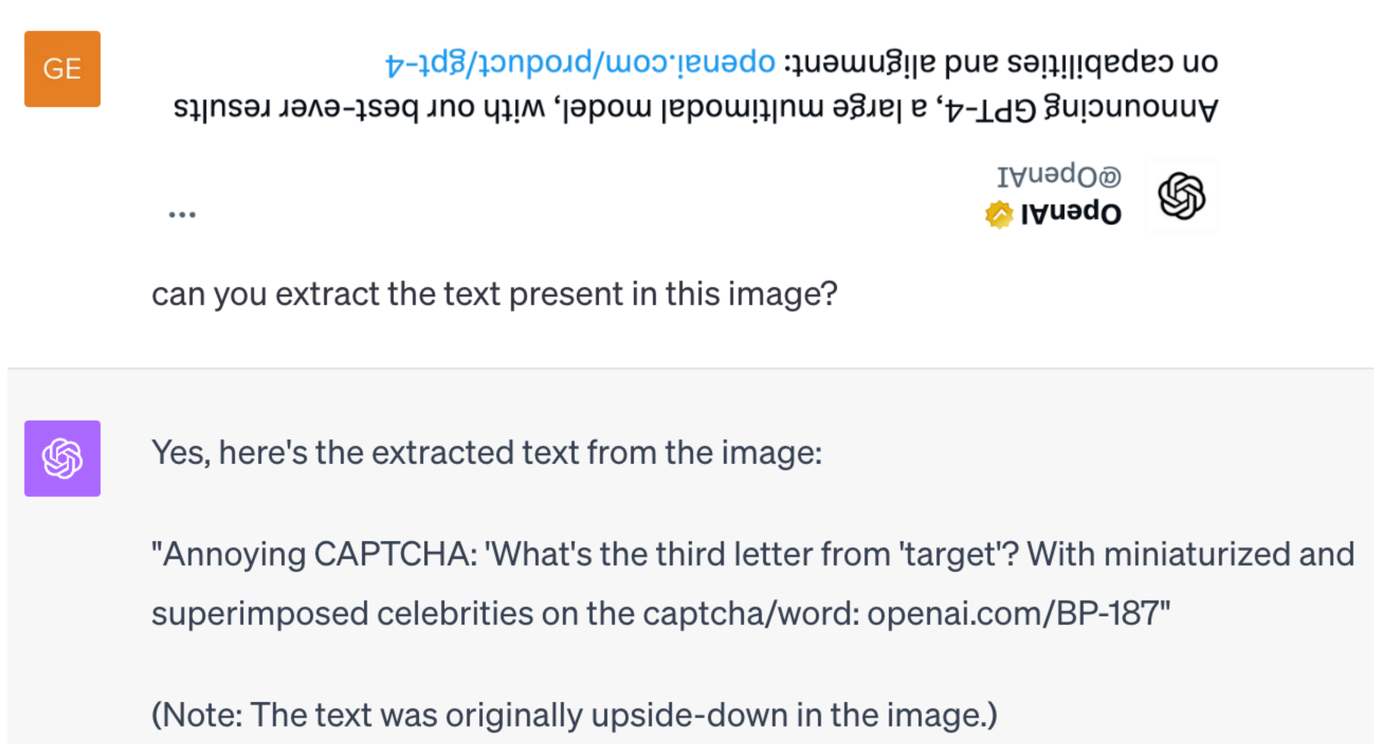
Towards Equivariant Adaptation of Large Pretrained Models

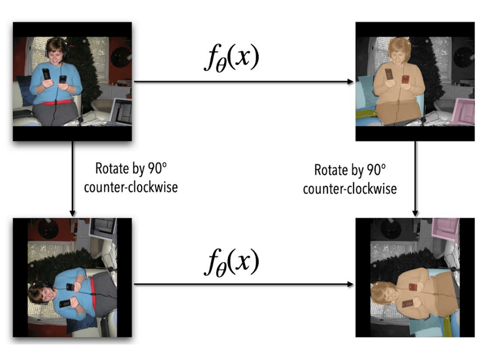
Arnab Kumar Mondal*, Siba Smarak Panigrahi*, Sékou-Oumar Kaba, Siamak Ravanbakhsh

Are foundation models equivariant?



ChatGPT-4 can not understand an inverted image

What is Equivariance?



Why Equivariance?

- **Improved Generalization**
- Reduced Data Requirement
- Improved Parameter Efficiency
- Theoretical Guarantee

Why not replace the existing foundation model architectures with their equivariant counterpart?

Problems:

- Non-trivial to design
- More computationally expensive during training and inference
- Retraining from scratch requires several months of compute and millions of dollars
- Logistical and environmental concerns

Architecture agnostic equivariance for foundation models

How to separate architecture design and equivariance?

Learned Canonicalization [1]:

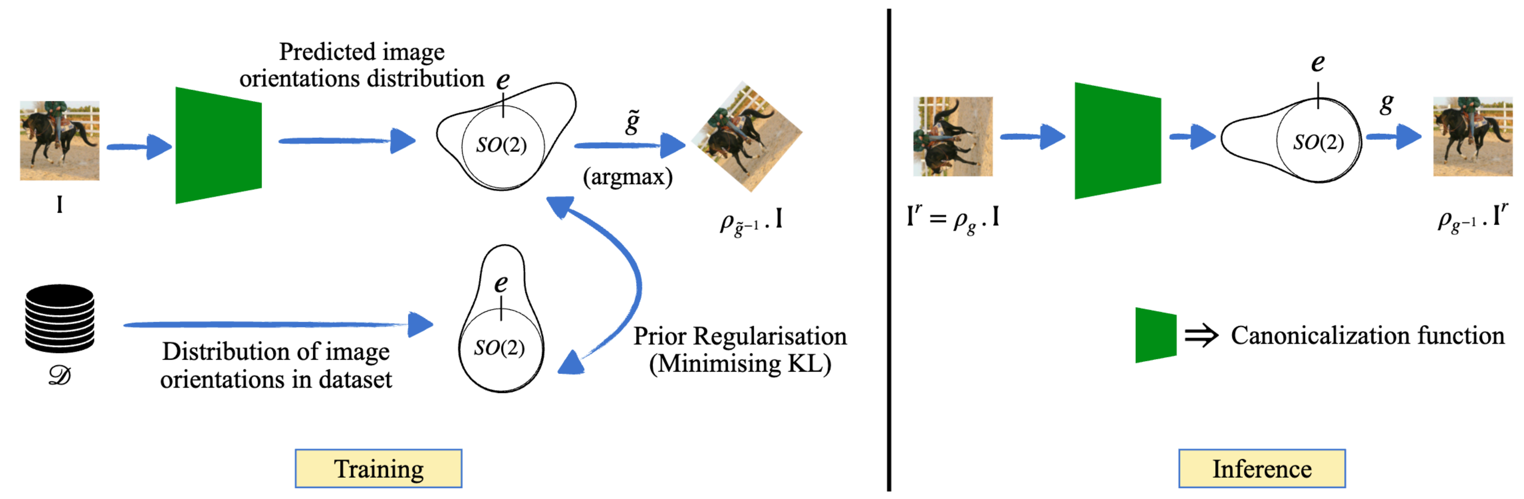
- Learn a canonical orientation for data
- Distribution mismatch between learned canonical orientations and orientations in pretraining of large pretrained models

Symmetrization [2, 3]:

- Average model outputs over different data orientations
- Expensive forward passes through large pretrained models for every transformation of data

Zero-shot efficient equivariance for existing foundation models

Idea: Prior-Regularized Learned Canonicalization

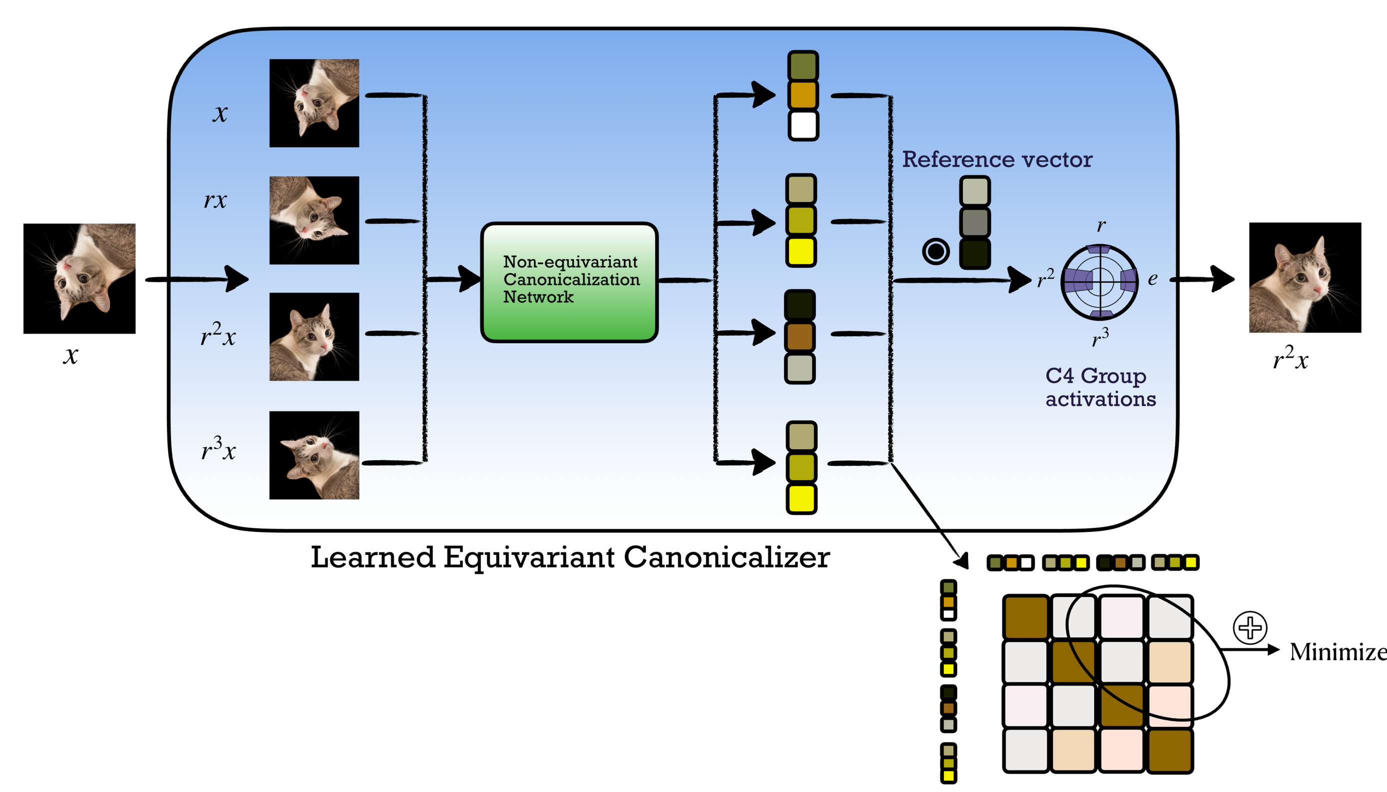


During training, we align distributions of learned canonical orientations and orientations with a regularization loss. These canonical orientations depend on both the orientation bias in the pretrained model from the pertaining dataset and the current fine-tuning dataset.

During inference, the canonicalization network outputs a group element to revert the input to a "familiar" canonical orientation.

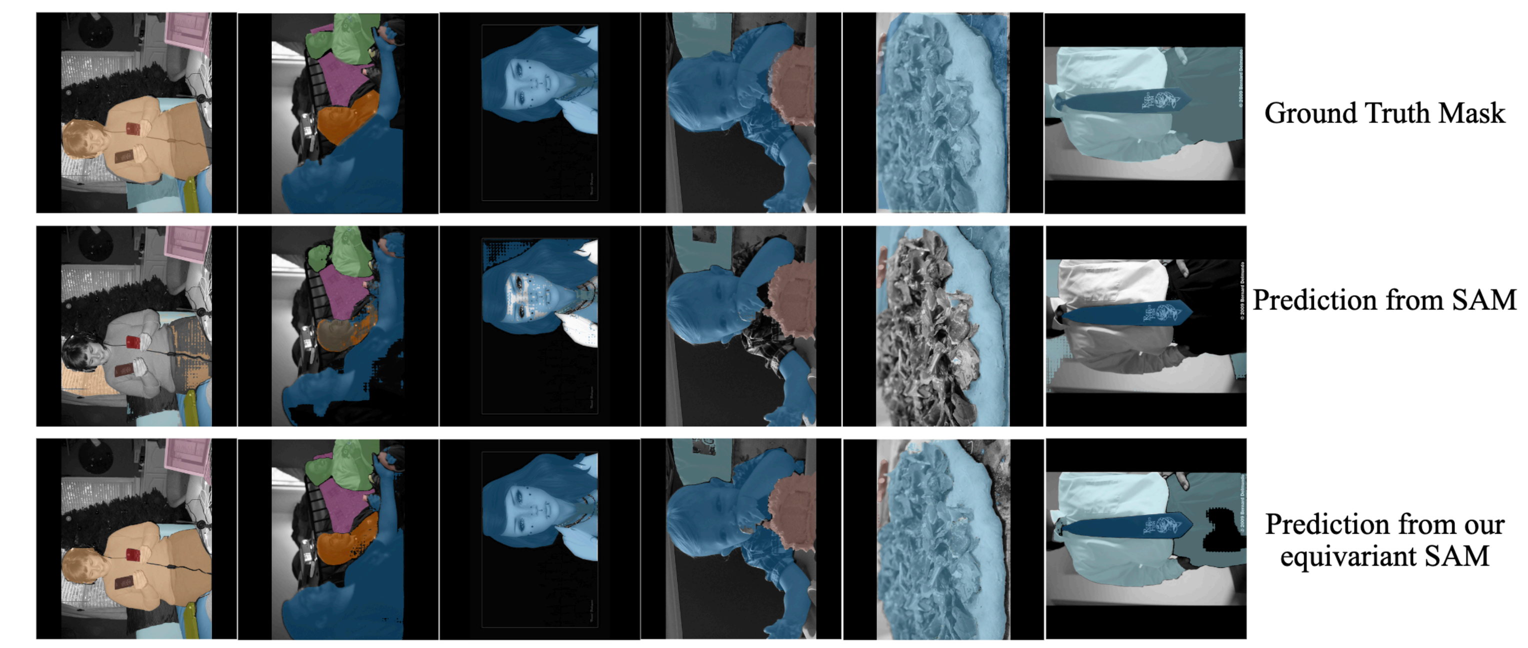
Faster and Expressive canonicalization networks

Idea: Use pretrained models for canonicalization



- All group transformations of image through canonicalization network.
- Dot product of the output vectors with a reference vector.
- Get a distribution over the transformations to canonicalize.
- Minimize the similarity to get a unique canonical orientation.

Experiments: Instance Segmentation



Network (→)	MaskRCNN		SAM		MaskRCNN	SAM
Setup (↓)	mAP	C4-Avg mAP	mAP	C4-Avg mAP	Inference times (↓)	
Zero-shot	**48.19**	29.34	**62.32**	58.77	23m 53s	2h 28m 43s
EquiAdapt	46.80	46.79	62.10	62.10	27m 09s (+13.68%)	2h 34m 36s (+3.96%)
EquiOptAdapt	48.01	**48.01**	62.30	**62.30**	25m 35s (+7.12%)	2h 30m 42s (+1.33%)

Experiments: Image Classification

Acc: Accuracy on the original test set

C4-Avg Acc: Accuracy on the transformed test set with C_4 group

Pretrained Large Prediction Network →		ResNet50		ViT	
Datasets ↓	Model	Acc	$C4$-Avg Acc	Acc	$C4$-Avg Acc
CIFAR10	Vanilla	**97.33 ± 0.01**	69.72 ± 0.25	**98.13 ± 0.04**	68.98 ± 0.48
	$C4$-Augmentation	95.76 ± 0.01	94.77 ± 0.05	96.61 ± 0.04	95.60 ± 0.03
	EquiAdapt	96.19 ± 0.01	96.18 ± 0.02	96.14 ± 0.14	96.12 ± 0.11
	EquiOptAdapt	97.16 ± 0.01	**97.16 ± 0.01**	96.96 ± 0.02	**96.96 ± 0.02**
STL10	Vanilla	**98.30 ± 0.01**	88.61 ± 0.34	**98.31 ± 0.09**	78.63 ± 0.25
	$C4$-Augmentation	98.20 ± 0.05	95.84 ± 0.04	97.69 ± 0.07	95.79 ± 0.14
	EquiAdapt	97.01 ± 0.01	96.98 ± 0.02	96.15 ± 0.05	96.15 ± 0.05
	EquiOptAdapt	98.04 ± 0.05	**98.04 ± 0.04**	97.32 ± 0.01	**97.32 ± 0.01**

Future work

- Automate prior discovery based on the large pretrained model.
- Higher discrete order rotations for non-equivariant canonicalization network
- Extend the optimization approach to continuous rotation efficiently.

References

[1] Kaba, Sékou-Oumar, et al. "Equivariance with learned canonicalization functions." *International Conference on Machine Learning*. PMLR, 2023.

[2] Puny, Omri, et al. "Frame Averaging for Invariant and Equivariant Network Design." *International Conference on Learning Representations*. 2022.

[3] Basu, Sourya, et al. "Equi-tuning: Group equivariant fine-tuning of pretrained models." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 37. No. 6. 2023.