A Appendix

A.1 Ablation study

Table 2: Test accuracy for the CIFAR-100 experiment with and without the standard cross-entropy (CE) loss (highlighted in magenta in the algorithm 1).

Dataset	**CutMix**	**Vanilla Mixup** $k=1$	**Region Mixup** $k=2$
w/o standard CE	78.27 ± 0.34	77.5 ± 0.22	76.36 ± 0.26
with standard CE	79.03 ± 0.30	78.1 ± 0.60	78.75 ± 0.28

A.2 Adversarial robustness

We assess the robustness of the trained models against adversarial samples. Adversarial examples are generated (in one single step) using the Fast Gradient Sign Method (FGSM) (Goodfellow et al., 2015), with the assumption that the adversary possesses complete information about the models, thereby conducting a white-box attack. Following Zhang et al. (2018), we constrain our experiment to basic FGSM attacks as the strength of iterative PGD attacks diminishes the practical relevance of any observed performance enhancements. For the black-box attack setting, we consider l_∞-square attack (Andriushchenko et al., 2020) with a constraint on the query budget limited to 100 queries. We use torchattack (Kim, 2020) to launch these attacks. We report test accuracies after the attack in Table 3 and Table 4.

Table 3: Test accuracy on white-box FGSM adversarial examples.

Dataset	**Model**	**CutMix**	**Vanilla Mixup** $k=1$	**Region Mixup** $k=2$
CIFAR-10		36.15±2.73	52.40±6.60	**58.96±5.23**
CIFAR-100	PreAct ResNet-18	13.08±0.87	18.56±1.22	**22.77±1.20**
Tiny ImageNet		**2.69 ± 0.19**	1.71±0.18	2.20±0.22

Table 4: Test accuracy on black-box Square Attack (l_∞). The black-box attacks are provided with a budget of 100 queries

Dataset	**Model**	**CutMix**	**Vanilla Mixup** $k=1$	**Region Mixup** $k=2$
CIFAR-10		31.76±1.70	**52.18±1.18**	51.35±1.99
CIFAR-100	PreAct ResNet-18	10.54±0.59	**20.03±0.46**	18.70±1.05
Tiny ImageNet		17.03±0.18	24.78±0.19	**25.02±0.69**

A.3 CLASS ACTIVATION MAPPINGS

We qualitatively compare Mixup, CutMix, and Region Mixup using class activation mappings (CAM) generated by Grad-CAM++ (Chattopadhay et al., 2018) on Tiny ImageNet dataset. We use the final residual block (layer4) of PreAct ResNet as the target layer to compute CAM.

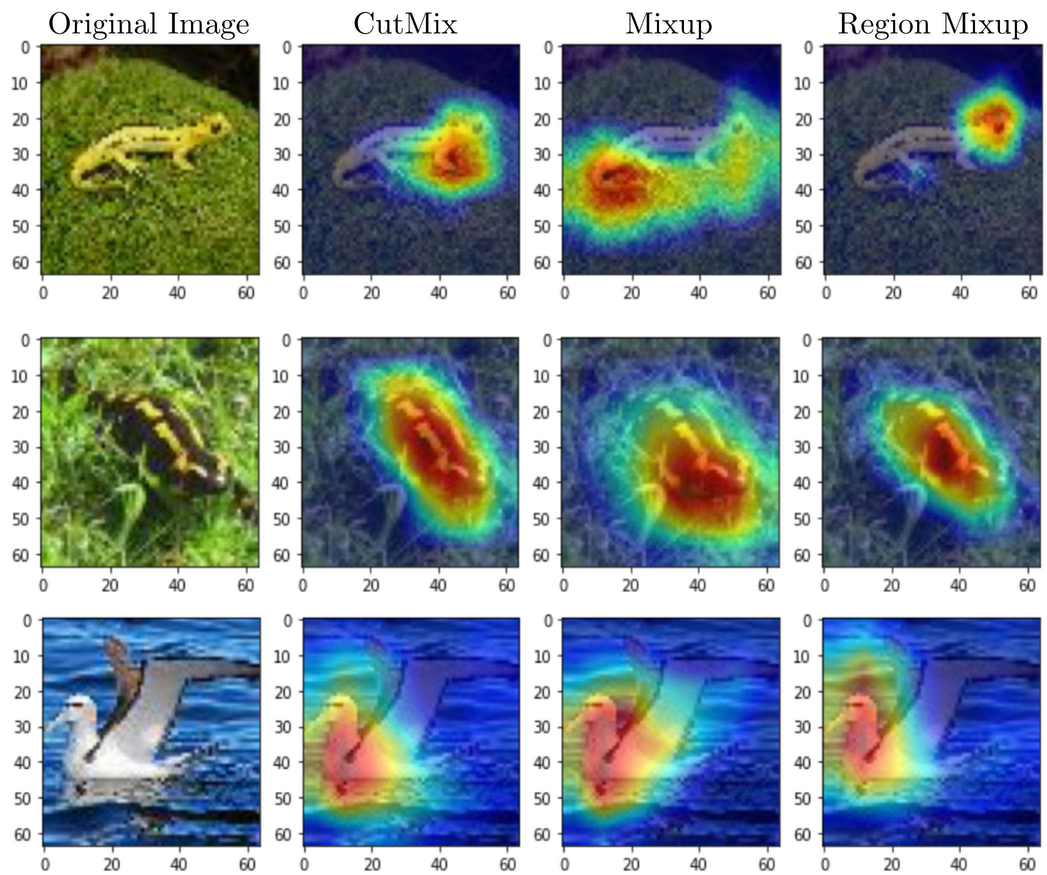

Figure 2: Class activation mapping (CAM) (Zhou et al., 2015) visualizations on Tiny ImageNet using Grad-CAM++ (Chattopadhay et al., 2018).