


Robust Representation Learning via Perceptual Similarity Metrics

Saeid Taghanaki*, Kristy Choi*, Amir Khasahmadi, Anirudh Goyal

Autodesk AI, Stanford University, MILA

ICML 2021

Background

Given one or a handful of examples, we are typically able to learn a concept and apply it across a variety of tasks and conditions (***generalization***).

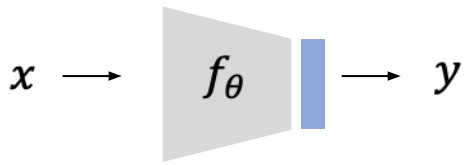
This becomes a challenge for machine learning models as they tend to overfit to ***spurious input features*** which results in failure in case of domain shift and low performance for rare subgroups present in data.

A wide range of methods tackle this problem by regularization, data augmentation, leveraging causal explanations, and self-training, however, they often require ***privileged information*** such as defining rare subgroups beforehand which is not trivial in a large dataset.

Inspired by the “robustness” of the human visual system, perceptual similarity metrics, and metric learning, we propose ***Contrastive Input Morphing (CIM)***. CIM has a small auxiliary network which is trained with a **triplet loss** that computes the **perceptual similarity** between sets of ***transformed inputs***, ***positive examples***, *and* ***negative examples***.

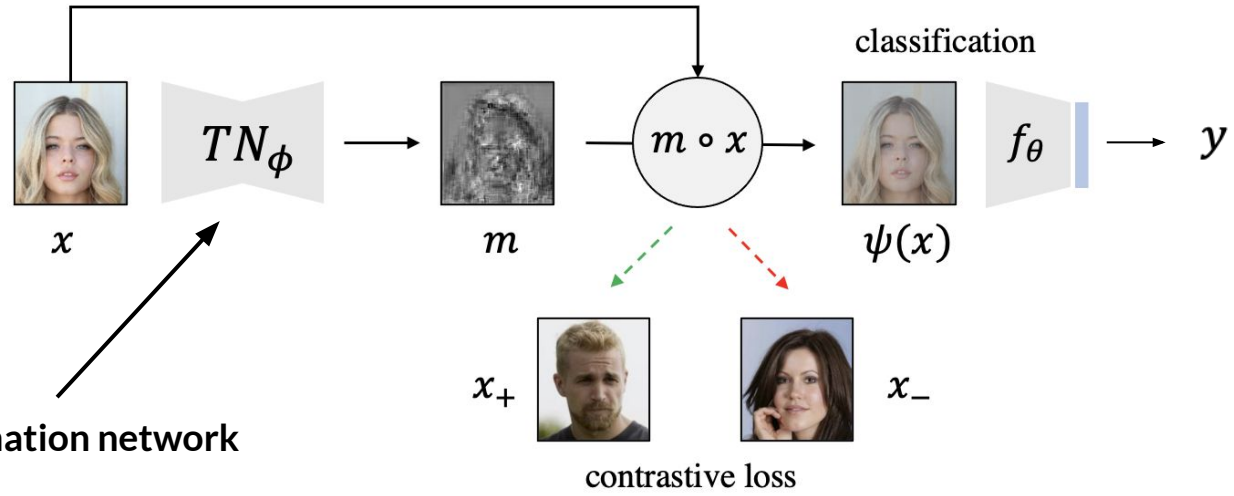
Contrastive Input Morphing (CIM)

vanilla classifier



$$\mathcal{L}_{\text{sup}}(\theta) = \mathbb{E}_{x,y\sim p_{\text{data}}(x,y)}[\ell(f_\theta(x), y)]$$

CIM



contrastive loss via MS-SSIM

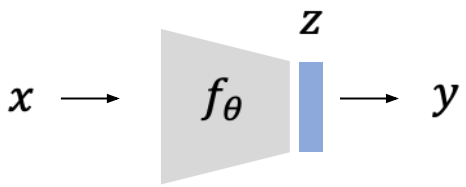
$$\mathcal{L}_{\text{con}}(\phi) = \min_{\phi} \text{MS}(\psi(x), x_+) - \text{MS}(\psi(x), x_-)$$

$$\mathcal{L}_{\text{CIM}}(\phi, \theta) = \lambda \mathcal{L}_{\text{con}}(\phi) + \mathcal{L}_{\text{sup}}(\theta)$$

Wang, Bovik, Sheikh, Simoncelli, *Image quality assessment: from error visibility to structural similarity.* IEEE 2004

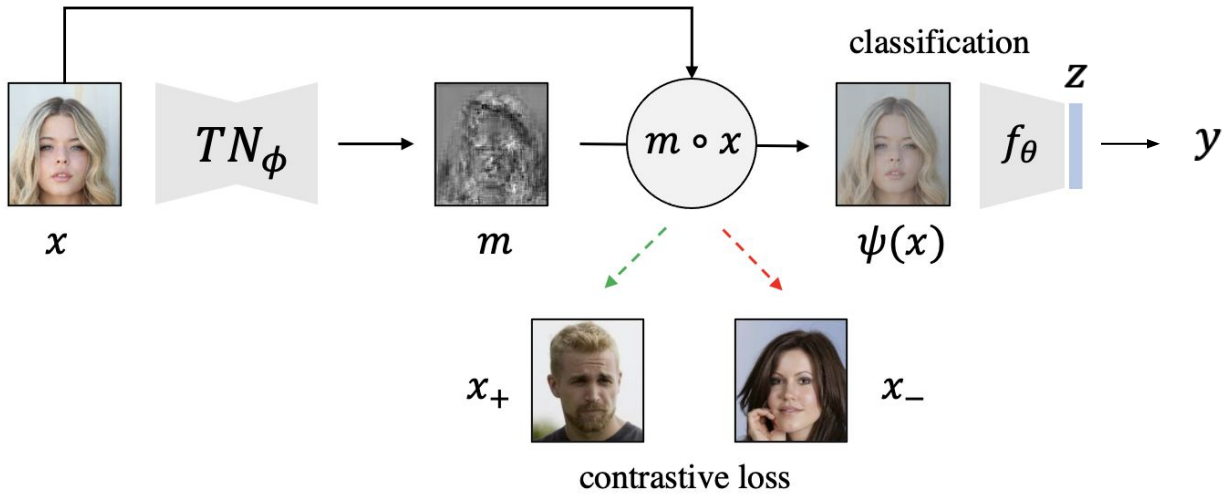
CIM + Variational Information Bottleneck (VIB)

VIB



$$\mathcal{L}_{\text{VIB}}(\theta) = \mathcal{L}_{\text{sup}}(\theta) + D_{\text{KL}}(Q(Z|X)||P(Z))$$

CIM + VIB

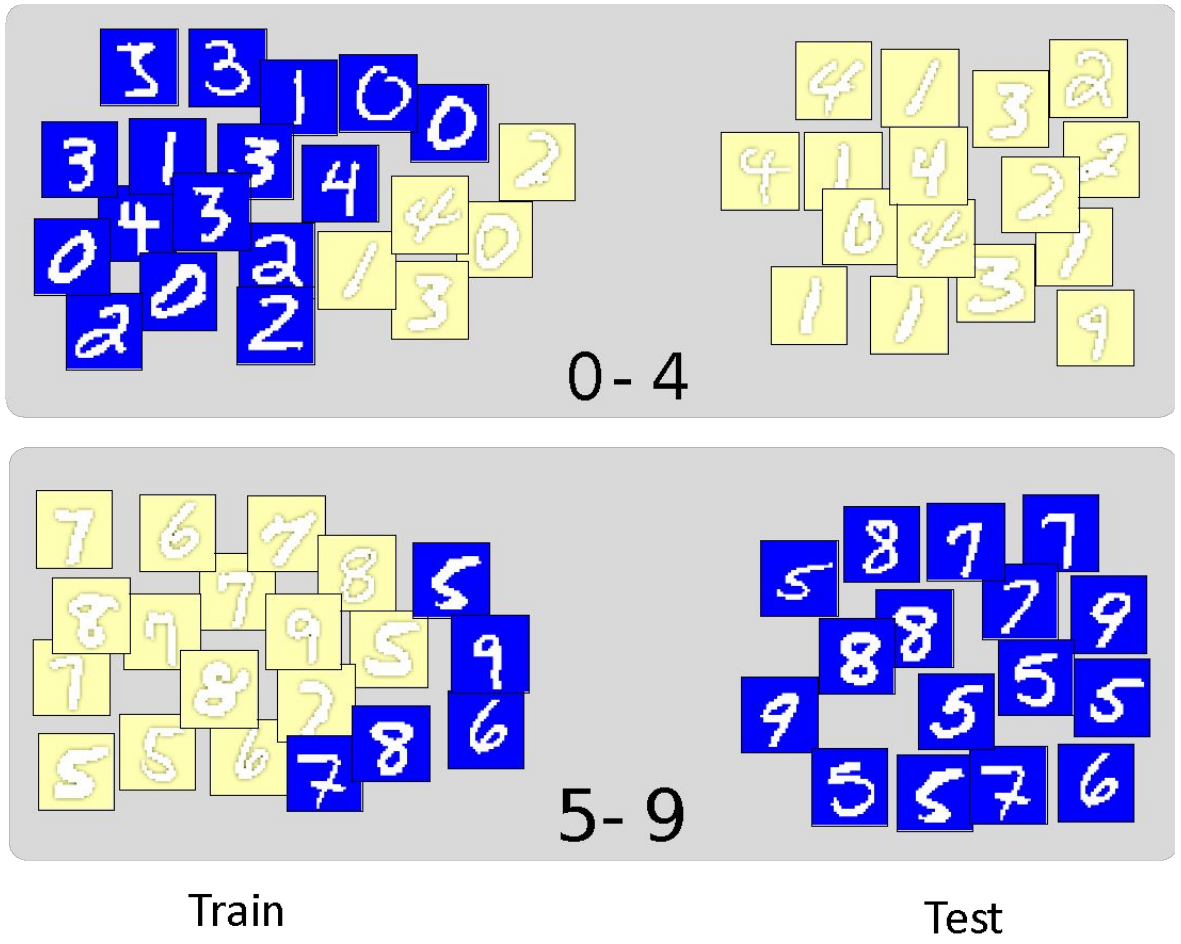


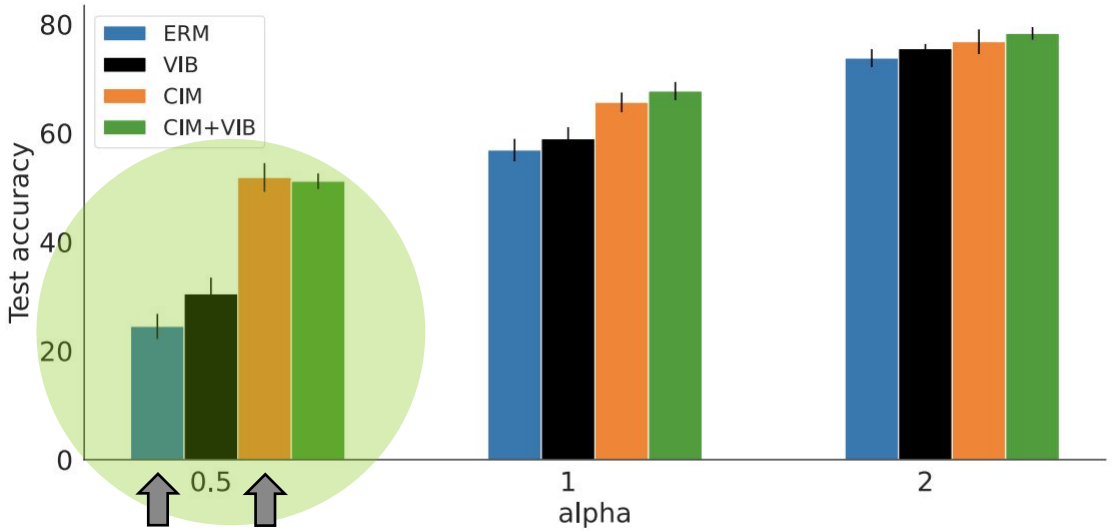
$$\mathcal{L}_{\text{CIM+VIB}}(\phi, \theta) = \lambda\mathcal{L}_{\text{con}}(\phi) + \mathcal{L}_{\text{sup}}(\theta) + D_{\text{KL}}(Q(Z|X)||P(Z))$$

CIM can easily be plugged in with existing mutual-information based representation learning approaches, such as VIB!

Alemi, Fischer, Dillon, Murphy, *Deep variational information bottleneck*. 2016

Classification with Nuisance Backgrounds

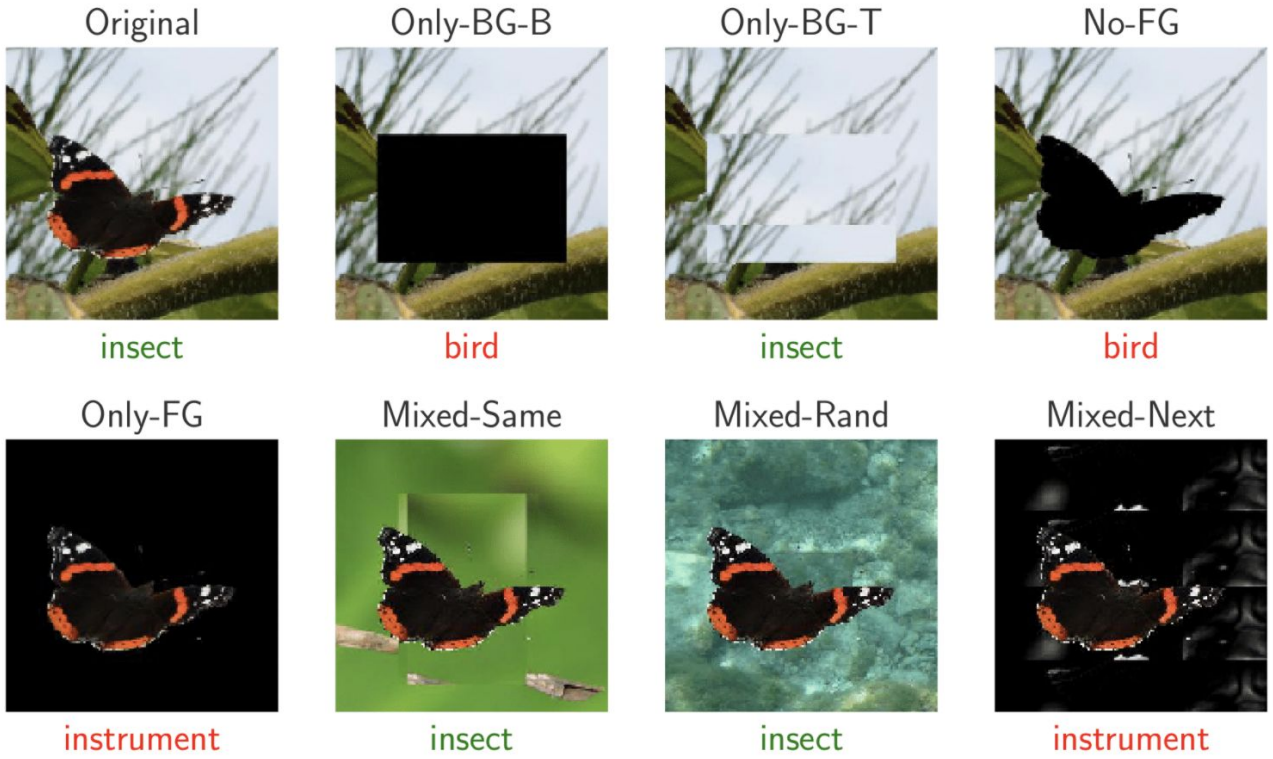




(b) Test accuracy in Colored MNIST experiments.

Learned representations rely less on background information for classifying MNIST digit classes

Background Challenge



	OR (↑)	MS (↑)	MR (↑)	BGp (↓)
Res50 (Xiao et al., 2020)	96.3	89.9	75.6	14.3
VIB (Alemi et al., 2016)	97.4	89.9	80.5	9.4
CIM (Ours)	97.7	89.8	81.1	8.8
CIM + VIB (Ours)	**97.9**	**90.2**	**82.2**	**8.0**

CIM helps improve downstream accuracy on classification tasks with nuisance information

Xiao, Engstrom, Ilyas, Madry, *Noise or signal: The role of image backgrounds in object recognition*. 2020

Out-of-Domain Generalization (VLCS)

Caltech101

LabelMe

SUN09

VOC2007

Method	Caltech	LabelMe	Pascal	Sun	Average
DeepC (Li et al., 2018b)	87.47	62.06	64.93	61.51	68.89
CIDDG (Li et al., 2018b)	88.83	63.06	64.38	62.10	69.59
CCSA (Motiian et al., 2017)	92.30	62.10	67.10	59.10	70.15
SLRC (Ding & Fu, 2017)	92.76	62.34	65.25	63.54	70.15
TF (Li et al., 2017a)	93.63	63.49	69.99	61.32	72.11
MMD-AAE (Li et al., 2018a)	94.40	62.60	67.70	64.40	72.28
D-SAM (D'Innocente & Caputo, 2018)	91.75	57.95	58.59	60.84	67.03
Shape Bias (Asadi et al., 2019)	98.11	63.61	74.33	67.11	75.79
VIB (Alemi et al., 2016)	97.44	66.41	73.29	68.49	76.41
SCL_{E2E} (Ours)	95.56	66.72	73.16	65.10	75.14
CIM (Ours)	98.21	**67.80**	73.97	69.01	77.25
CIM + VIB (Ours)	**98.81**	66.49	**74.89**	**70.13**	**77.58**

more robust representations → improved OOD generalization performance!

Torralba and Efros, *Unbiased look at dataset bias.* CVPR 2011

Preservation of Subgroup Performance

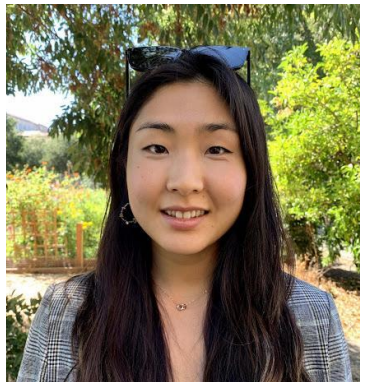
Dataset	Method	Unsupervised (subgroub-level)	Worst group acc.	Average acc.
CelebA	GDRO (Sagawa et al., 2019)	✗	**88.30**	91.80
	ERM	✓	41.10	94.80
	Baseline (Ours)	✓	70.31	93.98
	SCL_{E2E} (Ours)	✓	68.80	**95.80**
	VIB (Alemi et al., 2016)	✓	78.13	91.94
	CIM (Ours)	✓	81.25	89.24
	CIM + VIB (Ours)	✓	83.59	90.61
Waterbirds	GDRO (Sagawa et al., 2019)	✗	83.80	89.40
	CAMEL (Goel et al., 2020)	✗	**89.70**	90.90
	ERM	✓	60.00	**97.30**
	Baseline (Ours)	✓	62.19	96.42
	SCL_{E2E} (Ours)	✓	64.10	96.50
	VIB (Alemi et al., 2016)	✓	75.31	95.39
	CIM (Ours)	✓	73.35	89.78
	CIM + VIB (Ours)	✓	77.23	95.60

CIM helps preserve classification accuracy on rare subgroups of the data

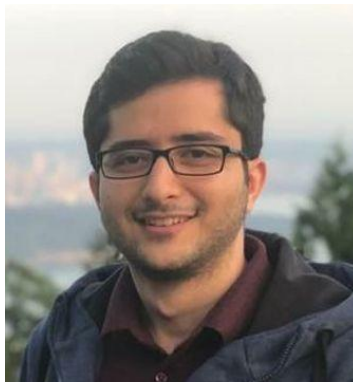
Sagawa, Koh, Hashimoto, Liang, *Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization.* ICLR 2020

Thank you!

Saeid Taghanaki



Kristy Choi



Amir Khasahmadi

Anirudh Goyal

arXiv: https://arxiv.org/pdf/2106.06620.pdf
Email: asgt.saeid@gmail.com