Beyond I-Con: A Roadmap for Representation Learning Loss Discovery

Jasmine Shone*, Zhening Li*, Shaden Alshammari, Mark Hamilton, William Freeman

We generalize the I-Con objective by replacing the KL divergence with **any positive definite divergence D**, enabling the discovery of **new loss functions** with distinct properties.

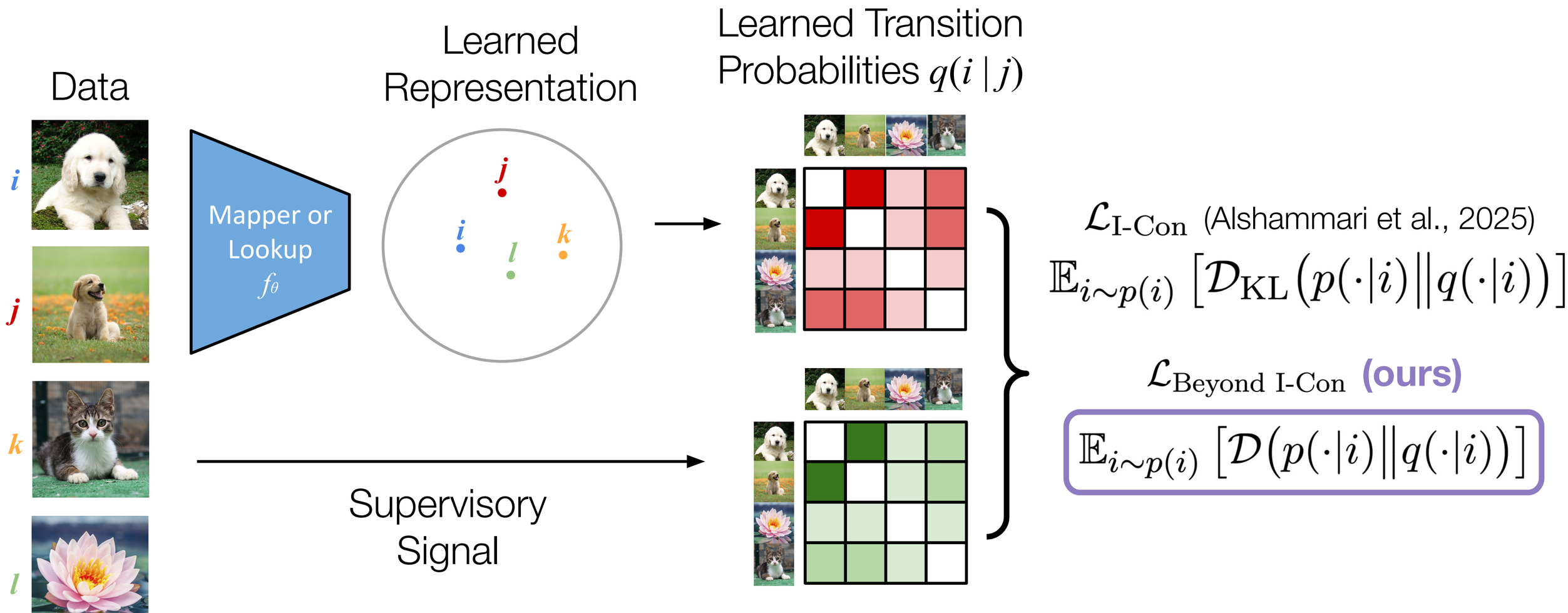

Figure adapted from I-Con (Alshammari et al., 2025)

Results

TV outperforms state of the art on ImageNet-1K clustering with DiNO ViT-B and ViT-L embeddings.

Method	DiNO ViT-S/14	DiNO ViT-B/14	DiNO ViT-L/14
k-Means	51.84	52.26	53.36
TEMI (Adaloglou et al., 2023)	56.84	58.62	—
Debiased InfoNCE Clustering	**57.8 ± 0.26**	64.75 ± 0.18	67.52 ± 0.28
JSD	53.50	63.80	66.60
TV	55.90	**65.13 ± 0.13**	**68.40 ± 0.29**
Hellinger	54.90	63.80	67.85

JSD and Hellinger outperform KL on downstream classification with SupCon-learned features (CIFAR-10).

Divergence	Linear probe test acc.	k-NN ($k = 7$) test acc.
KL	90.03 ± 0.14	89.61 ± 0.13
TV	83.23 ± 0.18	82.95 ± 0.16
Hellinger	90.47 ± 0.08	90.40 ± 0.09
JSD	**90.84 ± 0.11**	**90.62 ± 0.11**

TV/Hellinger/JSD gives better SNE visualization than KL on CIFAR-10