shiqing

June 4, 2024

$\tilde{x} = x + \epsilon \cdot sign(-\nabla_x \log p(x; D))$
$\tilde{x} = x - \epsilon \cdot sign(-\nabla_x \log p(x; D))$

ood dataset	auroc	aupr
svhn	0.9188	0.8935
lsun	0.9369	0.9531
cifar100	0.8871	0.8967
mnist	**0.9213**	**0.9389**
tiny-imagenet	0.8874	0.8946
svhn+grad	**0.9491**	**0.9268**
lsun+grad	**0.9430**	**0.9582**
cifar100+grad	**0.8884**	**0.9086**
mnist+grad	0.8908	0.928
tiny-imagenet+grad	**0.8874**	**0.9003**

Table 1: vgg16+cafar10,accuracy=0.9402, input grad

ood dataset	auroc	aupr
without noise	0.9210	0.9434
noise(fgsm)	**0.9613**	**0.9710**
noise(bim)	0.9613	0.9710
noise(pgd)	0.9594	0.9705

Table 2: vgg16+cafar10,ood:svhn,epsilon=0.001

ood dataset	auroc	aupr
svhn	0.9107	0.8645
lsun	0.9057	0.9244
cifar100	0.8690	0.8836
mnist	0.9260	0.9473
tiny-imagenet	0.8563	0.8927
svhn+grad	**0.9639**	**0.9402**
lsun+grad	**0.9440**	**0.9561**
cifar100+grad	**0.8869**	**0.9019**
mnist+grad	**0.9474**	**0.9641**
tiny-imagenet+grad	**0.8842**	**0.8927**

Table 3: resnet50+cafar10,accuracy=0.9461,input grad