Learning visual representations for transfer learning by suppressing texture

Shlok Mishra[1], Anshul Shah[2], Ankan Bansal[1], Janit Anjaria[1], Jonghyun Choi[3], Abhinav Shrivastava[1], Abhishek Sharma[4], David Jacobs[1]
[1]University of Maryland, College Park [2]Johns Hopkins University [3]Yonsei University [4]Axogyan AI

BMVC 2022

Problem Addressed & Contributions

Suppressing texture improves the transfer learning performance.

- CNN overemphasize on texture at expense of learning shape (high-level information).
- Analysed a few techniques to suppress texture.
- Empirically, Anisotropic Diffusion gave the best results likely due to its edge preserving denoising. Retaining edges is important for downstream tasks.
- We show improved performances across self-supervised learning and supervised learning on various datasets and frameworks.

Overemphasis on Texture

(a) Texture image
81.4% Indian elephant
10.3% indri
8.2% black swan

(b) Content image
71.1% tabby cat
17.3% grey fox
3.3% Siamese cat

(c) Texture-shape cue conflict
63.9% Indian elephant
26.4% indri
9.6% black swan

Texture suppressing methods

Object Detection & Semantic Segmentation

Methods	Dataset	AP_{50}	AP	AP_{75}	mIoU (SS)
Stylized ImageNet		43.5	28.80	33.7	-
Supervised ImageNet		81.6	54.2	59.8	59.8
MoCo V2	ImageNet	82.4	57.0	63.6	67.5
MoCo V2 Anistropic (Ours)	ImageNet	**83.7**	**58.2**	**64.8**	**67.8**
Dense-CL	ImageNet	82.8	58.7	65.2	69.4
Dense-CL Anistropic (Ours)	ImageNet	**83.5**	**59.6**	**66.4**	**70.5**
Dense-CL CC	COCO	81.7	56.7	63.0	67.5
Dense-CL CC Anistropic (Ours)	COCO	**83.1**	**57.9**	**64.2**	**68.6**

Supervised Learning

Method	# Iterations	Top-1 Acc	Top-5 Acc	Object Detection
Baseline Supervised	-	76.13	92.98	70.7
Stylized ImageNet	-	76.72	93.27	75.1
Perona Malik with Pix2Pix	20	**76.95**	**93.36**	**75.21**
Perona Malik	20	76.71	93.26	74.37
Perona Malik	50	76.32	92.96	73.80
Robust AD	20	76.58	92.96	73.33
Robust AD	50	76.64	93.09	73.57
Gaussian Blur	-	76.21	92.64	73.26
Cartoon ImageNet	-	76.22	93.12	72.31
Bilateral ImageNet	-	75.99	92.90	71.34

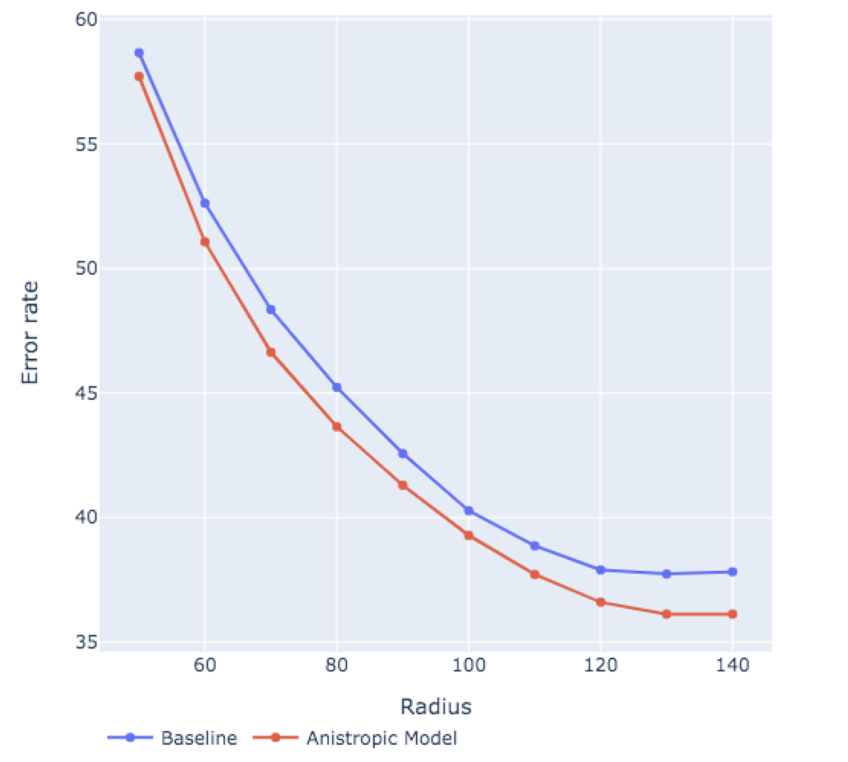
Learning Better Shape Representations

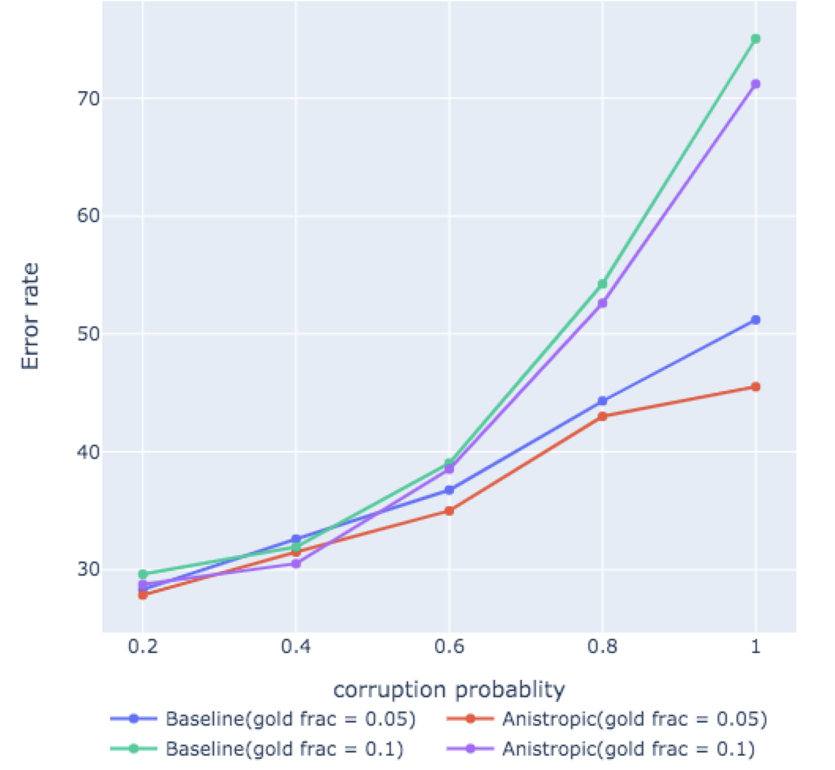
Method	Top-1 Acc	Top-5 Acc
ImageNet Baseline	13.00	26.24
Stylized Baseline	16.36	31.56
Anisotropic (Ours)	**24.49**	**41.81**

- We evaluate on Sketch ImageNet to show that we learn better shape representation as compared to the baseline.

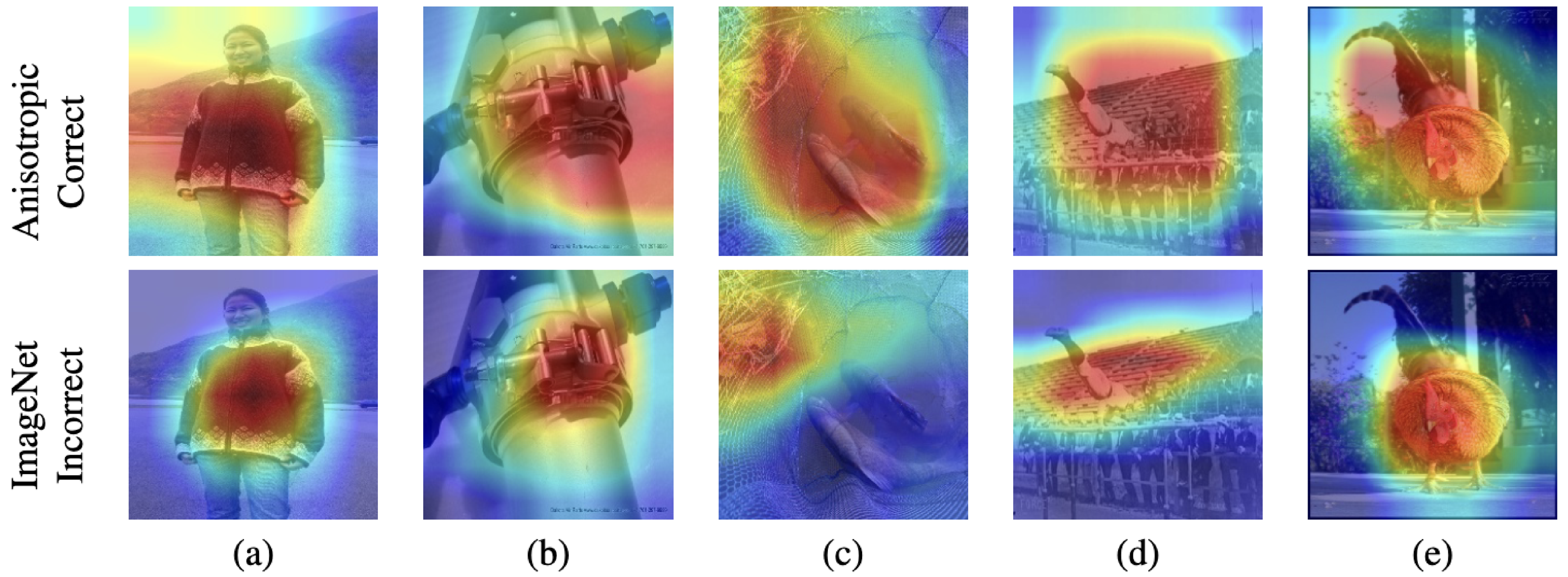
Analysis

- Our model is less reliant on high-frequency information.
- More robust to common corruptions and is more confident in making the right predictions.
- On label corruption task we consistently outperform the baseline with a larger improvement upon increasing the corruption probability.





- Anisotropic model has saliency maps that spread over bigger area and include the outline of the objects.



- Empirical results suggest that using the proposed data augmentation for pretraining self-supervised models and for training supervised models gives improvements across ten diverse datasets.