shiqing

June 5, 2024

$$\tilde{x} = x + \epsilon * \exp(\lambda * \max\{(norm - threshold), 0\}) \cdot sign(-\nabla_x \log p(x; D))$$

$threshold$:在训练集上统计得到所有关于输入的梯度的范数，计算分位数得到

通过gridSearch得到超参数:$\epsilon = 0.01, \lambda = 50$

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

ood dataset	auroc	aupr
svhn	0.9214	0.9434
lsun	0.9376	0.9538
cifar100	0.8871	0.8967
mnist	0.9213	0.9389
tiny-imagenet	0.9304	0.9527
svhn+noise	**0.9902**	**0.9924**
lsun+noise	**0.9979**	**0.9983**
cifar100+noise	**0.9405**	**0.9564**
mnist+noise	**0.9240**	**0.9406**
tiny-imagenet+noise	**0.9380**	**0.9527**

Table 1: vgg16+cafar10,accuracy=0.9405

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