

When and How Does CLIP Enable Domain and Compositional Generalization?

Elias Kempf*



**Simon
Schrodi***



**Max
Argus**

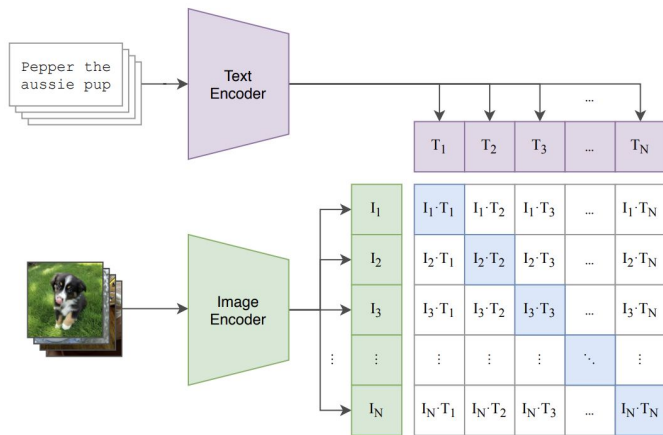


**Thomas
Brox**



*** Equal contribution**
University of Freiburg

CLIP is robust to natural distribution shifts



[Radford et al, ICML 2021]

Natural images



Sketches



Paintings



cat

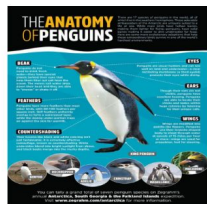
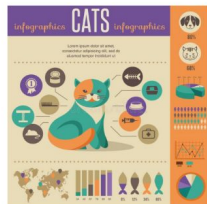
→ CLIP's training distribution is the cause [Fang et al, ICML 2022]

Controlling for the training distribution

Clipart



Infograph



Painting



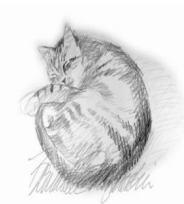
Quickdraw



Natural



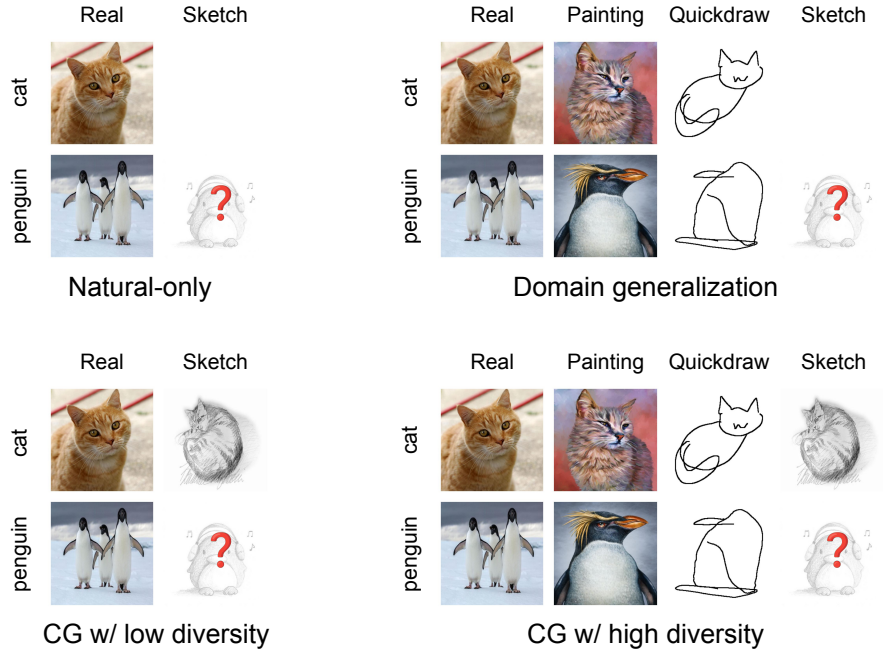
Sketch



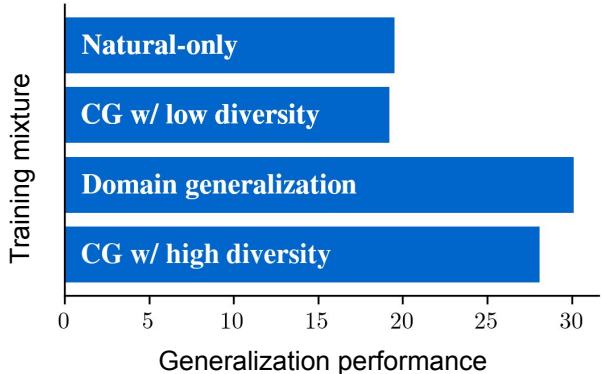
[DomainNet; Peng et al, ICCV 2019]

Domain diversity drives CLIP's robustness

Domain Mixtures for CLIP Training



Unseen domain-class combinations



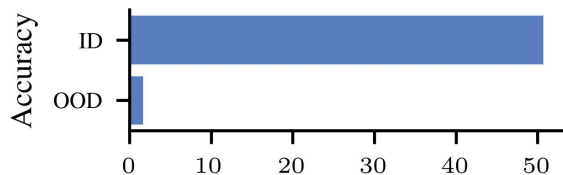
CLIP can fail to generalize to certain domains

**CLIP fails to generalize
to Quickdraw images!**

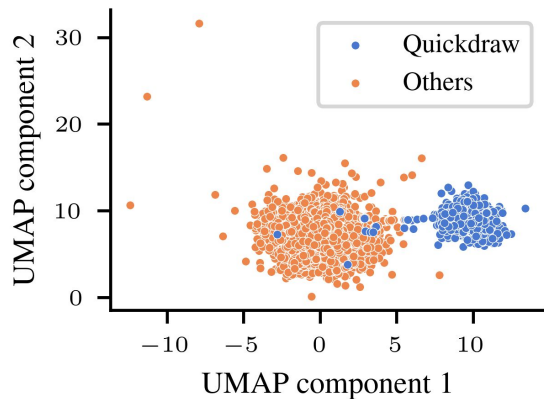
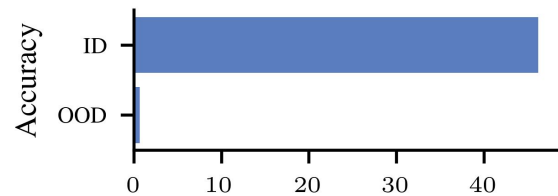


[DomainNet; Peng et al, ICCV 2019]

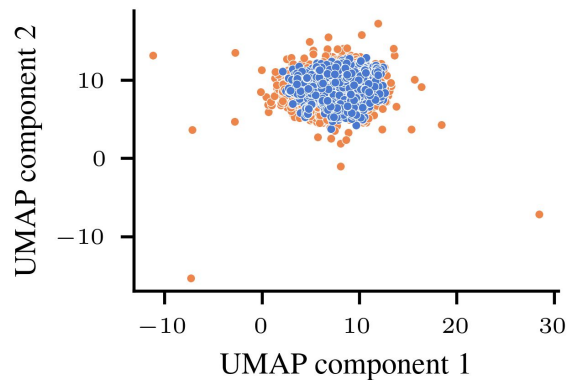
Why does CLIP not generalize to Quickdraw?



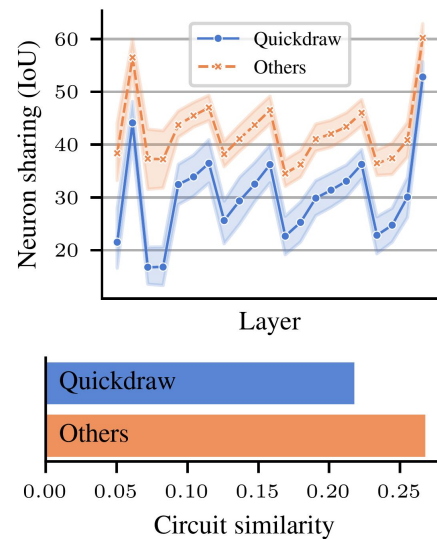
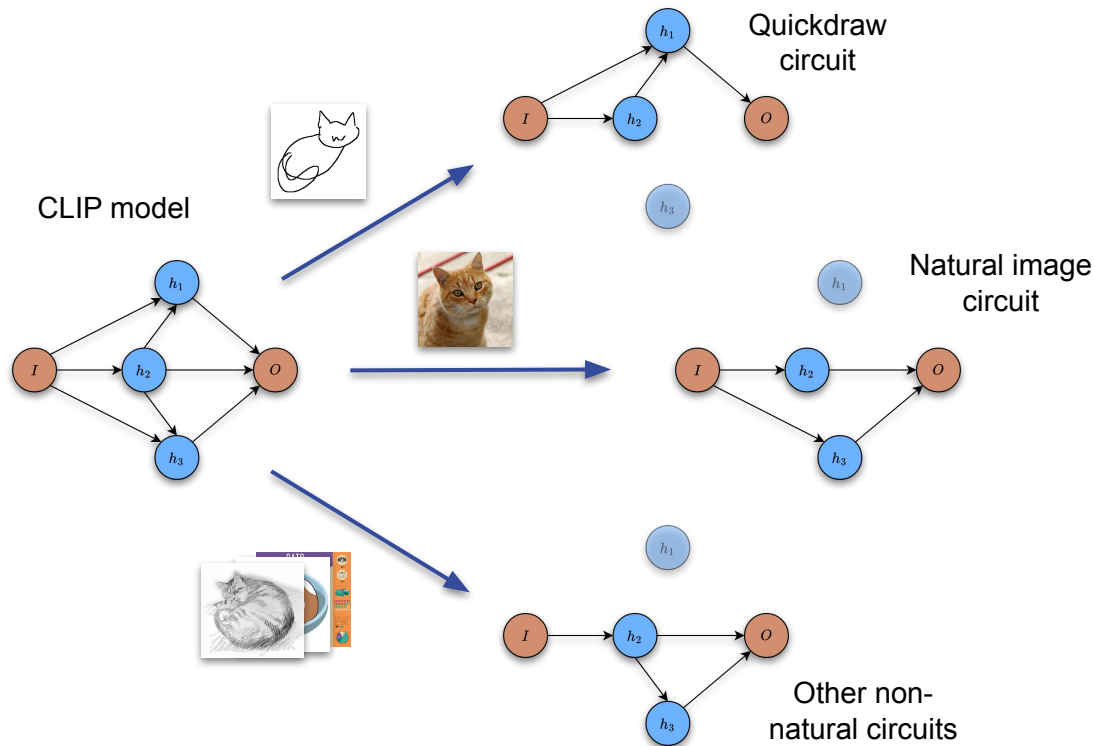
Alignment improves,
generalization does not ↓



adapt captions to
avoid separation

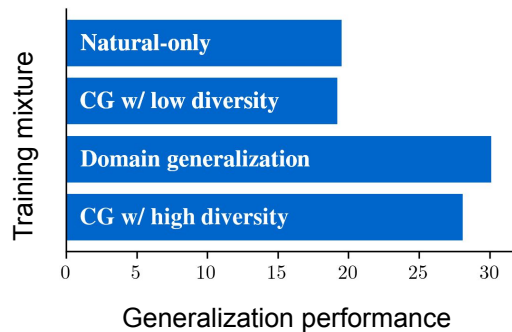


Generalization requires shared circuitry



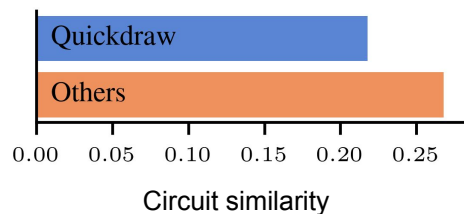
Quickdraw circuits are the least similar to other domains!

Summary



Domain diversity drives CLIP's robustness

CLIP sometimes generalizes better to entirely unseen domains

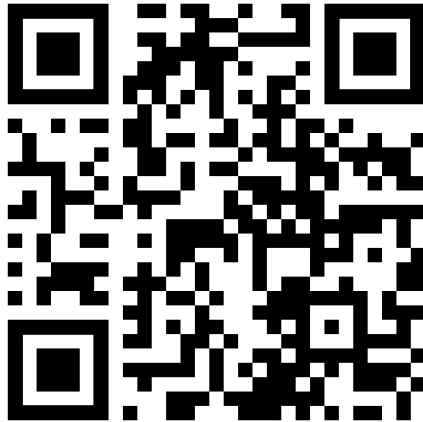


Generalization requires sufficiently shared circuitry across domains

Thank You For Your Attention!



Paper



Code

