深度神经网络结合蚁群算法的躲避攻击多目标对抗方法

魏焕新[1] 张宏国[2]

[1](湖南机电职业技术学院信息工程学院 湖南 长沙 410151)

[2](哈尔滨理工大学软件工程系 黑龙江 哈尔滨 150080)

摘 要 针对深度神经网络在躲避攻击多目标对抗方法中输入的数据易导致机器误解码,提出一种深度神经网络结合蚁群算法的躲避攻击多目标对抗方法。设计一种与变换器和多个模型组成的体系结构,利用变换器生成一个多目标的对抗性样本,利用深度学习训练的分类器对输入值进行分类;引入蚁群算法,利用蚂蚁互相交流学习的正反馈原理保证算法的收敛性和寻优速度;融合两种算法的优势,实现躲避攻击的多目标对抗。实验结果表明,相比其他现有方法,该方法在躲避攻击多目标对抗方面更具优势,实现了 100% 的攻击成功率。

关键词 深度神经网络 躲避攻击 对抗样本 机器学习 蚁群算法 多目标对抗

中图分类号 TP391 **文献标志码** A DOI:10.3969/j.issn.1000-386x.2020.11.047

MULTI-TARGET COUNTERMEASURE METHOD OF AVOIDING ATTACK BASED ON DEEP NEURAL NETWORK AND ANT COLONY ALGORITHM

Wei Huanxin[1] Zhang Hongguo[2]

[1](*College of Information Engineering*, *Hunan College of Electrical and Mechanical Technology*, *Changsha 410151*, *Hunan*, *China*)

[2](*Department of Software Engineering*, *Harbin University of Science and Technology*, *Harbin 150080*, *Heilongjiang*, *China*)

Abstract Aiming at the fact that the data input by deep neural network in the multi-target antagonistic method of evading attack can easily lead to machine misunderstanding codes, a new method of evading attack multi-target antagonistic method based on deep neural network and ant colony algorithm is proposed. An architecture composed of a converter and multiple models was designed. A multi-target antagonistic sample was generated by the converter, and the input value was classified by the classifier of deep learning training. Then, the ant colony algorithm was introduced to ensure the convergence and optimization speed of the algorithm. The advantages of the two algorithms were fused to realize the multi-target confrontation of evading attack. Experimental results show that, compared with other existing methods, the proposed method has more advantages in avoiding attack against multiple targets and achieves 100% attack success rate.

Keywords Deep neural network Evade attack Counter sample Machine learning Ant colony algorithm Multi-objective confrontation

0 引 言

随着新兴计算技术的发展,机器学习技术作为分类方案发挥了关键作用。探索性攻击的对抗性样本引起了人们对深度神经网络安全性的关注。深度神经网络广泛用于图像识别、语音识别、模式分析和入侵检测等领域[1-3]。单目标对抗方法通常使用有针对性的对抗性样本,只允许识别一个类,容易错误地将攻击类识别为非攻击类。因此,研究多目标对抗方法具有很好的现实意义和实用价值[4-5]。

国内外许多专家及学者围绕多目标对抗方法进行了深入研究。文献[6]研究了机器学习中的几个安全问题,对机器学习的攻击归类为致病攻击、影响学习、

收稿日期:2019-05-16。国家自然科学基金项目(51375128);湖南省教育厅科学研究项目(15C0490);全国教育科学规划课题(EJA17450)。**魏焕新**,副教授,主研领域:网络安全,智能算法优化。**张宏国**,教授。

控制训练数据,以及作为探索性攻击导致错误分类,但不影响训练过程。文献[7-8]提出了一项关于对抗性实例的研究,使用对抗性样本的主要目的是通过向原始图像添加少量噪声来使深度神经网络犯错误。文献[9]提出了一种平滑梯度法来阻止攻击梯度,作为抵御黑匣子袭击的最先进的防御,它由一个探测器和一个改造者组成,以抵抗对抗性的样本攻击。文献[10]通过使用多个局部模型来攻击其他模型,提出了一种集合对抗性样本的方法。以上方法存在无法区分原始图像和失真图像之间差异的问题,仍有一定的改进空间[11-13]。

基于上述分析,本文提出一种基于深度神经网络躲避攻击的多目标对抗方法,并结合蚁群算法实现躲避攻击的多目标对抗。其主要创新点为:

(1) 现有的大多数方法中,对抗性样本只分为有针对性的对抗性样本和无针对性的对抗性样本,而本文方法设计一种与变换器和多个模型组成的体系结构,利用变换器生成一个多目标的对抗性样本,利用多个模型预训练的分类器对输入值进行分类。

(2) 现有的大多数方法中针对性对抗样本是单个目标攻击,只允许识别一个类,而提出的多目标对抗样本,使用单个修改后的图像攻击每个目标类中的多个模型,为了产生这样的样本,本文方法进行了一次变换,从而通过多个模型最大化不同目标类别的概率。

(3) 现有的大多数方法采用深度神经网络进行多目标对抗,而本文方法在深度神经网络的基础上引入了蚁群算法,利用蚂蚁互相交流学习的正反馈原理保证算法的收敛性和寻优速度,融合两种算法的优势,解决深度神经网络输入数据导致机器误解数据的问题,实现躲避攻击的多目标对抗。

1　提出的多目标对抗方法

图 1 为提出的多目标对抗方法的示意图。

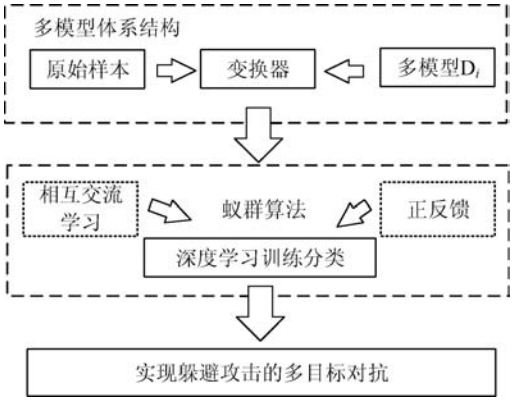


图 1　多目标对抗方法示意图

1.1　多目标对抗

生成对抗性样本的基本体系结构由两个元素组成:目标模型和变换器。变换器将原始样本 x 和目标类 y 作为输入数据,然后创建输出,变换的示例为 $x^* = x + w$,其中噪声值 w 被添加到原始样本 x 中。变换后的样本 x^* 作为输入数据提供给目标模型。目标模型为变换器提供变换后的样本的类概率结果。变换器更新变换的样本 $x^* = x + w$ 中的噪声值 w,使得其他类概率高于原始类概率,同时最小化 x^* 和 x 之间的失真距离。

通过对抗训练、过滤方法、防御性蒸馏和磁铁法等方法研究了对抗性样本的防御。为了阻止攻击梯度,这种防御性蒸馏具有两个神经网络,其中分类器的输出类概率被用作第二级分类器训练的输入。同样,这种方法有两个步骤:首先,检测器过滤失败的对抗性样本作为预学习步骤,包括原始样本和对抗性示例之间的差异;然后,重整器找到原始样本,该样本可以转换为对抗性样本的小扰动。使用过滤方法、防御蒸馏和磁铁方法进行测试发现,系统的被欺骗率可高达 100%。

对抗性样本攻击的相关工作可以分为四类:目标模型信息、距离度量、对抗性样本识别和生成方法[8]。

1.2　基于深度神经网络的多目标样本优化

深度神经网络广泛用于图像识别、语音识别、模式分析和入侵检测等领域,具有全局搜索能力强的优点。为了生成多目标的对抗性样本,设计一种与变换器和多个模型组成的体系结构,利用变换器生成一个多目标的对抗性样本,利用深度网络学习训练的分类器对输入值进行分类,提出一种由变换器和多个模型 D_i ($1 \leqslant i \leqslant n$) 组成的体系结构,如图 2 所示。变换器的作用是生成一个多目标的对抗性样本[14]。多个模型 D_i 是预训练的分类器,其对输入值进行分类。

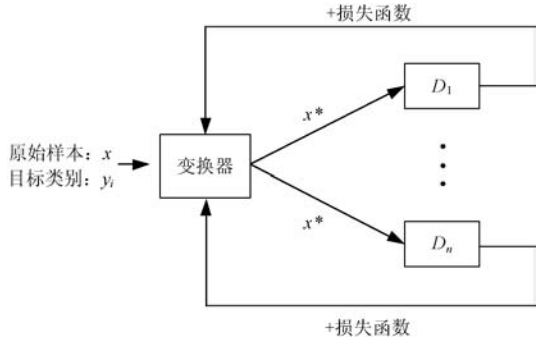


图 2　由变换器和多个模型组成的体系结构

若 DNN 模型将样本作为输入值,则它提供每个类的分类得分的输出值,所提深度神经网络模型将样本分类为输出值最高的类。表 1 显示了原始样本“2”和

目标对抗性样本“3”的分类得分示例。可以看出,原始样本“2”可分类为得分 29 的类别,对抗性样本“3”将分类为得分 18.703 的类别。

表 1 对抗样本“2”→“3”的分类分数

描述	原始的“2”	对抗性样本“3”
分类	[4 8 **29** -5 -8 -16	[-1 2 18.701 18.703
得分	3 -1 -3 -9]	-11 -4 -16 3 2 -3]

深度神经网络的变换器将原始样本 $x \in X$ 和目标类 $y_i \in Y(1 \leqslant i \leqslant n)$ 作为输入值,并生成变换后的样本 x^* 作为输出值。y_i 是攻击者为每个模型 D_i 选择的目标类。多个模型 D_i 将变换后的样本 x^* 作为输入值,并向变换器提供反馈,该反馈是分类的结果。

本文方案的目标是生成多目标对抗性样本 x^*,其被多个模型 D_i 中的每一个错误分类为每个目标类别 y_i,同时最小化计算原始样本 x 的像素距离。在数学表达式中,D_i 的运算函数分别由 $f_i(x)(1 \leqslant i \leqslant n)$ 表示。给定预训练模型 D_i、原始样本 $x \in X$ 和目标类 y_i,这是一个生成多目标对抗性样本 x^* 的优化问题:

$$x^*: \underset{x^*}{\operatorname{argmin}} L(x, x^*) \quad \text{s.t. } f_i(x^*) = y_i \quad 1 \leqslant i \leqslant n \tag{1}$$

式中:$L(\cdot)$ 是原始样本 x 和多目标对抗性样本 x^*。

为了实现这一目标,该过程包括两个步骤:创建预训练模型 D_i 和生成多目标对抗性样本 x^*。首先,模型 D_i 通过学习过程将原始样本 x 将其正确分类为其原始类 y^{org}:

$$f_i(x) = y^{\text{org}} \in Y \tag{2}$$

式中:y^{org} 是原始类。在实验中,训练模型 D_i 对原始样本进行分类时,其准确度可以达到 99%。

其次,变换器接受原始样本 x 和目标类别 y_i 作为输入值,并生成变换后的样本 x^*。在这项研究中,修改了变压器架构,x^* 被定义为:

$$x^* = \frac{\tanh(x + w)}{2} \tag{3}$$

式中:w 用作调整项以最佳地最小化梯度;tanh 用于软化梯度。模型 D_i 的 x^* 的分类损失返回到变换器。然后变换器通过上面的迭代过程更新变换的示例 x^*,以计算总损耗 $loss_{\text{T}}$ 并且最小化总损耗 $loss_{\text{T}}$。$loss_{\text{T}}$ 定义为:

$$loss_{\text{T}} = loss_{\text{distortion}} + \sum_{i=1}^{n} c_i \times loss_i \tag{4}$$

式中:$loss_{\text{distortion}}$ 是变换样本的失真;$loss_i$ 是 D_i 的分类损失;c_i 是模型 D_i 的损失权重。损失权重的初始值是 1。$loss_{\text{distortion}}$ 是原始样本 x 和变换的样本 x^* 之间的距离。

$$loss_{\text{distortion}} = \sqrt{\left(x^* - \frac{\tanh(x)}{2}\right)^2} \tag{5}$$

为了满足 $f_i(x^*) = y_i(1 \leqslant i \leqslant n)$,需要最小化 $\sum_{i=1}^{n} loss_i$:

$$\sum_{i=1}^{n} loss_i = \sum_{i=1}^{n} g_i(x^*) \tag{6}$$

式中:$g_i(k) = \max\{Z_i(k)_j; j \neq y_i\} - Z_i(k)_{y_i}$,其中 Z_i 是模型 D_i 预测类的概率。通过最优化最小化损失 $loss_i$,$f_i(x^*)$ 的预测具有比其他类别更高的预测目标类别 y_i 的概率。

算法 1 在变换器中生成一个多目标对抗性样本

输入:原始样本 x,目标类别 $y_i(1 \leqslant i \leqslant n)$,迭代次数 r,损失权重 c_i。

对目标对抗性样本生成:

$w \leftarrow 0$

$t \leftarrow y_i$

$x^* \leftarrow x$

For r step do

$x^* \leftarrow \frac{\tanh(x^* + w)}{2}$

$loss_1 \leftarrow \sqrt{\left(x^* - \frac{\tanh(x)}{2}\right)^2}$

$loss_2 \leftarrow c_i \sum_{i=1}^{n} \max\{Z_i(x^*)_j; j \neq t\} - Z_i(x^*)_t$

$loss_3 \leftarrow loss_1 + loss_2$

通过最小化 $loss_3$ 的梯度更新 w

End for

Return x^*

1.3 蚁群算法

多目标分配模型可以使用智能优化算法进行求解。蚁群算法是一种基于种群寻优的启发式智能优化算法,利用蚁群中蚂蚁间互相交流学习的正反馈原理保证了算法的收敛性和寻优速度[14-15]。借鉴蚁群算法解决问题的思路,设计面向对抗资源多目标分配的蚁群算法,并改进蚁群算法流程,以求解多目标分配模型。算法实现步骤如下:

(1) 随机生成初始蚁群 POP,种群数为 N,求出目标函数值 $f_i(x)$,$i = 1, 2, \cdots, J$。

(2) 初始化外部集合 BP,即 $X_f = \{x \in POP \mid e(x) \leqslant 0\}$,$BP = \{x \in X_f \mid \exists x' \in X_f$,有 $x' < x\}$。

(3) 设置迭代次数 $N_C = 0$。

(4) 令 $i = 1$。

(5) 随机产生一个[0, 1]范围内的随机数 p,当 $p \leqslant p_0$ 时,对蚂蚁 i 用基于全局最有经验指导的方式寻优;当 $p > p_0$ 时,使用信息素交流方式寻优。

(6) 在可行域内移动蚂蚁 i,并在最终位置上加入随机扰动因子 ϕ_0,再次评价蚂蚁 i,求得目标函数值以及约束函数值。

(7) 更新 BP,并删除 BP 中被 i 所支配的解。

(8) $i = i + 1$,如果 $i \leqslant N$,则转至(5)。

(9) $N_C = N_C + 1$,如果 $N_C \leqslant N_{CMax}$,则转至(4);否则,算法结束。

2 实　验

通过实验展示一种生成多目标对抗性样本,该样本被多个模型中的每个模型错误分类为每个目标类,同时能够最小化原始样本的计算像素距离。在实验中使用具有 Xeon E5 -2609 1.7 GHz 处理器的服务器。

2.1 实验方法

实验采用 MNIST 数据集,它是一个标准的数据集,包含从 0 到 9 的手写数字图像。实验方法为对 D_i ($1 \leqslant i \leqslant n$) 进行预训练并生成多目标对抗样本。

D_i 是常见的卷积神经网络。为了有足够的方法训练多个模型 D_i,使用了 60 000 个训练数据样本。不失一般性,令实验中模型数量 $n = 5$。通过学习不同的训练数据创建了若干个 D_i 的模型,同时保持原始样本的准确度超过 99%。在使用 10 000 个测试数据的实验结果中,D_i 正确地将原始样本分类到原始类别中,准确度超过 99%。

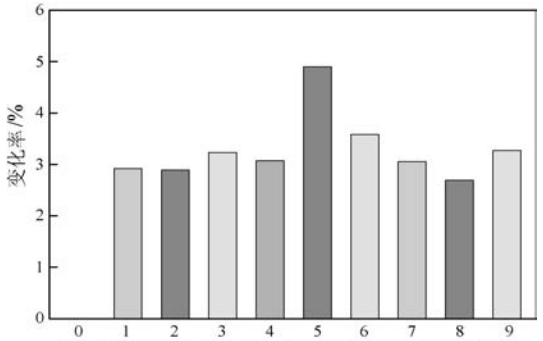
将 Adam 算法作为优化器来生成多目标对抗性样本,并使总损失最小化,参照文献[16]中所提参数设置方法,设定学习率为 1×10^{-2},初始常数为 1×10^{-3}。对于给定次数的迭代,变换器更新变换后的样本将其提供给模型 D_i,并从这些模型接收反馈。在迭代集合结束时,具有目标针对性的攻击成功。使用所需迭代次数、人类识别度和失真量来评估变换后的样本 x^*。使用均方误差将失真测量修改后的样本与原始样本之间的像素距离缩小。

2.2 实验结果分析

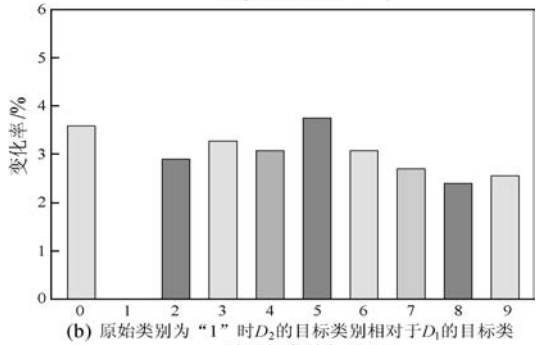
实验结果展示了多目标对抗性样本的图像、每个目标类的失真、多目标对抗性样本的类别得分、可伸缩性分析和多个目标类的信息。

不失一般性,以模型 D_1 和 D_2 为例进行相关结果分析与说明。对于模型 D_1,目标类别的得分"0"(5.09)略高于原始类(5.03)。对于模型 D_2,目标类别"8"(5.08)的得分略高于原始类别(5.03)的得分。为了减少多目标对抗性示例的失真,该结果表明多目标对抗性样本已经被修改,直到模型 A 和 B 的每个目标类别的得分高于原始类别的得分。

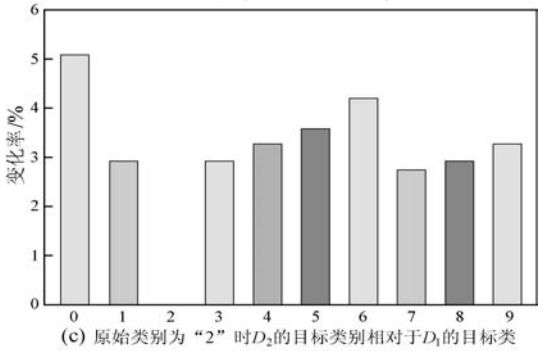
为了分析多目标对抗性样本的可扩展性,每组模型有 900 个随机多目标对抗性示例,针对其 100% 目标攻击成功率的平均失真和迭代次数进行分析比较。迭代模式说明,随着模型数量的增加,多目标对抗性样本需要更多的学习过程来攻击多个模型。另一方面,平均失真的模式表明随着模型数量的增加,失真增加,但变化率降低。从图 3 和图 4 中可以看到,每个目标类的多目标对抗性样本的失真是不同的。如果攻击者需要具有最小失真的多目标对抗样本,则此信息非常有用。



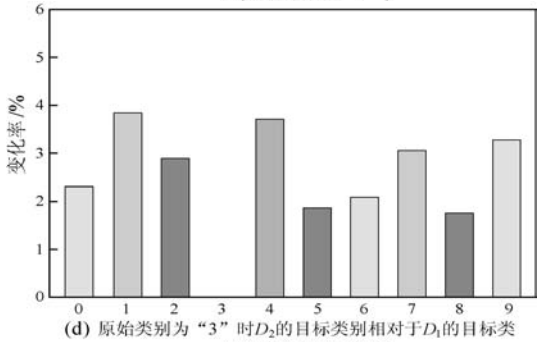
(a) 原始类别为"0"时D_2的目标类别相对于D_1的目标类别(目标类别为"9")



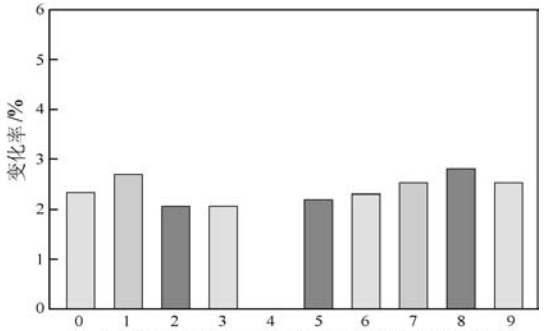
(b) 原始类别为"1"时D_2的目标类别相对于D_1的目标类别(目标类别为"8")



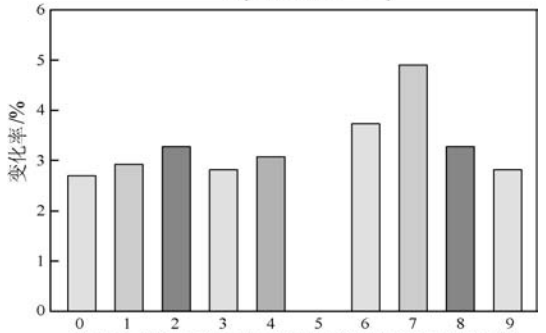
(c) 原始类别为"2"时D_2的目标类别相对于D_1的目标类别(目标类别为"7")



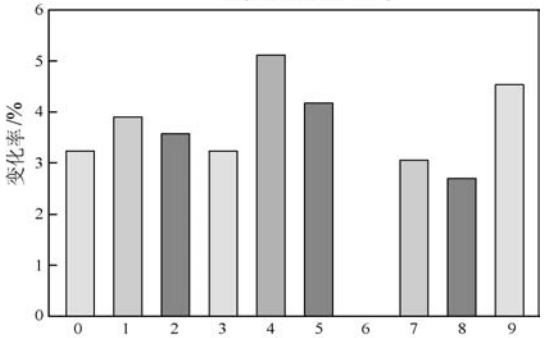
(d) 原始类别为"3"时D_2的目标类别相对于D_1的目标类别(目标类别为"6")



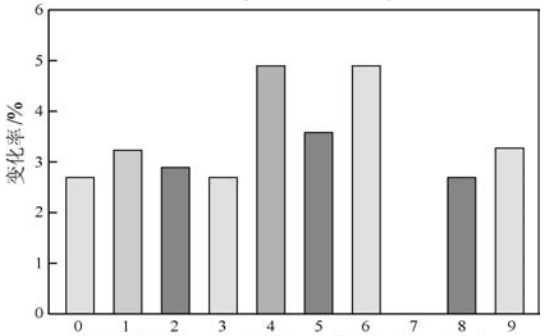
(e) 原始类别为“4”时D_2的目标类别相对于D_1的目标类别(目标类别为“5”)



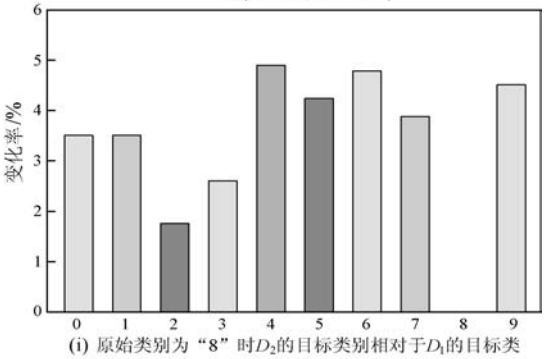
(f) 原始类别为“5”时D_2的目标类别相对于D_1的目标类别(目标类别为“4”)



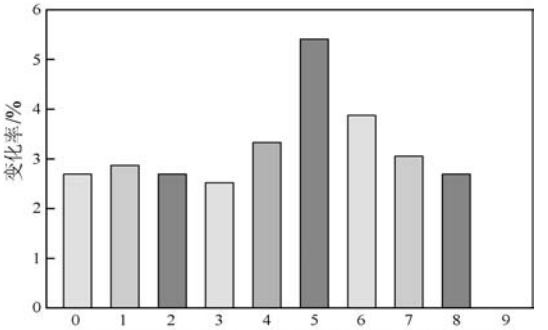
(g) 原始类别为“6”时D_2的目标类别相对于D_1的目标类别(目标类别为“3”)



(h) 原始类别为“7”时D_2的目标类别相对于D_1的目标类别(目标类别为“2”)

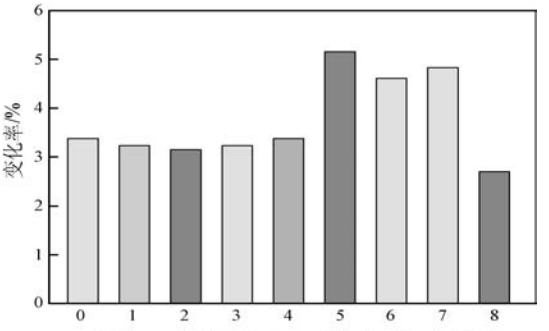


(i) 原始类别为“8”时D_2的目标类别相对于D_1的目标类别(目标类别为“1”)

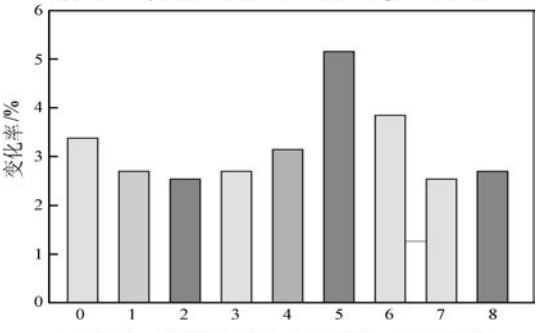


(j) 原始类别为“9”时D_2的目标类别相对于D_1的目标类别(目标类别为“0”)

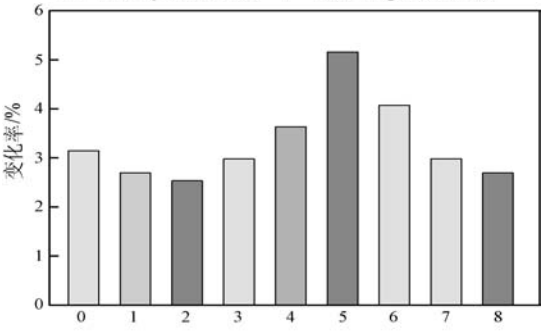
图 3 对于 D_2 和 D_1 的每个目标类别相对于原始样本的平均变化率(1 000 次迭代)



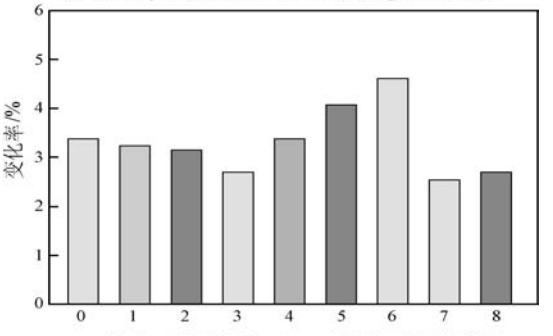
(a) 模型D_1目标类别为“0”时模型D_2的目标类别



(b) 模型D_1目标类别为“1”时模型D_2的目标类别



(c) 模型D_1目标类别为“2”时模型D_2的目标类别



(d) 模型D_1目标类别为“3”时模型D_2的目标类别

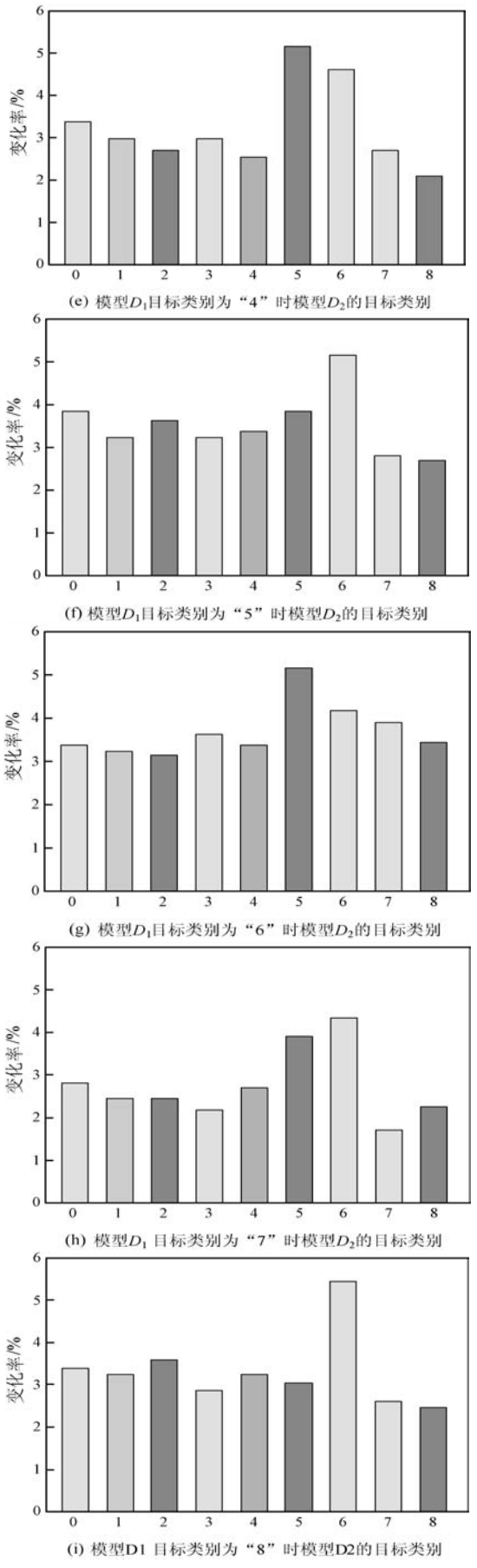


图 4　对于原始样本“9”,每个 D_1 和 D_2 模型的目标类别的平均变化率

2.3　实验结果对比

本文提出的深度神经网络结合蚁群算法的多目标对抗方法与文献[6]、文献[10]提出的多目标对抗方法相比,具有较大的优势。同样采用 2.1 节中数据集,表 2 和表 3 是采用文献[6]、文献[10]方法的多目标针对性对抗样本对于模型 D_1、D_2、D_3、D_4 和 D_5 中的每个目标类的类别得分。表 4 是本文提出的深度神经网络结合蚁群算法的多目标对抗方法中多目标对抗性样本模型 D_1、D_2、D_3、D_4 和 D_5 中的每个目标类的多目标对抗性样本的类别得分。通过对比可以看出,本文方法的得分整体高于文献[6]和文献[10]方法,多目标对抗性更好,更具有优势。这是因为引入蚁群算法,利用蚂蚁互相交流学习的正反馈原理保证算法的收敛性和寻优速度,保证方法的全局优化性能。

表 2　文献[6]方法对于模型 D_1、D_2、D_3、D_4 和 D_5 中的每个目标类的类别得分

描述	多目标对抗性样本的类别得分
D_1(目标“9”)	[1.232 −0.67 0.13 1.12 −1.33 −1.23 −1.19 1.06 1.13 **1.101**]
D_2(目标“0”)	[**5.11** −4.28 3.60 4.51 −8.29 2.78 0.57 1.91 −0.61 −3.36]
D_3(目标“1”)	[−2.21 **3.22** 3.11 3.86 −0.82 −1.62 2.35 −3.14 3.21 −6.69]
D_4(目标“2”)	[−2.61 −0.24 **5.98** 4.34 −7.19 −4.31 −2.11 2.11 2.11 −5.01]
D_5(目标“3”)	[−0.68 −1.06 5.06 **4.22** −5.99 1.22 −0.66 0.40 0.40 −4.97]

表 3　文献[10]方法对于模型 D_1、D_2、D_3、D_4 和 D_5 中的每个目标类的类别得分

描述	多目标对抗性样本的类别得分
D_1(目标“9”)	[1.162 −0.56 0.13 1.18 −1.36 −1.37 −1.39 1.06 1.13 **1.168**]
D_2(目标“0”)	[**5.52** −5.38 3.60 4.51 −8.29 1.78 0.59 1.82 −0.61 −4.36]
D_3(目标“1”)	[−2.29 **3.89** 3.11 3.86 −0.82 −1.62 2.35 −3.14 3.21 −7.69]
D_4(目标“2”)	[−2.67 −0.24 **6.68** 4.34 −7.89 −4.35 −2.12 2.18 2.13 −5.01]
D_5(目标“3”)	[−0.68 −1.06 5.06 **5.52** −6.98 1.48 −0.76 0.40 0.40 −4.97]

表 4 本文方法中多目标对抗性样本模型 D_1、D_2、D_3、D_4 和 D_5 中的每个目标类的类别得分

描述	多目标对抗性样本的类别得分
D_1（目标“9”）	[1.287 −0.61 0.16 1.27 −1.56 −1.38 −1.48 1.10 1.19 **1.288**]
D_2（目标“0”）	[**5.62** −5.46 3.70 4.59 −8.34 1.84 0.65 1.87 −0.67 −4.48]
D_3（目标“1”）	[−2.58 **3.99** 3.15 3.96 −0.89 −1.67 2.52 −3.19 3.21 −7.87]
D_4（目标“2”）	[−2.76 −0.27 **6.72** 4.45 −7.92 −4.55 −2.22 2.23 2.47 −5.06]
D_5（目标“3”）	[−0.72 −1.00 5.06 **5.62** −6.87 1.68 −0.67 0.39 0.41 −4.79]

3 结 语

本文提出一种深度神经网络结合蚁群算法躲避攻击的多目标对抗方法。实验结果表明，该方法比现有的躲避攻击多目标对抗方法具有明显的优势，实现100%的攻击成功率。未来的研究方向是继续将实验扩展到其他标准图像数据集中，并进行隐身中的失真与人类感知之间的相关性分析。利用软件方法进行全系统的闭环虚拟现实仿真，是一项行之有效的方法，它与利用硬件进行全系统的仿真相辅相成，未来将对这方面的工作进行更加深入的研究，同时关注本文算法对其他数据集的适用性。

参 考 文 献

[1] 钱铁云，王毅，张明明，等. 基于深度神经网络的入侵检测方法[J]. 华中科技大学学报（自然科学版），2018，46（1）：6－10.

[2] 王靖宇，王霰禹，张科，等. 基于深度神经网络的低空弱小无人机目标检测研究[J]. 西北工业大学学报，2018，36（2）：268－263.

[3] 杜海文，崔明朗，韩统，等. 基于多目标优化与强化学习的空战机动决策[J]. 北京航空航天大学学报，2018，44（11）：2247－2256.

[4] 姚宗信，李明，陈宗基. 多空中作战平台协同对抗多目标态势分析方法[J]. 系统工程与电子技术，2008，30（2）：292－296.

[5] 毛存礼，余正涛，沈韬，等. 基于深度神经网络的有色金属领域实体识别[J]. 计算机研究与发展，2015，52（11）：2451－2459.

[6] Belmecheri F，Prins C，Yalaoui F，et al. Particle swarm optimization algorithm for a vehicle routing problem with heterogeneous fleet，mixed backhauls，and time windows[J]. Journal of Intelligent Manufacturing，2013（24）：775－789.

[7] Kritikos M N，Ioannou G. The heterogeneous fleet vehicle routing problem with overloads and time windows[J]. International Journal of Production Economics，2013，144（1）：68－75.

[8] 闫明. 基于 DCT 变换的对抗样本防御方法研究[D]. 哈尔滨：哈尔滨工业大学，2018.

[9] Leung S C H，Zhang Z，Zhang D，et al. A meta-heuristic algorithm for heterogeneous fleet vehicle routing problems with two-dimensional loading constraints[J]. European Journal of Operational Research，2013，225（2）：199－210.

[10] Li X，Leung S C H，Tian P. A multistart adaptive memory-based tabu search algorithm for the heterogeneous fixed fleet open vehicle routing problem[J]. Expert Systems with Applications，2012，39（1）：365－374.

[11] 邢健飞，罗志增，席旭刚. 基于深度神经网络的实时人脸识别[J]. 杭州电子科技大学学报，2013（6）：107－110.

[12] Alshehri M D，Almutairi A T，Alomran A M，et al. Over-the-counter and prescription medications for acne：A cross-sectional survey in a sample of university students in saudi arabia[J]. Indian Dermatology Online Journal，2017，8（2）：120－132.

[13] 吕刚，郝平，盛建荣. 一种改进的深度神经网络在小图像分类中的应用研究[J]. 计算机应用与软件，2014，31（4）：182－184，213.

[14] 李健伟，曲长文，彭书娟，等. 基于生成对抗网络和线上难例挖掘的 SAR 图像舰船目标检测[J]. 电子与信息学报，2019，41（1）：143－149.

[15] 唐建强，李昊. 基于蚁群算法的对抗资源多目标分配[J]. 电子信息对抗技术，2018，33（2）：48－52.

[16] 杨观赐，杨静，李少波，等. 基于 Dopout 与 ADAM 优化器的改进 CNN 算法[J]. 华中科技大学学报（自然科学版），2018，46（7）：122－127.

（上接第 274 页）

[21] Yue L，Weibin Z，Lin S. Really should we pruning after model be totally trained? Pruning based on a small amount of training[EB]. arXiv：1901.08455，2019.

[22] Lee N，Ajanthan T，Torr P H S. SNIP：Single-shot network pruning based on connection sensitivity[EB]. arXiv：1810.02340，2018.

[23] Dong X Y，Huang J S，Yang Y，et al. More is less：A more complicated network with less inference complexity[C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. IEEE，2017：5840－5848.