FoMo: Forking Moment in Generative Trajectory as a Perceptual Distance

Jaihyun Lew · Mingi Jung · Minjun Park · Wooseok Song · Sungroh Yoon† (†Corresponding Author)

Project page
jhlew.github.io/FoMo
`pip install fomo-iqa`
NeurIPS 2026

DSAIL Data Science & AI Laboratory

NEURAL INFORMATION PROCESSING SYSTEMS

1 Human labels are scarce

KADID-10K
10,125 images, 30 ratings each

480k FoMo pairs

0 human ratings

Area = dataset size

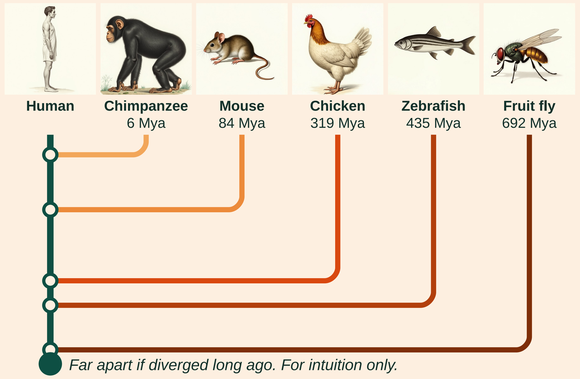
2 Diffusion labels itself



Reference | Re-noise to step s | Denoise with FLUX

Label = s, no annotator

Side note: an unofficial, conceptual analogy

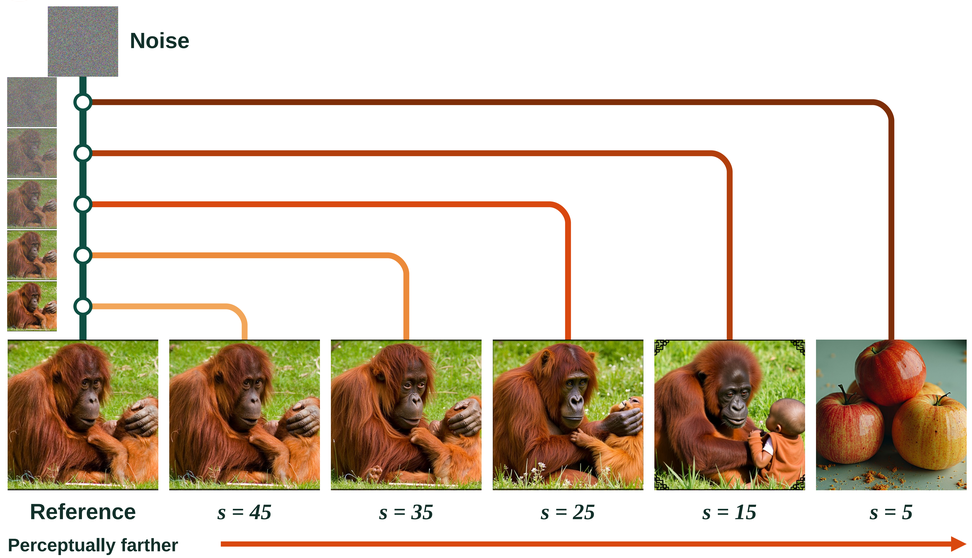


Far apart if diverged long ago. For intuition only.

Noise ≈ common ancestor **Fork step** ≈ divergence time

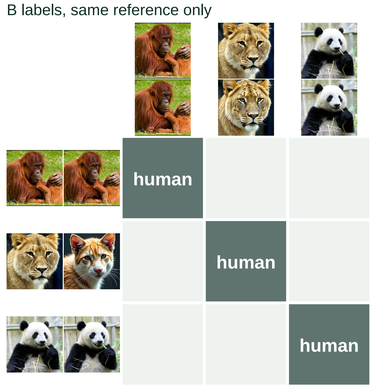
Times: TimeTree, square-root scale. Illustrations: FLUX.1-dev.

3 Fork early: far apart. Fork late: close.



5 RankBCE: rank every pair

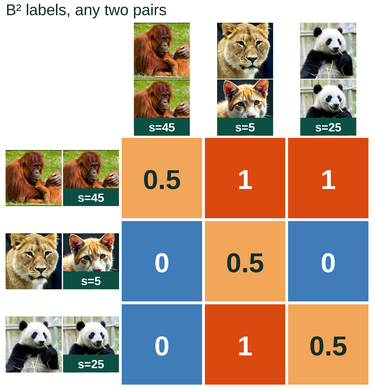
2AFC
B labels, same reference only



human vote ☐ no label

64
comparisons per batch

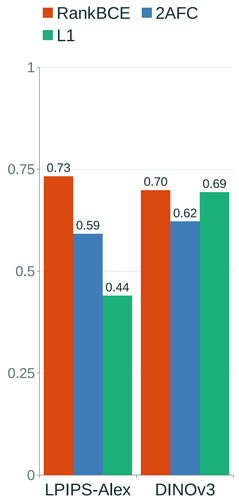
RankBCE
B² labels, any two pairs



row closer | column closer | tie

4,096
comparisons per batch

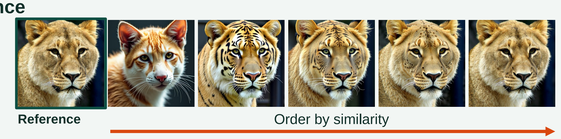
Same FoMo labels, PIPAL SROCC



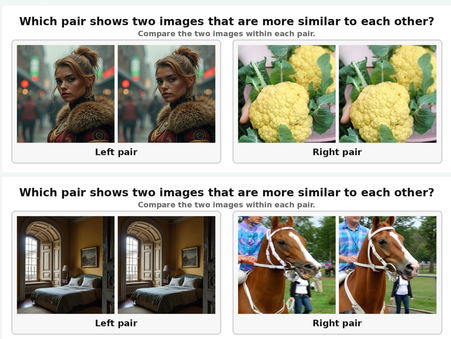
4 People agree, in two studies

Study 1: same reference

0.970
rank correlation, 1,982 annotations



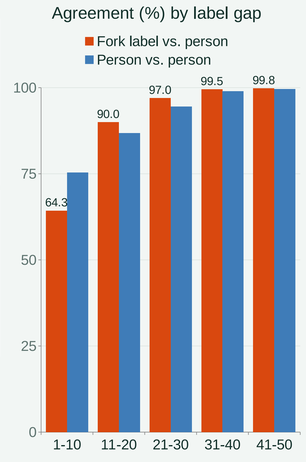
Study 2: across references



90.2%
of 5,713 annotations

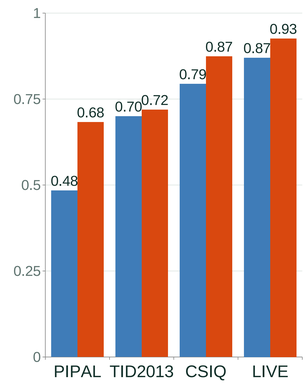
92.8%
of items, majority vote

Agreement (%) by label gap



6 Ahead of human labels on average

Mean SROCC, 7 backbones



SROCC: FoMo minus best human-labeled set

	Alex	VGG	DISTS	DINOv3	CLIP	MAE	DreamSim
PIPAL	+0.111	+0.048	+0.051	+0.374	+0.332	+0.294	+0.016
TID2013	−0.008	−0.094	−0.143	+0.100	+0.217	+0.048	−0.014
CSIQ	−0.007	−0.037	−0.020	+0.195	+0.297	+0.097	−0.018
LIVE	−0.004	−0.006	+0.004	+0.132	+0.098	+0.110	+0.003

FoMo ahead | Human labels ahead

7 / 7
PIPAL backbones

12 / 12
frozen-ViT cells

18 / 28
all cells