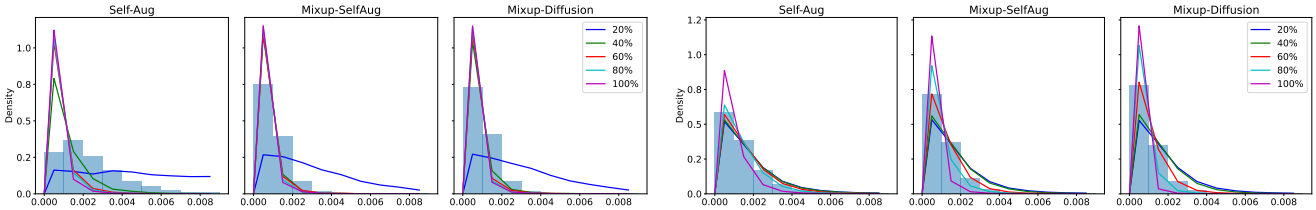
Figure 1: Norms of gradients before clipping and adding noise on CIFAR-10(a) and Caltech256(b) with fining tuning Vit-B-16 model with $\varepsilon = 1$ and $\delta = 10^{-5}$. The different curves represent different training stages. The histogram shows average norm of gradients. Gradients with our proposed Mixup-SelfAug and Mixup-Diffusion are more concentrated suggesting more stable training and faster convergence.



(a) CIFAR-10 **(b)** Caltech256

Table 1: Fine-tune Vit-B-16 model and ConvNext models on Caltech256, SUN397 and Oxford-IIIT Pet datasets with different ε. We set $\delta = 10^{-5}$. We can observe that our proposed methods improves performance in all cases.

Model	Dataset	Method	ε=1	ε=2	ε=4	ε=8
Vit-B-16	Caltech256	Mixed Ghost clipping	79.74%(±0.15%)	88.21%(±0.21%)	91.42%(±0.24%)	92.27%(±0.16%)
		Self-Aug	80.36%(±0.11%)	89.67%(±0.16%)	92.01%(±0.08%)	93.17%(±0.15%)
		Mixup-SelfAug	**81.21%**(±0.15%)	**90.12%**(±0.17%)	**92.17%**(±0.21%)	**93.39%**(±0.08%)
		Mixup-Diffusion	**85.82%**(±0.12%)	**91.32%**(±0.11%)	**92.86%**(±0.14%)	**93.87%**(±0.10%)
	SUN397	Mixed Ghost clipping	70.67%(±0.17%)	71.21%(±0.13%)	72.24%(±0.15%)	72.51%(±0.19%)
		Self-Aug	72.65%(±0.09%)	76.02%(±0.14%)	78.05%(±0.11%)	79.54%(±0.15%)
		Mixup-SelfAug	**73.19%**(±0.13%)	**76.45%**(±0.17%)	**78.67%**(±0.16%)	**79.57%**(±0.14%)
		Mixup-Diffusion	**75.12%**(±0.17%)	**77.78%**(±0.12%)	**79.47%**(±0.18%)	**80.57%**(±0.09%)
	Oxford-IIIT Pet	Mixed Ghost clipping	71.21%(±0.18%)	79.12%(±0.15%)	80.37%(±0.22%)	81.02%(±0.19%)
		Self-Aug	72.21%(±0.21%)	82.11%(±0.19%)	85.84%(±0.25%)	88.23%(±0.11%)
		Mixup-SelfAug	**72.45%**(±0.24%)	**82.51%**(±0.21%)	**86.75%**(±0.17%)	**88.70%**(±0.15%)
		Mixup-Diffusion	**73.67%**(±0.25%)	**84.57%**(±0.18%)	**88.14%**(±0.21%)	**89.25%**(±0.17%)
ConvNext	Caltech256	Mixed Ghost clipping	79.21%(±0.18%)	87.37%(±0.19%)	87.96%(±0.14%)	88.17%(±0.17%)
		Self-Aug	79.98%(±0.15%)	88.15%(±0.11%)	91.02%(±0.14%)	92.17%(±0.10%)
		Mixup-SelfAug	**81.07%**(±0.11%)	**88.69%**(±0.12%)	**91.34%**(±0.08%)	**92.38%**(±0.09%)
		Mixup-Diffusion	**85.08%**(±0.14%)	**90.74%**(±0.16%)	**92.27%**(±0.11%)	**93.27%**(±0.06%)
	SUN397	Mixed Ghost clipping	64.27%(±0.22%)	65.05%(±0.15%)	65.25%(±0.12%)	65.33%(±0.14%)
		Self-Aug	72.21%(±0.14%)	76.45%(±0.10%)	77.95%(±0.16%)	78.93%(±0.11%)
		Mixup-SelfAug	**72.47%**(±0.07%)	**76.82%**(±0.11%)	**78.52%**(±0.04%)	**79.45%**(±0.08%)
		Mixup-Diffusion	**74.95%**(±0.16%)	**77.51%**(±0.13%)	**79.33%**(±0.10%)	**79.98%**(±0.12%)
	Oxford-IIIT Pet	Mixed Ghost clipping	65.21%(±0.23%)	78.19%(±0.18%)	79.12%(±0.24%)	79.85%(±0.21%)
		Self-Aug	68.11%(±0.18%)	81.29%(±0.21%)	85.47%(±0.14%)	86.96%(±0.11%)
		Mixup-SelfAug	**68.75%**(±0.24%)	**81.65%**(±0.22%)	**86.25%**(±0.17%)	**87.67%**(±0.16%)
		Mixup-Diffusion	**71.91%**(±0.19%)	**83.55%**(±0.22%)	**87.94%**(±0.21%)	**88.56%**(±0.18%)

Table 2: FID values between the train and test sets, and train set and text-to-images diffusion model generated images.

FID model	Measurement	CIFAR-10	CIFAR-100	EuroSAT	Caltech256
InceptionV3 / Vit-B-16	Between training and testing	3.15 / 0.42	3.57 / 0.49	7.41 / 7.26	6.78 / 1.64
	Between training and text-diffusion	30.53 / 30.10	19.79 / 21.85	**164.00** / **82.88**	25.14 / 37.39

Table 3: FID values for different methods' generated images compared to the training set and test set.

Compared with	Method	CIFAR-10	CIFAR-100	EuroSAT	Caltech256
Train / Test	Self-Aug	2.64 / 3.08	3.09 / 3.53	2.18 / 4.12	1.02 / 2.53
	Self-Aug +	5.95 / 6.37	6.05 / 6.47	6.08 / 8.27	2.88 / 4.23
	Mixup-SelfAug	3.07 / 3.49	3.34 / 3.79	2.99 / 6.14	1.31 / 2.77
	Mixup-Diffusion	3.26 / 3.67	3.46 / 3.92	6.57 / 10.89	1.54 / 2.85

Table 4: Vit-B-16 model performance with Self-Augs+ on CIFAR-10,CIFAR-100 and EuroSAT with different ε

Dataset	Method	ε=1	ε=2	ε=4	ε=8
CIFAR-10	Mixed Ghost clipping	95.08%(±0.14%)	95.00%(±0.23%)	95.28%(±0.42%)	95.33%(±0.21%)
	Self-Aug	96.49%(±0.12%)	96.98%(±0.02%)	97.06%(±0.03%)	97.23%(±0.03%)
	Mixup-SelfAug	**97.21%**(±0.27%)	**97.35%**(±0.18%)	**97.42%**(±0.22%)	**97.55%**(±0.25%)
	Self-Aug+	96.30%(±0.09%)	96.49%(±0.14%)	96.80%(±0.10%)	96.88%(±0.08%)
CIFAR-100	Mixed Ghost clipping	78.16%(±0.35%)	78.53%(±0.05%)	78.36%(±0.32%)	78.43%(±0.10%)
	Self-Aug	79.28%(±0.18%)	83.18%(±0.33%)	83.47%(±0.30%)	84.18%(±0.08%)
	Mixup-SelfAug	**81.75%**(±0.15%)	**83.54%**(±0.12%)	**84.52%**(±0.05%)	**84.58%**(±0.20%)
	Self-Aug+	78.15%(±0.21%)	81.10%(±0.24%)	81.52%(±0.17%)	83.20%(±0.15%)
EuroSAT	Mixed Ghost clipping	84.02%(±0.09%)	84.82%(±0.15%)	84.85%(±0.07%)	85.04%(±0.15%)
	Self-Aug	93.28%(±0.15%)	94.13%(±0.23%)	95.34%(±0.21%)	95.49%(±0.15%)
	Mixup-SelfAug	**94.32%**(±0.11%)	**94.91%**(±0.15%)	**95.56%**(±0.21%)	**95.58%**(±0.13%)
	Self-Aug+	92.20%(±0.14%)	93.87%(±0.11%)	94.27%(±0.13%)	94.33%(±0.15%)

Table 5: Change the number of diffusion samples(k") for Mixup-Diffusion by using Vit-B-16 model.We set $\delta = 10^{-5}$ and $\varepsilon = 1$.

# of diffusion samples	CIFAR-100	EuroSAT	Caltech 256	Oxford-IIIT PET
2	82.02% (±0.11%)	92.58%(±0.14%)	85.82%(±0.12%)	73.67%(±0.25%)
4	81.91%(±0.15%)	92.31%(±0.18%)	86.68% (±0.15%)	77.07% (±0.22%)
8	81.86%(±0.12%)	90.93%(±0.08%)	88.59% (±0.07%)	78.59% (±0.18%)
16	81.26%(±0.18%)	89.38%(±0.12%)	89.28% (±0.11%)	83.32% (±0.23%)

Table 6: Increase K from 16 to 36 for Self-Aug. We conduct experiments on CIFAR-10 with $\varepsilon = 8$ and $\delta = 10^{-5}$ in two settings: train a WRN-16-4 model from scratch and fine-tune a pretrained Vit-B-16. We can observe that for both cases, there are no substantial performance improvements when increasing K.

Training method	K	WRN-16-4	Vit-B-16 (pretrained)
Self-Aug	16	78.74% (±0.45%)	97.23% (±0.03%)
	24	78.49% (±0.42%)	96.95% (±0.11%)
	32	78.40% (±0.34%)	97.04% (±0.05%)
	36	78.55% (±0.38%)	96.96% (±0.08%)
Mixup-SelfAug	32	**79.83%**(±0.32%)	**97.55%**(±0.25%)