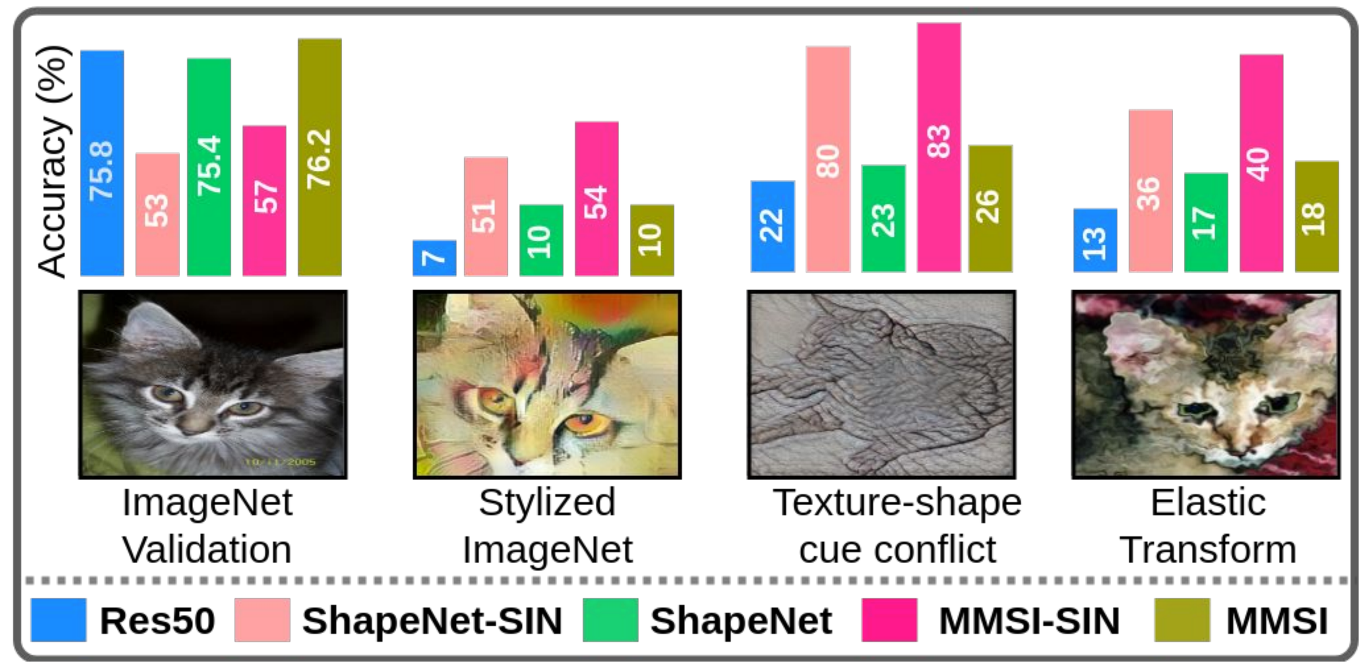
Maximizing Mutual Shape Information

Md Amirul Islam[1,7] Matthew Kowal[2,7] Patrick Esser[6]
Björn Ommer[4] Konstantinos G. Derpanis[2,5,7] Neil D. B. Bruce[3,7]

[1]Ryerson University, [2]York University, [3]University of Guelph, [4]University of Munich, [5]SAIC Toronto, [6]Runway, [7]Vector Institute

BMVC 2022

Motivation

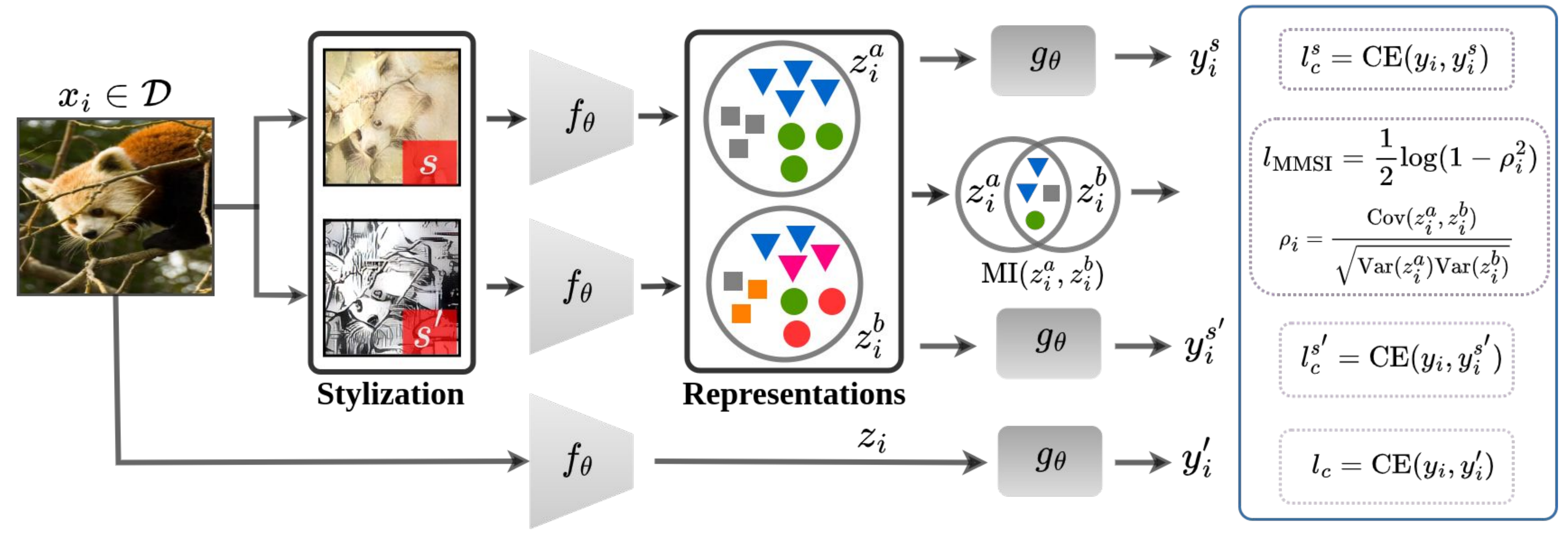


- ➢ Recent works have shown that models which strongly rely on texture information to make categorical decisions **perform poorly on out-of-domain examples**.
- ➢ Alternatively, networks which classify based on object shape have **improved generalization ability and robustness**.

Contributions

- ➢ We propose a novel training loss for increasing a neural network's ability to encode shape information.
- ➢ Our shape biased models are more robust to adversarial attacks and distorted images, and generalizes better to out-of-distribution examples.
- ➢ We obtain all these benefits without sacrificing overall performance on ILSVRC2012 ImageNet and transfer learning on downstream tasks.

Maximizing Mutual Shape Information



- ➢ Given an input image, we create two stylized images, by applying a style transfer algorithm (i.e., two different styles on the same image) such that the images share the same shape.
- ➢ We feed the original and the stylized images to the same CNNs to generate latent representations
- ➢ Then we maximize the number of neurons which encode shape by computing the mutual information between the latent representations of the pair of images which share a semantic concept (i.e., shape).
- ➢ The three representations are then passed through a shared classifier to obtain the classification outputs which are used to compute the cross entropy losses.

Influence of Shape on Adversarial Attacks

Method	*Clean*	*FGSM*	*I-FGSM*
ShapeNet	91.9	2.3	13.5
MMSI	92.5	37.7	25.2

Method	*Clean*	*PGD*	*MIM*
ShapeNet	91.9	22.4	0.7
MMSI	92.5	42.5	10.9

- ➢ Evaluated on NIPS-2017-non-targeted-adversarial attack dataset.
- ➢ Our method is significantly more robust than the ShapeNet.

Results on ILSVRC2012 ImageNet

Method	*Top1 Acc.*	*Top5 Acc.*
ResNet50	75.8	92.7
ShapeNet	75.4	92.5
MMSI	76.2	92.9

- ➢ Our method marginally outperforms the baselines on ILSVRC2012 ImageNet dataset.

CAM Visualization of Learned Representations

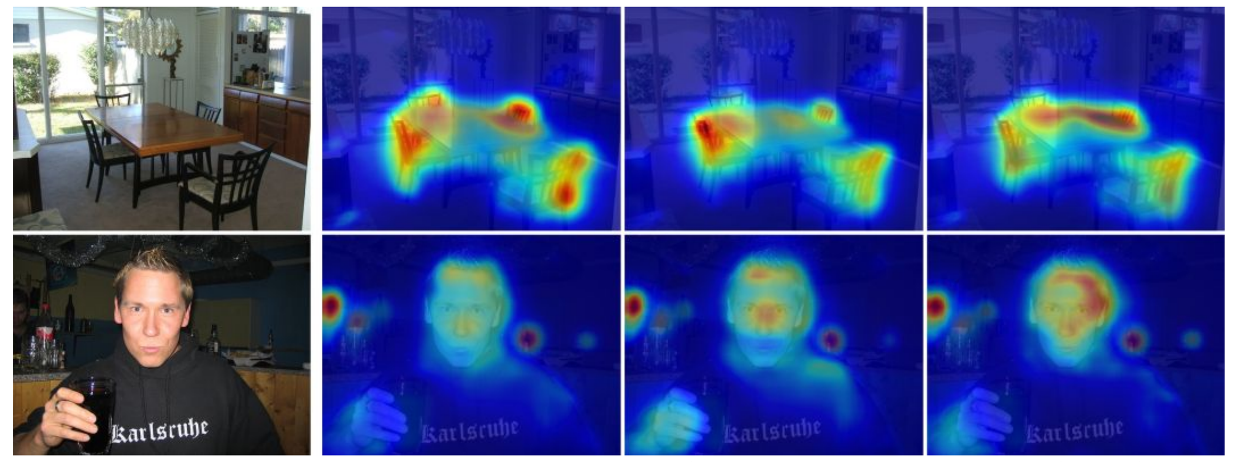


Image ResNet50 ShapeNet MMSI (Ours)

- ➢ MMSI (ours) can take advantage of global object shape to produce a more reliable larger shape cue.

Shape Bias and Dimensionality

Method	*Bias* *Shape*	*Bias* *Texture*	*Factor* *Shape*	*Factor* *Texture*
ShapeNet	22.9	77.1	18.4	31.2
MMSI (Ours)	26.1	73.9	18.8	30.2

- ➢ MMSI has more shape encoding neurons and higher bias than ShapeNet.

Adaptability of Learned Representations

Method	Image Classification VOC2007 mAP(%)	Image Classification Caltech-101 Top-1 (%)	Image Classification CIFAR100 Top-1 (%)	Image Classification DTD Top-1 (%)	Semantic Segmentation VOC12 FT	Semantic Segmentation VOC12 Freeze	CAMs VOC12 mIoU (%)
ShapeNet-SIN	90.7	**91.0**	80.8	**67.0**	55.7	40.0	42.6
MMSI-SIN	**91.3**	90.0	**81.1**	**67.0**	**56.2**	**40.8**	**44.6**
ResNet-IN	93.9	94.3	82.6	68.1	62.7	49.6	48.3
ShapeNet	93.8	94.2	82.7	67.6	62.5	47.4	48.3
MMSI	**94.3**	**95.0**	**82.8**	**68.7**	**62.8**	**50.0**	**48.9**

- ➢ We directly assess the quality of the shape-based representations by fine-tuning on different downstream tasks.
- ➢ Our approach marginally outperforms the baselines across various tasks.

Conclusions

- ➢ We presented a simple and effective strategy to learn shape-centric representations for object recognition while improving the network's robustness and generalization.
- ➢ We achieved robustness without sacrificing overall classification performance.
- ➢ Our model showed a significant improvement in its robustness to various attacks and distortions.