一种利用类别显著性映射生成对抗样本的方法

叶启松，戴旭初

（中国科学技术大学 网络空间安全学院，安徽 合肥 230026）

摘 要：如果对抗样本的迁移性越强，则其攻击结构未知的深度神经网络模型的效果越好，所以设计对抗样本生成方法的一个关键在于提升对抗样本的迁移性。然而现有方法所生成的对抗样本，与模型的结构和参数高度耦合，从而难以对结构未知的模型进行有效攻击。类别显著性映射能够提取出样本的关键特征信息，而且在不同网络模型中有较高的相似度。基于显著性映射的这一特点，在样本生成过程中，引入类别显著性映射进行约束，实验结果表明，该方法生成的对抗样本具有较好的迁移性。

关键词：深度学习；安全；对抗样本；迁移性

中图分类号：TP181　　文献标识码：A　　DOI：10.19358/j.issn.2096-5133.2021.06.002

引用格式：叶启松，戴旭初. 一种利用类别显著性映射生成对抗样本的方法[J]. 信息技术与网络安全，2021，40(6)：9-14.

An adversarial example generation method based on class activation map

Ye Qisong，Dai Xuchu

（School of Cyberspace Security，University of Science and Technology of China，Hefei 230026，China）

Abstract：The adversarial examples，if their transferability is stronger，will be more effective to attack models with unknown structure. Therefore，a key to design adversarial examples generation method is to improve the transferability of adversarial examples. However，the existing method for generating adversarial examples are highly coupled with the structure and parameters of the local model，which make the generated adversarial examples difficult to attack other models. The class activation map can extract the key feature information of the example，and it has high similarity in different neural network models. Based on this observation，the class activation map is used to constrain the process of example generation. Experimental results show that the adversarial examples generated by this method have good transferability.

Key words：deep learning；security；adversarial example；transferability

0 引言

深度学习技术在计算机视觉、语音识别、自然语言处理等各个领域有着广泛的应用，然而有研究表明，深度神经网络具有一定的脆弱性[1]，该脆弱性使得深度神经网络容易受到攻击，这一问题引起了广泛的重视。对抗样本攻击是攻击深度神经网络的主要方法，该方法通过对原样本添加微小的、不可察觉的扰动生成对抗样本，使得深度神经网络对该样本做出错误的预测。

对抗样本的迁移性指针对结构已知的深度神经网络模型生成的对抗样本，能使得结构未知的深度神经网络模型对该样本做出错误预测。如果对抗样本有更好的迁移性，其就能更好地攻击结构和参数未知的模型，这也是利用对抗样本进行攻击的主要应用场景。攻击者在拥有深度神经网络模型的结构和参数信息的前提下进行的对抗样本攻击，称为在白盒条件下的对抗样本攻击。现有的白盒条件下的对抗样本攻击方法虽然有较高的攻击成功率，但是其生成的对抗样本的迁移性较差，在主要的应用场景中并不适用。迁移性差的主要原因在于，这类方法所生成的对抗样本与模型的结构和参数高度耦合，其扰动难以对结构和参数不同的其他模型进行有效的干扰。迁移性差的这一缺点在目标神经网络引入了防御方法时表现得更为明显。

为了提升对抗样本的迁移性，文献[2]通过训练生成式神经网络，同时针对多个目标模型生成对

抗样本以提升对抗样本的迁移性，该方法在对抗样本的生成阶段不涉及任何梯度计算，生成过程较快。文献[3]以对抗训练的方式训练生成式对抗网络，对抗训练使得生成器能够学习到攻击结构未知模型的策略，利用该生成器生成对抗扰动，添加在原样本中得到对抗样本，该对抗样本具有较好的迁移性。文献[4]利用演化算法，得到关键位置(在样本中对结果影响较大的像素位置)，并在关键位置添加扰动得到对抗样本，该对抗样本在扰动量较小的同时拥有较好的迁移性。

基于梯度权重的类别显著性映射(Gradient-weighted Class Activation Map，Grad-CAM)[5]能够提取出样本的关键特征信息，该关键特征信息在不同模型中有较高的相似度，并且对模型的预测结果影响较大。利用 Grad-CAM 这一重要特性，本文提出一种生成对抗样本的新方法——基于 Grad-CAM 和生成式神经网络的对抗样本生成方法(Adversarial Example Generator by Using Class Activation Map，AEG-GC)，该方法的特点是生成的对抗样本与原样本相差较小，但它们的显著性映射相差较大。另外，实验表明，该方法所得到的对抗样本具有很好的迁移性。

1 相关工作简介

1.1 基于梯度权重的类别显著性映射

提取样本的关键特征信息的主要方法有 Guided-Backprop[6]、CAM[7]、Grad-CAM[5]。Guided-Backprop 获取关键特征信息的过程涉及两步：(1)在卷积神经网络中，利用卷积层替换池化层。(2)利用反向传播算法，计算模型的输出结果对输入的梯度。该梯度信息能够表示所提取的关键特征。该方法提取出的关键特征信息示意图如图 1(b)和图 1(d)所示，这两张图分别以猫和狗作为研究对象，虽然 Guided-Backprop 能够较好地提取出样本的关键特征信息，但是所提取出的特征信息也容易受到非研究对象的干扰，例如图 1(b)中以猫作为研究对象，仍然受狗这一类别的影响严重。CAM 方法获取关键特征信息的过程涉及两步：(1)将卷积神经网络中的全连接层替换为全局平均池化层(Global Average Pooling，GAP)[8]。(2)以全局平均池化层的结果作为系数，对最后一层卷积层的每个通道进行加权求和，加权求和的结果即为 CAM 所提取的关键特征信息。CAM 方法提取的关键特征信息如图 1(c)和图 1(e)所示，由图可见猫和狗的轮廓信息更加清晰。然而，Guided-Backprop 和 CAM 方法都有一个共同的缺点，它们在提取关键特征信息时，需要对原模型进行改造，并重新训练新的卷积神经网络，这破坏了原网络模型的结构和参数信息，同时这一过程也需要花费较大的计算代价。Grad-CAM 在 CAM 的基础上进行改进，其建立显著性映射的过程，不需要重新训练原卷积神经网络。本文将在 2.1 节详细介绍 Grad-CAM 方法的具体过程。

(a)原样本　(b)利用 Guided-Backprop 方法提取猫的关键特征信息　(c)利用 CAM 方法提取猫的关键特征信息

(d)利用 Guided-Backpro 方法提取狗的关键特征信息　(e)利用 CAM 方法提取狗的关键特征信息

图 1　关键特征提取的示意图

1.2 对抗样本生成方法

对于一个给定的分类器 $f(\boldsymbol{x}):\boldsymbol{x}\in\mathbf{R}^{H\times W\times 3}\rightarrow\boldsymbol{y}\in\mathbf{R}^{C\times 1}$，$\boldsymbol{y}$ 为以 $\boldsymbol{x}$ 为输入对应的输出，对抗攻击的目的是在 $\boldsymbol{x}$ 的邻域内找到一个样本 $\tilde{\boldsymbol{x}}$，使得分类器分类错误。而对抗样本攻击又可分为有目标攻击和无目标攻击，对于无目标攻击，其目的是在 $\boldsymbol{x}$ 的邻域内找到一个样本 $\tilde{\boldsymbol{x}}$，使得 $f(\tilde{\boldsymbol{x}})\neq\boldsymbol{y}$。而对于有目标攻击，其目的是在 $\boldsymbol{x}$ 的邻域内找到一个样本 $\tilde{\boldsymbol{x}}$，使得 $f(\tilde{\boldsymbol{x}})=\tilde{\boldsymbol{y}}$，其中 $\tilde{\boldsymbol{y}}$ 为攻击者自定义的目标类别，并且有 $\tilde{\boldsymbol{y}}\neq\boldsymbol{y}$。通常要求 $\tilde{\boldsymbol{x}}$ 与 $\boldsymbol{x}$ 足够接近，在本文中即要求对于给定值 $\varepsilon>0$，有 $||\boldsymbol{x}-\tilde{\boldsymbol{x}}||_2\leqslant\varepsilon$。

基于最优化的对抗样本攻击方法[1]利用受限拟牛顿梯度法(L-BFGS)，通过求解如下最优化表达式生成对抗样本：

$$\arg\min_{x^*} \lambda \cdot ||\tilde{x}-x||_2 - J(\tilde{x},y) \tag{1}$$

因为该方法并没有做出$||x-x^*||_2 \leqslant \varepsilon$的限定，所以通过该方法生成的对抗样本扰动量通常较大。

基于损失函数梯度方法主要是以增大损失$J(\tilde{x},y)$的值为目标构造对抗样本$\tilde{x}$，其中J为交叉熵损失。主要方法为快速梯度法(Fast Gradient Method，FGM)[9]，该方法通过如下方式得到对抗样本：

$$\tilde{x}=x+\varepsilon \cdot \frac{\nabla_x J(x,y)}{||\nabla_x J(x,y)||_2} \tag{2}$$

PGD[10]为FGM的迭代版本，其迭代方式如下：

$$\tilde{x}_0=x,\ \tilde{x}_{t+1}=\tilde{x}_t+\alpha \cdot \frac{\nabla_x J(\tilde{x}_t,y)}{||\nabla_x J(\tilde{x}_t,y)||_2} \tag{3}$$

为了使得$\tilde{x}$满足$||x-\tilde{x}||_2 \leqslant \varepsilon$，通常设$\alpha=\varepsilon/T$，其中$T$为总的迭代次数。PGD虽然拥有更高的攻击成功率，但是其生成的对抗样本的迁移性更差[11]。

为了解决PGD生成的对抗样本迁移性差的问题，MI-FGSM[12]在PGD基础上做出了改进，具体的迭代过程如下所示：

$$g_{t+1}=\mu \cdot g_t+\frac{\nabla_x J(\tilde{x}_t,y)}{||\nabla_x J(\tilde{x}_t,y)||_2} \quad (0<\mu<1) \tag{4}$$

$$\tilde{x}_0=x,\ \tilde{x}_{t+1}=\tilde{x}_t+\alpha \cdot \mathrm{sign}(g_{t+1}) \tag{5}$$

其中，g_t的迭代过程，利用归一化梯度进行加权平均，这一过程提升了梯度下降的稳定性，使得损失函数能在更少迭代周期的情况下逼近极值点。相比于PGD方法，MI-FGSM生成的对抗样本的迁移性有明显提升[12]。

而ATA[13]是将Grad-CAM方法与PGD方法相结合以提升对抗样本的迁移性，其迭代过程如下：

$$J_{ata}=J(\tilde{x}_t,y)+\lambda \cdot J_{Grad-CAM}(x,\tilde{x}) \tag{6}$$

$$\tilde{x}_0=x,\ \tilde{x}_{t+1}=\tilde{x}_t+\alpha \cdot \frac{\nabla_x J_{ata}}{||\nabla_x J_{ata}||_2} \tag{7}$$

其中$J_{Grad-CAM}(x,\tilde{x})$表示$x$和$\tilde{x}$之间的显著性映射的差异，将在2.2节进行详细介绍。

2 AEG-GC

AEG-GC方法结合Crad-CAM训练生成式神经网络，生成对抗样本。以下将分别介绍Grad-CAM计算方法和训练生成式神经网络的具体过程。

2.1 Grad-CAM计算方法

设A表示卷积神经网络中间层的特征矩阵，A^k表示其第k个通道对应的矩阵，矩阵大小为$N \times N$，$A^k_{i,j}$表示矩阵A^k在位置(i,j)上对应的值。y^c表示类别c对应的分类得分。Grad-CAM方法建立显著性映射的公式如下所示：

$$a_k^c=\frac{1}{N*N}\sum_i^N\sum_j^N\frac{\partial y^c}{\partial A^k_{i,j}} \tag{8}$$

$$L^c_{Grad-CAM}=\mathrm{Relu}\left(\sum_k^K a_k^c A^k\right) \tag{9}$$

计算所得的a_k^c为A^k对类别c影响的权重大小，$L^c_{Grad-CAM}$为Grad-CAM方法针对类别c建立的显著性映射矩阵，该矩阵中数值越大，表示对y^c的影响越显著。

2.2 AEG-GC的方法设计

本文所提出的AEG-GC方法，其训练框架如图2所示，其中生成式神经网络G以原样本x作为输入，输出对应的对抗扰动ρ，其添加在原样本中得到对抗样本$\tilde{x}=x+\rho$。

AEG-GC方法通过最小化损失J来训练生成式神经网络G：

$$J_{noise}(\rho)=\max(0,||\rho||_2-c) \tag{10}$$

$$J_{loss}=-\mathrm{loss}_F \tag{11}$$

$$J_{Grad-CAM}(x,\tilde{x})=-||L_{Grad-CAM}(x)-L_{Grad-CAM}(\tilde{x})||_2^2 \tag{12}$$

$$J=J_{noise}+J_{loss}+\lambda \cdot J_{Grad-CAM} \tag{13}$$

其中，J_{noise}使用Soft Hinge损失对扰动大小作出约

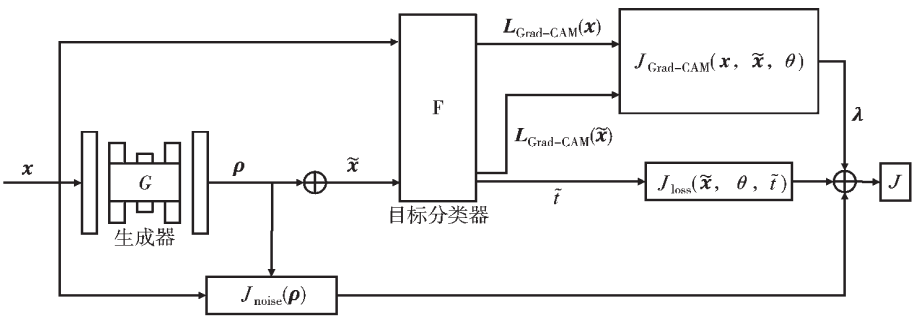

图2 AEG-GC训练框架示意图

束，式中 c 为自定义的扰动系数，最小化 J_{noise} 使得 G 能够生成不易被察觉的扰动。$loss_F$ 为目标分类器 F 对应的损失函数，可选择例如均方差损失、交叉熵损失、Soft Max 交叉熵损失等，最小化 J_{loss} 使得对抗样本能够最大限度地增加目标分类器的损失，也即使得目标分类器 F 越容易对 $\tilde{x}$ 分类错误。$L_{Grad-CAM}(x)$ 表示使用 Grad-CAM 方法针对目标分类器 F 提取出的关于 x 的显著性映射矩阵，如 1.1 节所述，显著性映射很好地刻画出了样本的关键特征信息，并且其对模型的预测有较大的影响，因此，最小化 $J_{Grad-CAM}$ 使得 x 与 $\tilde{x}$ 在目标分类器中的显著性映射的差异增大，增大了目标分类器对 $\tilde{x}$ 分类错误的概率。式中 λ 为系数，用以调整 $J_{Grad-CAM}$ 的权重。

3 实验

3.1 模型和数据集

本文将 AEG-GC 与 PGD 和 ATA 方法进行了比较，实验在 MNIST 和 CIFAR10 两个数据集上进行，MNIST 数据集大小为60 000，每个样本为28×28×1 的手写数字图片，共 10 个分类。CIFAR10 数据集大小为 60 000，每个样本大小为 32×32×3，共 10 个分类。

本实验在数据集 MNIST 上使用的模型为文献[14]所提供的 CNN_{nature}、CNN_{adv}、CNN_{secret}，在数据集 CIFAR10 上使用的模型为文献[15]所提供的 $Resnet_{nature}$、$Resnet_{adv}$、$Resnet_{secret}$。其中，nature 表示基于原数据集训练得到的模型，adv 和 secret 为使用了对抗样本进行对抗训练[9]所得到的模型，对抗训练表示在模型的训练阶段，将针对模型生成的对抗样本作为训练数据的一部分加入到训练过程。被对抗训练过的模型，能学习到抵御对抗扰动的策略，可以有效降低对抗样本的攻击成功率。因此，这两个模型在抵御对抗扰动方面，将比 nature 更为出色。

3.2 实验结果

实验中生成式神经网络模型结构来源于文献[16]，$loss_F$ 为交叉熵损失，共进行了 100 个训练周期，c 用以约束对抗样本扰动大小，在 MNIST 中为 0.3，在 CIFAR10 中为 8。

对比实验结果如表 1 和表 2 所示。利用 AEG-GC、ATA 和 PGD 三种对抗样本生成方法，针对原模型生成对抗样本攻击目标模型，表中的数值表示目标模型对对抗样本分类的准确率，准确率越低，表示相应对抗样本的攻击能力越强。当原模型与目标模型相同时，对应的数值表示白盒攻击的实验结果，其他的数值表示黑盒攻击的实验结果，黑盒攻击时，分类准确率越低，说明该对抗样本的迁移性越强。

表 1 MNIST 对抗样本攻击实验结果

			目标模型		
			CNN_{nature}	CNN_{adv}	CNN_{secret}
原模型	CNN_{nature}	AEG-GC	0.15	0.943 0	0.939 9
		ATA	0.07	0.958 7	0.956 0
		PGD	0	0.967 1	0.969 5
	CNN_{adv}	AEG-GC	0.822 6	0.904 5	0.930 6
		ATA	0.840 0	0.900 1	0.969 9
		PGD	0.854 2	0.938 0	0.978 1
	CNN_{secret}	AEG-GC	0.826 5	0.931 2	0.891 2
		ATA	0.854 4	0.960 4	0.902 4
		PGD	0.931 8	0.956 9	0.943 2

表 2 CIFAR10 对抗样本攻击实验结果

			目标模型		
			$Resnet_{nature}$	$Resnet_{adv}$	$Resnet_{secret}$
原模型	$Resnet_{nature}$	AEG-GC	0.18	0.653 3	0.643 3
		ATA	0	0.687 0	0.668 0
		PGD	0	0.861 3	0.859 5
	$Resnet_{adv}$	AEG-GC	0.533 0	0.448 0	0.618 8
		ATA	0.610 0	0.457 3	0.632 0
		PGD	0.786 0	0.472 7	0.634 1
	$Resnet_{secret}$	AEG-GC	0.522 9	0.599 3	0.445 4
		ATA	0.603 1	0.621 1	0.460 8
		PGD	0.790 0	0.635 0	0.479 1

设 $<Model1, Model2>$ 表示针对原模型 Model1 生成的对抗样本，攻击目标模型 Model2 时所对应的分类准确率。针对 MNIST 数据集，在白盒条件下，$<CNN_{nature}, CNN_{nature}>$、$<CNN_{adv}, CNN_{adv}>$ 和 $<CNN_{secret}, CNN_{secret}>$ 对应的实验数据中，AEG-GC 和 ATA 方法对应的准确率差距较小，因此 AEG-GC 方法和 ATA 方法的攻击性能比较近似。然而在黑盒攻击条件下，AEG-GC 对应的准确率更低，意味着 AEG-GC 方法生成的对抗样本的迁移性更强。针对 CIFAR10 数据集，在白盒攻击条件和黑盒攻击条件下，AEG-GC 攻击性能具有一定优势，并且，在 $<Resnet_{adv}, Resnet_{nature}>$ 和 $<Resnet_{secret}, Resnet_{nature}>$ 对应的实验中，AEG-GC 方法对应的分类准确率优势较为明显。总体上来说，AEG-GC 方法生成的对抗样本在白盒条件下的攻击性能较好，并且在黑盒条件下的攻击性能具有一

定的优势，其生成的对抗样本迁移性更强。

3.3 参数λ的影响

AEG-GC 方法中，λ 越大，生成器的训练过程越倾向于生成与原样本显著性映射差异更大的对抗样本。如图 3 所示，表示针对 MNIST 和 CIFAR10 数据集，使用 AEG-GC 方法，针对 adv 模型生成对抗样本攻击 secret 模型时的攻击成功率。

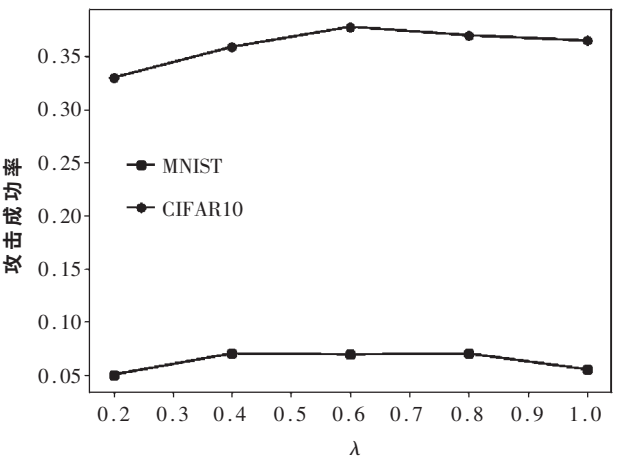

图 3 λ 的值对攻击成功率的影响

当增大 λ 时，MNIST 和 CIFAR10 数据集下，攻击成功率会有一定程度的提升，说明 Grad-CAM 的引入，有助于提升对抗样本的迁移性。然而随着 λ 的增大，相对来说，会减小训练过程中 J_{noise} 对于噪声大小的约束，此时对抗样本的噪声会增大，而目标模型预测过程中，首先会对超过一定阈值大小的像素值进行过滤处理，此时，噪声过大的对抗样本将不能对目标模型进行很好的攻击。

当 $\lambda=0.6$ 时，对抗样本有较高的攻击成功率，即此时对抗样本的迁移性较好，因此 3.2 节的实验中，λ 设为 0.6。

4 结论

本文提出一种生成对抗样本的方法——AEG-GC 方法，该方法利用 Grad-CAM 对生成式神经网络的训练过程进行约束，旨在生成对图片的关键特征干扰性更强的对抗样本，实验表明，对于 MNIST 和 CIFAR10 数据集，相比于 ATA 和 PGD 方法，AEG-GC 所生成的对抗样本具有更好的迁移性。与此同时，对于一个已训练好的 AEG-GC，其生成对抗样本的过程不需要进行梯度计算，因此，其实现的复杂度比 ATA 和 PGD 方法更低。但是针对不同场景的数据集，AEG-GC 需要重新训练不同的模型，所以该方法的局限性在于其只适用于在确定的场景进行的对抗样本攻击。为了解决这一局限性，增加该方法的应用场景，下一步工作可以以多个模型组成的集成模型作为目标，研究针对集成模型的关键特征提取方法，并在生成式神经网络中针对集成模型进行攻击。

参考文献

[1] SZEGEDY C, ZAREMBA W, SUTSKEVER I, et al. Intriguing properties of neural networks[EB/OL]. (2013-12-21)[2021-03-01]. https://arxiv.org/abs/1312.6199.

[2] BALUJA S, FISCHER I. Adversarial transformation networks: learning to generate adversarial examples[EB/OL]. (2017-03-28)[2021-03-01]. https://arxiv.org/abs/1703.09387.

[3] YOO Y J, PARK S, CHOI J, et al. Butterfly effect: bidirectional control of classification performance by small additive perturbation[EB/OL]. (2017-11-27)[2021-03-01]. https://arxiv.org/abs/1711.09681.

[4] NARODYTSKA N, KASIVISWANATHAN S P. Simple black-box adversarial perturbations for deep networks[EB/OL]. (2016-12-19)[2021-03-01]. https://arxiv.org/abs/1612.06299.

[5] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 618-626.

[6] SPRINGENBERG J T, DOSOVITSKIY A, BROX T, et al. Striving for simplicity: the all convolutional net[EB/OL]. (2015-04-13)[2021-03-01]. https://arxiv.org/abs/1412.6806.

[7] ZHOU B, KHOSLA A, LAPEDRIZA A, et al. Learning deep features for discriminative localization[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2921-2929.

[8] LIN M, CHEN Q, YAN S C. Network in network[EB/OL]. (2014-03-04)[2021-03-01]. https://arxiv.org/abs/1312.4400.

[9] GOODFELLOW I J, SHLENS J, SZEGEDY C. Explaining and harnessing adversarial examples[EB/OL]. (2015-03-20)[2021-03-01]. https://arxiv.org/abs/1412.6572.

[10] KURAKIN A, GOODFELLOW I, Bengio S. Adversarial examples in the physical world[EB/OL]. (2017-02-17)[2021-03-01]. https://arxiv.org/abs/1607.02533.
[11] SUTSKEVER I, MARTENS J, DAHL G, et al. On the importance of initialization and momentum in deep learning[C]. International Conference on Machine Learning, 2013: 1139-1147.
[12] DONG Y P, LIAO F Z, PANG T Y, et al. Boosting adversarial attacks with momentum[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 9185-9193.
[13] WU W B, SU Y X, CHEN X X, et al. Boosting the transferability of adversarial examples via attention[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020: 1161-1170.
[14] TSIPRAS D, SCHMIDT L. MNIST adversarial examples challenge[EB/OL]. [2020-10-28]. https:// -github.com/MadryLab/mnist_challenge.
[15] TSIPRAS D, SCHMIDT L. CIFAR10 adversarial examples challenge[EB/OL]. [2020-09-08]. https:// -github.com/MadryLab/cifar10_challenge.
[16] ISOLA P, ZHU J Y, ZHOU T, et al. Image-to-image translation with conditional adversarial networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1125-1134.

(收稿日期: 2021-03-13)

作者简介:

叶启松(1995-),男,硕士研究生,主要研究方向:对抗样本攻击和防御。

戴旭初(1963-),通信作者,男,教授,主要研究方向:信息安全、信号处理。E-mail: daixc@ustc.edu.cn。

版权声明

经作者授权，本论文版权和信息网络传播权归属于《信息技术与网络安全》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《信息技术与网络安全》编辑部

中国电子信息产业集团有限公司第六研究所