Appendix – Edges to Shapes to Concepts: Adversarial Augmentation for Robust Vision

Model	E	E + C	E+ M	E+ TSD	E+C+M
IN-1k	80.6	81.3	81.2	80.4	**81.40**
IN-R	44.7	48.1	48.0	45.1	**45.14**
IN-A	28.3	28.9	28.7	26.3	**29.28**

Table 1. ViT-S model is trained using ELEAS along with other augmentation techniques. ViT-S trained using ELEAScombined with Cutmix and Mixup (E+C+M) gives the best performance.

A. ELEAS with other techniques

ELEAS (E)works well in conjunction with other established image augmentation techniques. We applied ELEAS on top of other augmentation techniques with a probability of 0.5. Table 1 shows that employing CutMix (C) [4] and MixUp (M) [5] together with ELEAS further improves the model's performance.

B. Variance analysis / Statistical significance

In order to evaluate the consistency of the results obtained by ELEAS, we trained the ResNet101 model for five independent runs. Table 2 demonstrates the consistent results of our method, ELEAS, across multiple model and setup instantiations.

Model	IN-1K	IN-A	IN-R	IN-Sketch
Res101	78.65±0.08	13.4±0.03	44.4±0.06	32.4±0.03

Table 2. Variance calculation using 5 independent training runs. As observed, the variance is negligible across these runs.

C. Hyperparameter ablations

We sweep hyperparameters for the ratio of clean vs ELEAS image loss (eq. 2 in the paper) and β in $\mathtt{Beta}(1, \beta)$ (eq. 1 in the paper) (see Table 3). The loss ratio is used as a proxy for augmented image frequency. We chose Laplacian edge detection for image edge detection because it's fast (single convolution pass) and efficient enough for online computation, if needed, without impacting total training time.

Model	IN-1K	IN-A	IN-R	IN-Sketch
Res101 ($\eta = 0.6$)	78.54	13.3	44.2	32.2
Res101 ($\eta = 0.65$)	78.65	13.4	44.4	32.4
Res101 ($\eta = 0.7$)	78.50	13.1	44.0	32.1
Res50 ($\beta = 0.4$)	76.64	4.8	41.1	29.5
Res50 ($\beta = 0.5$)	77.10	5.4	41.7	29.7
Res50 ($\beta = 0.6$)	76.69	4.9	40.5	29.2

Table 3. Hyperparamter change experiments of loss weighting ratio η and edge & texture mixing beta distribution parameter β.

Method	IN-A(↑)	IN-R(↑)	IN-C(↓)	IN-Sketch(↑)	IN-1K(↑)
Vanilla	2.0	36.2	75.0	23.5	76.4
TSD [3]	3.3	40.8	67.5	28.3	76.9
AugMix [2]	3.7	35.2	64.9	28.6	**77.6**
SIN [1]	1.9	41.5	73.8	26.6	76.7
ELEAS	**5.4**	**41.7**	**58.5**	**29.7**	77.1

Table 4. **Performance comparison on the ImageNet and the robustness datasets for Resnet50.** The models trained using ELEAS show an improvement in the ImageNet performance along with better robustness. Except for IN-C the performance is measured in Accuracy@1 (**higher is better**). For IN-C, the performance is measured in mean corruption error (mCE) (**lower is better**).

D. Comparison with AugMix

AugMix is a data-augmentation strategy proposed in [2] that stochastically samples augmentation techniques like rotation or posterization and mix the random augmentation sequence using a convex combination. AugMix does not enforce the increase in the model's shape sensitivity; however, it was proposed to increase the robustness of the model to shift in data distribution. Therefore we have compare ELEAS with the AugMix and reported the results in Table 4. The proposed method (i.e., ELEAS) significantly outperforms the AugMix augmentation on all the robustness benchmarks. Although the AugMix gets the best performance on the clean imagenet datasets, it performs weaker on harder IN-A and IN-R datasets than other baseline methods.

E. Qualitative samples of images generated by ELEAS

We have added some qualitative examples of image samples generated by **E**dge **Lea**rning for **S**hape sensitivity (ELEAS) in Figure 1.

References

[1] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A. Wichmann, and Wieland Brendel. Imagenet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. In *International Conference on Learning Representations*, 2019.

[2] Dan Hendrycks, Norman Mu, Ekin Dogus Cubuk, Barret Zoph, Justin Gilmer, and Balaji Lakshminarayanan. Augmix: A simple data processing method to improve robustness and uncertainty. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020.

[3] Yingwei Li, Qihang Yu, Mingxing Tan, Jieru Mei, Peng Tang, Wei Shen, Alan Yuille, and cihang xie. Shape-texture debiased neural network training. In *International Conference on Learning Representations*, 2021.

[4] Sangdoo Yun, Dongyoon Han, Sanghyuk Chun, Seong Joon Oh, Youngjoon Yoo, and Junsuk Choe. Cutmix: Regularization strategy to train strong classifiers with localizable features. In *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019*, pages 6022–6031. IEEE, 2019.

[5] Hongyi Zhang, Moustapha Cissé, Yann N. Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net, 2018.

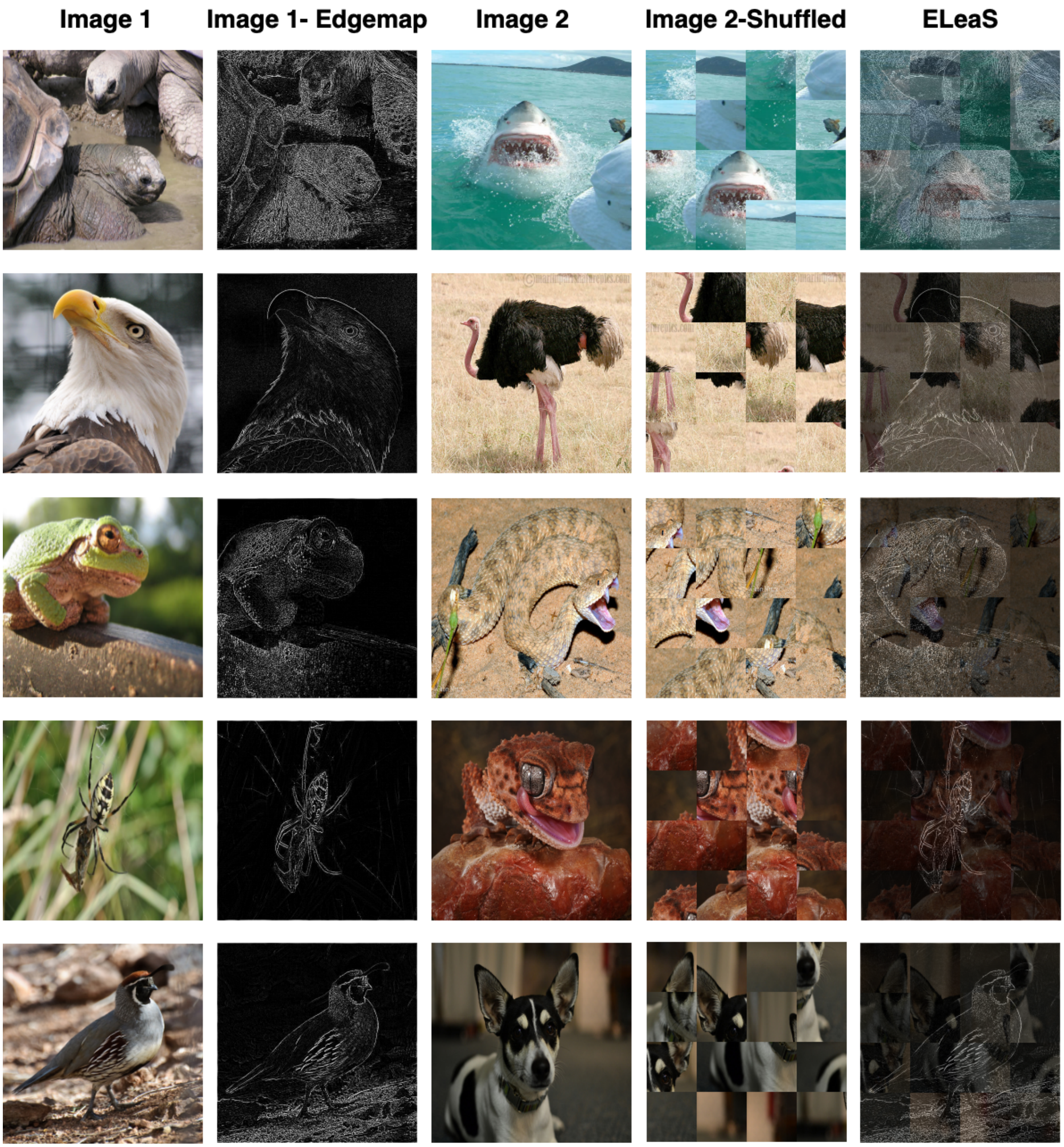

Figure 1. **The augmented image samples generated by** ELEAS We first create edgemap from Figure 1 and patch-shuffled image from Figure 2 and then generate their superposition using weights sampled from `Beta(4,1)` distribution.(**Best viewed in color.**)