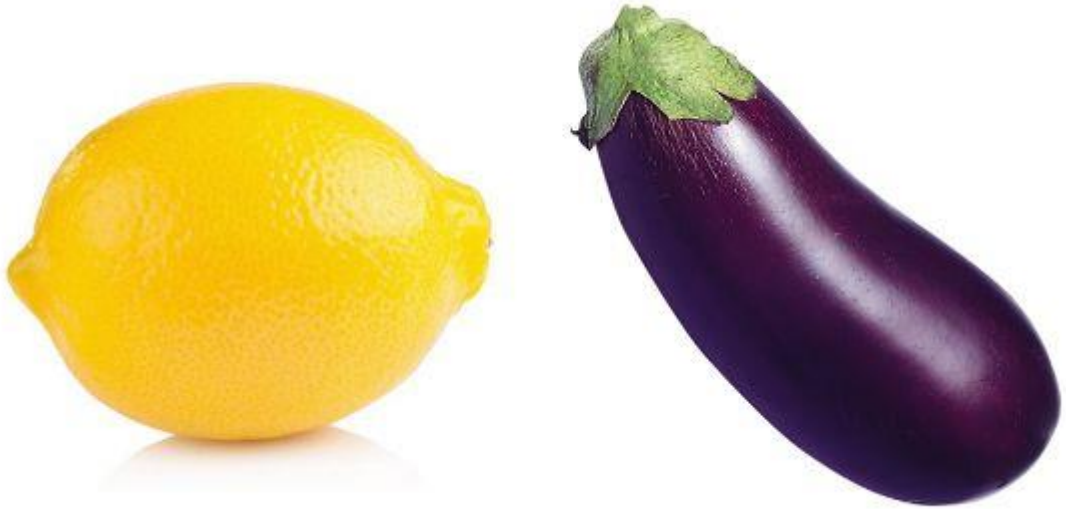
Algorithm	PACS	VLCS	OfficeHome	TerraInc	DomainNet
ERM (w/ ResNet-50 V2)	87.2	78.2	68.7	49.9	45.3
Fishr (w/ ResNet-50 V2)	87.5	77.9	70.4	**51.7**	47.0
ViT-S/16	86.2	79.7	72.2	42.0	47.1
GMoE-S/16	**88.1**	**80.1**	**74.2**	48.5	**48.7**

Computational Cost Comparison

Table 16: Comparison of training/inference iteration time and run-time memory for a mini-batch. A mini-batch is formed with 160 images in 224×224 resolutions from DomainBed. For both metrics, lower is better.

Training	**ERM**	**DANN**	**IRM**	**Fish**	**Fishr**	**SWAD**	**VIT-S/16**	**GMoE-S/16**
Step Time (s) ↓	1.01	1.02	1.10	2.79	1.10	1.21	0.90	0.98
Run-time Memory (GB) ↓	13.40	13.42	13.40	3.41	15.25	14.32	11.15	12.28
Inference	**ERM**	**DANN**	**IRM**	**Fish**	**Fishr**	**SWAD**	**VIT-S/16**	**GMoE-S/16**
Step Time (s) ↓	0.32	0.33	0.32	0.33	0.34	0.33	0.28	0.30
Run-time Memory (GB) ↓	1.82	1.83	1.82	1.82	1.83	1.84	0.76	1.05

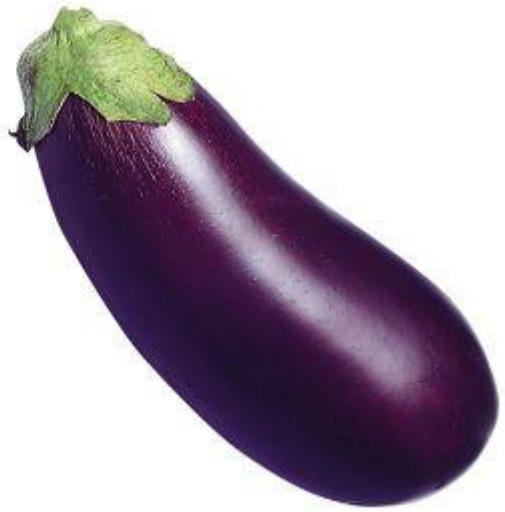
When are Lemons Purple? The Concept Association Bias of CLIP

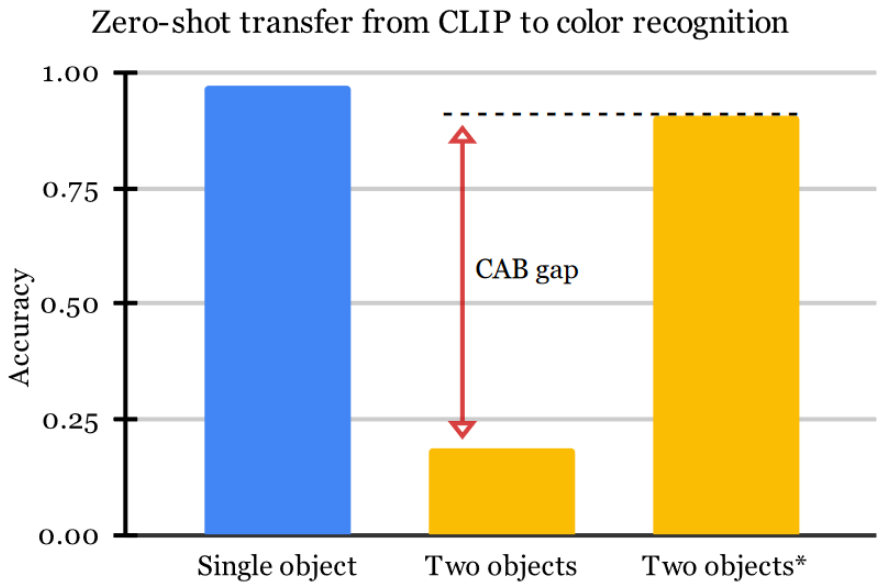


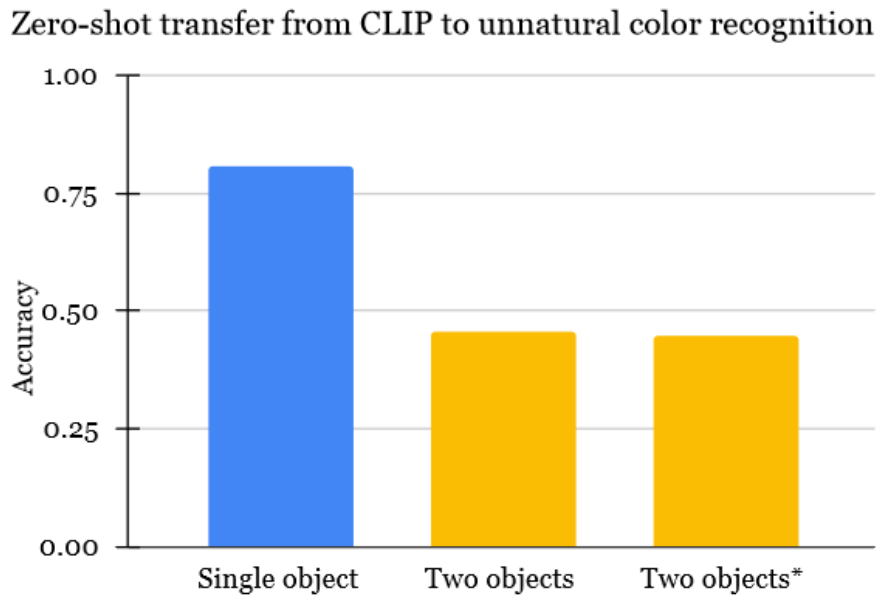
Q: The color of the lemon is [mask].

CLIP’s answer: Purple

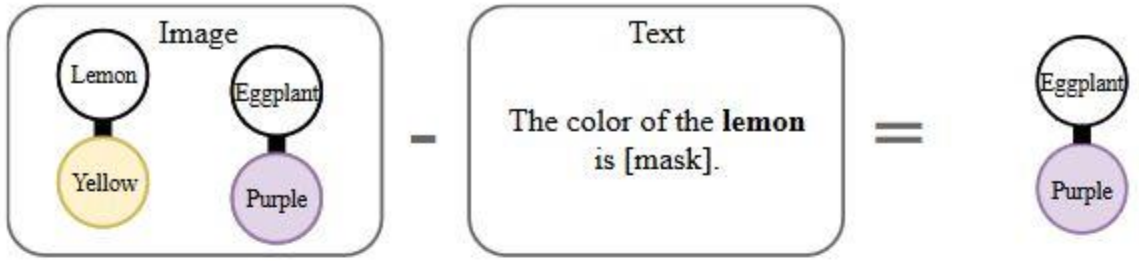
When are Lemons Purple? The Concept Association Bias of CLIP



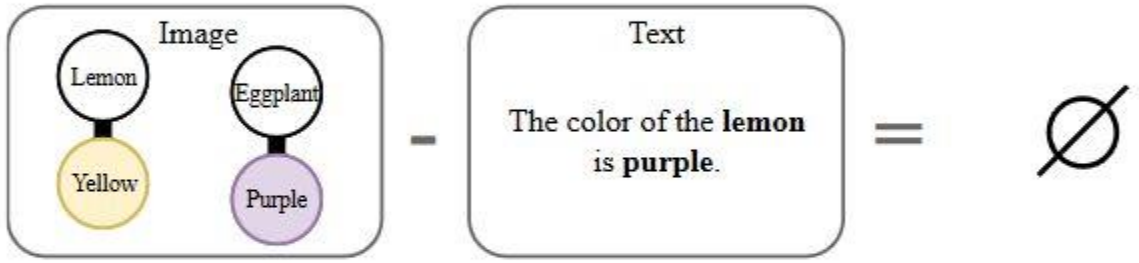




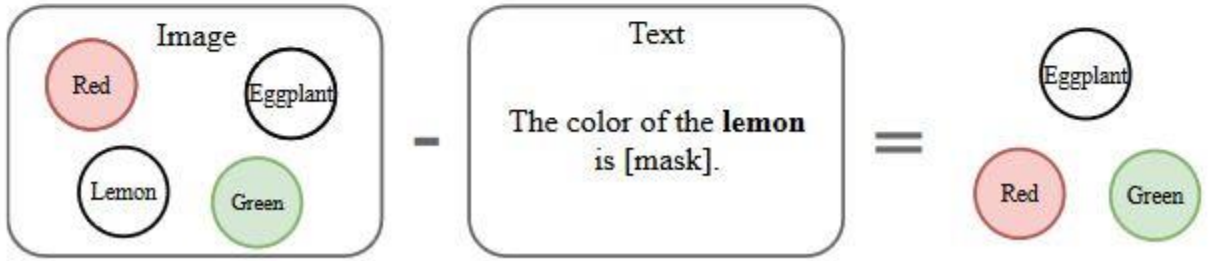
When are Lemons Purple? The Concept Association Bias of CLIP



(a) Natural color

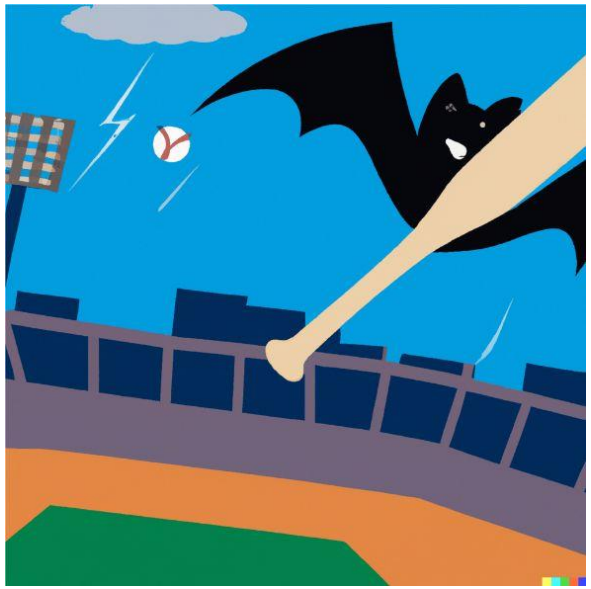


(b) Natural color ([mask] = purple)

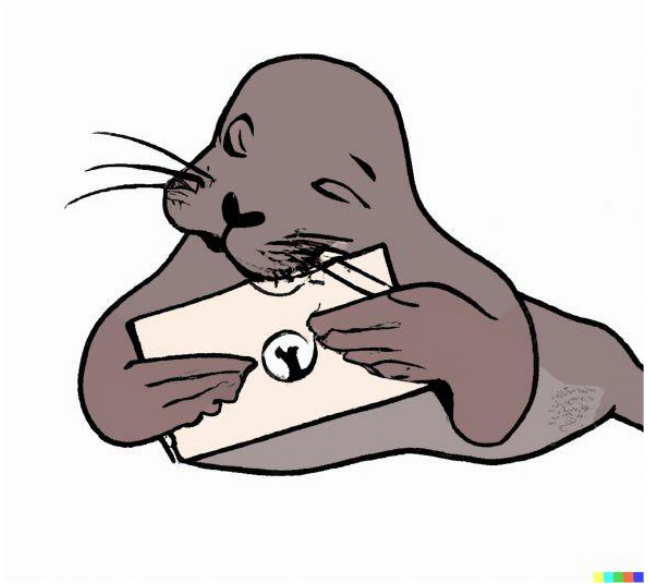


(c) Unnatural color

DALLE-2 is Seeing Double: Flaws in Word-to-Concept Mapping in Text2Image Models



a bat is flying over
a baseball stadium



a seal is opening a letter

a bass lounging in
a tropical resort in space

DALLE-2 is Seeing Double: Flaws in Word-to-Concept Mapping in Text2Image Models

a fish and a gold ingot

a fish and an ingot

a zebra and a street

a zebra and a gravel street