

面向图像识别的卷积神经网络鲁棒性研究进展

林点, 潘理, 易平

(上海交通大学网络空间安全学院, 上海 200240)

摘要: 卷积神经网络是目前人工智能领域在图像识别与处理相关应用中的关键技术之一, 广泛的应用使其鲁棒性研究的重要性不断凸显。以往对于卷积神经网络鲁棒性的研究较为笼统, 且多集中在对抗鲁棒性方面。这难以更深入地研究神经网络鲁棒性的发生机制, 已经不适应人工智能的发展。引入神经科学的相关研究, 提出了视觉鲁棒性的概念, 通过研究神经网络模型与人类视觉系统的相似性, 揭示了神经网络鲁棒性的内在缺陷。回顾了近年来神经网络鲁棒性的研究现状, 并分析了神经网络模型缺乏鲁棒性的原因。神经网络缺乏鲁棒性体现在其对于微小扰动的敏感性, 其原因在于神经网络会更倾向于学习人类难以感知的高频信息用于计算和推理。而这部分高频信息很容易被扰动所破坏, 最终导致模型出现判断错误。传统鲁棒性的研究大多关注模型的数学性质, 无法突破神经网络的天然局限性。视觉鲁棒性在传统鲁棒性的概念上进行拓展。传统鲁棒性概念衡量模型对于失真变形的图像样本的辨识能力, 失真样本与原始干净样本在鲁棒模型上都能保持正确的输出。视觉鲁棒性衡量模型与人类判别能力的一致性。这需要将神经科学和心理学的研究方法、成果与人工智能相结合。回顾了神经科学在视觉领域的发展, 讨论了认知心理学的研究方法在神经网络鲁棒性研究上的应用。人类视觉系统在学习和抽象能力上具有优势, 神经网络模型在计算和记忆速度方面强于人类。人脑的生理结构与神经网络模型的逻辑结构的差异是导致神经网络鲁棒性问题的关键因素。视觉鲁棒性的研究需要对人类的视觉系统有更深刻的理解。揭示人类视觉系统与神经网络模型在认知机制上的差异, 并对算法进行有效的改进, 这是神经网络鲁棒性乃至人工智能算法的主要发展趋势。

关键词: 卷积神经网络; 图像识别; 鲁棒性; 对抗样本; 人类视觉

中图分类号: TP183

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2022037

Research on the robustness of convolutional neural networks in image recognition

LIN Dian, PAN Li, YI Ping

School of Cyber Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

Abstract: Convolutional neural network is one of the key technologies in the application of image recognition and

收稿日期: 2021-05-24; 修回日期: 2021-11-02

通信作者: 易平, yiping@sjtu.edu.cn

基金项目: 国家自然科学基金 (62172278)

Foundation Item: The National Natural Science Foundation of China (62172278)

引用格式: 林点, 潘理, 易平. 面向图像识别的卷积神经网络鲁棒性研究进展[J]. 网络与信息安全学报, 2022, 8(3): 111-122.

Citation Format: LIN D, PAN L, YI P. Research on the robustness of convolutional neural networks in image recognition[J]. Chinese Journal of Network and Information Security, 2022, 8(3): 111-122.

processing in artificial intelligence. Its wide application makes researches on its robustness more and more important. Previous researches on robustness of neural networks were too sweeping and most of them focused on adversarial robustness, which causes difficulty in further study in the mechanism of neural network robustness. The related researches of neuroscience were introduced and the concept of visual robustness was put forward. By studying the similarity and difference between neural network models and human visual system, the internal mechanism and faults of neural network robustness were revealed. The researches of neural network robustness in recent years were reviewed, and the reasons for the lack of robustness of neural network models were analyzed. The lack of robustness of neural networks is reflected in their sensitivity to small perturbations. The reason is that neural networks tend to learn high-frequency information for calculation and inference, which is difficult for humans to recognize. High-frequency information is easily affected by perturbations, and eventually causes mistakes of models. Previous researches on robustness mostly focused on mathematical properties of models, and were limited in the natural faults of neural networks. Visual robustness extends the traditional concept of robustness. The traditional concept of robustness measures the discrimination ability of models for distorted image examples. Distorted examples and clean examples can get correct outputs through robust models. Visual robustness measures the consistency between models and humans in discrimination ability. Visual robustness combines the research methods and achievements of neuroscience and psychology with artificial intelligence. The development of neuroscience in the field of vision were reviewed, and the application of research methods of cognitive psychology in neural network robustness were discussed. Human visual system has advantages in learning and abstract ability, while neural network models have better performance in calculation speed and memory. The difference between the physiological structure of human brain and the logical structure of neural network models is the key factor leading to the problem of robustness of neural networks. The research of visual robustness requires deeper understanding of human visual system. Revealing the differences in cognitive mechanism between human visual system and neural network models and effectively improving the algorithm are the development trends of neural network robustness and even artificial intelligence.

Keywords: convolutional neural network, image recognition, robustness, adversarial example, human vision

0 引言

机器学习的理论从 20 世纪中叶就已经被提出和实践,但是由于机能限制,当时机器学习在复杂问题上表现不尽如人意。而随着计算机科学的不断发展,计算机算力不断提升,以及大数据时代的到来,限制机器学习发展的因素被逐个解决。1998 年,LeCun^[1]提出了卷积神经网络(CNN, convolutional neural network),并使用一种 7 层的卷积神经网络结构 LeNet-5 成功解决了手写数字识别的难题。这意味着机器学习可以解决以图像识别为代表的复杂问题。2012 年, Krizhevsky 等^[2]提出了一种新的卷积神经网络结构 AlexNet,在 ImageNet 图像识别竞赛上以高正确率获得冠军。从此以后,以卷积神经网络为代表的深度学习概念被广泛研究和应用,各种新的网络结构和训练方法被提出,其图像识别能力也不断提高。

大量围绕卷积神经网络的研究被开展,大量以卷积神经网络为基础的应用被开发,这引出了人们对其鲁棒性的思考。对于神经网络而言,鲁棒性意味着模型在面对异常输入时仍有输出正确结果的能力。对抗样本^[3]的发现让所有神经网络暴露在被攻击的风险中,研究者开始致力于提出抵抗对抗样本的方法,这也是目前鲁棒性研究的主要方向。

神经网络就是用计算机来模拟人脑神经元,并将其相互连接构建一个模拟神经网络,以此实现“类人工智能”。神经网络算法使计算机能够在一定程度上模拟人类的记忆、运算、推理和识别能力。显然,提高神经网络与人脑的相似性是神经网络自诞生以来就始终追求的目标。理论上,神经网络的判别结果应当以人类的判别结果为准绳,本文提出“视觉鲁棒性”的概念,以衡量神经网络与人类视觉的相似性。

本文致力于对神经网络的鲁棒性做一个全面的讨论。对抗鲁棒性是目前研究的热点,相关的学术论文和综述讨论也较多,但相关研究正逐渐陷入困境。视觉鲁棒性是本文提出的用以区别传统鲁棒性的概念,并从神经科学与人工智能两个方向介绍其进展。本文对神经网络鲁棒性的发展现状进行总结,并对其发展趋势进行讨论与展望。此外,本文关注的是图像识别应用场景下的鲁棒性问题,该应用场景下所使用的网络结构绝大多数是卷积神经网络及其变体,如果不做额外说明,本文提到的神经网络均指卷积神经网络。

1 神经网络鲁棒性的研究现状

1.1 对抗样本和对抗鲁棒性

Szegedy 等^[3]发现,可以给输入图像上添加精心设计过而人类无法察觉的扰动,从而使神经网络模型发生误判,被添加过扰动的输入称作“对抗样本”。对抗鲁棒性是神经网络模型在对抗样本攻击下保持自身输出正确结果的能力,一般使用对抗样本的攻击成功率来衡量神经网络模型的对抗鲁棒性。

1.1.1 对抗样本的生成方法

(1) 白盒攻击场景

白盒攻击场景下,攻击者拥有目标模型的全部信息,包括训练集、模型参数、模型梯度、模型输出等。Goodfellow 等^[4]提出了快速梯度符号法(FGSM, fast gradient sign method)算法,通过计算模型损失函数的梯度符号并依此为图像添加扰动,以最小化分类模型的损失函数。此后,研究者提出了C&W算法^[5]、投影梯度下降(PGD, projected gradient descent)算法^[6]等攻击性和隐蔽性更强的算法。白盒攻击方法的存在,反映了模型鲁棒性上存在的重大缺陷,以至于攻击者可以无视人类对图像的视觉感知,只通过对梯度计算等数学方法使模型的判断出现错误,这是模型鲁棒性研究的开端。

(2) 黑盒攻击场景

Papernot 等^[7]提出了“黑盒攻击”的概念,使攻击者能够在不掌握目标模型具体信息的情况下生成对抗样本,并提供了两种核心思路:基于

对抗样本迁移性的算法和基于查询的算法。对抗样本迁移性指一些对抗样本能同时作用于不同神经网络模型的现象^[3,4,7],往往通过集成^[8]的方法实现。基于查询的算法则在于利用从目标模型接口查询得到的信息^[9]。相比白盒攻击,黑盒攻击更加贴合一些实用场景。黑盒攻击算法的存在,说明对抗鲁棒性无法通过保密模型信息的方式实现,必须提出有效的方法提高模型自身的鲁棒性。

1.1.2 对抗样本的防御方法

提升对抗鲁棒性,即防御对抗样本有两条基本的思路:其一是削弱对抗样本的攻击性,间接地提高模型防御对抗样本的能力;其二是通过对抗训练、修改网络结构的方式直接提高神经网络模型的对抗鲁棒性。对抗鲁棒性提升方法如表1所示。

(1) 消除对抗扰动

对抗扰动在图像上往往以高频噪声的方式呈现,所以采用降噪方法可以消除对抗扰动,常见的方法有图像压缩^[10]、去噪自编码器^[11]等。

(2) 对抗样本检测

有研究者认为,对抗样本相比干净样本在特征层面具有显著差异,可以据此识别出对抗样本,如对降噪前后的样本进行比较^[12],或直接使用深度神经网络进行检测^[13],但该方法对攻击力较强的攻击方法效果较差^[14]。

(3) 对抗训练

对抗训练^[4]核心思想来源于数据增强^[15],将对抗样本加入训练数据,改变训练集的数据分布,增强神经网络模型的鲁棒性。Tsipras 等^[16]认为对抗训练会提高模型鲁棒性,但会降低模型精度。Xie 等^[17]提出,对抗训练可以提高模型的精度,但是由于对抗样本具有与正常样本不同的分布,所以应在训练过程中给予两类样本不同的批标准化参数,且对抗训练要达到理想的鲁棒性,所需的模型深度远远超出一般模型。

(4) 生物启发模型

相比神经网络,人类视觉更具有鲁棒性,因此更接近人类视觉系统结构的模型应当更具有鲁棒性。Reddy 等^[18]在网络结构中加入了模拟人类视网膜和视皮层注视点的结构。Kim 等^[19]提出了模拟人类大脑的循环反馈结构的生物启发模型。

表 1 对抗鲁棒性提升方法比较
Table 1 Comparison of adversarial robustness enhancement methods

方法名称	优点	缺点
消除对抗扰动 ^[10-11]	与模型结构无关, 可以作为任何模型的输入层扩展模块, 时间开销小	难以控制消除扰动的程度, 程度过高会使图像丢失信息, 过低则防御能力不足
对抗样本检测 ^[12-13]	与模型结构无关, 可以作为任何模型的输入层扩展模块, 时间开销小	存在误报率和漏报率的冲突; 对于较强的攻击算法效果较差 ^[14]
对抗训练 ^[4]	目前最为有效的对抗鲁棒性提升方法, 理论上可以应对所有已知攻击	训练成本高, 可能使模型对于对抗样本过拟合, 要达到理想的鲁棒性需要更深的网络结构 ^[16-17]
生物启发模型 ^[18-19]	与人类视觉相适应的结构能够有效提高鲁棒性	模型往往较为复杂, 难以推广; 缺乏更深入的有效性研究; 需要脑科学领域的进一步验证

1.2 针对常见图像失真的鲁棒性

对抗样本必须有一个恶意的攻击者人为地生成, 而图像失真则无处不在, 如有损压缩、噪声、曝光异常、颜色失真、图像旋转等问题, 它们同样会导致神经网络模型出现误判。

几何变换失真包含图像的平移、旋转、缩放、翻转、变形等, 一般不改变图像的像素值, 但会改变图像像素的分布情况。几何不变性反映了一个模型接收经过几何变换后的输入时保持自身输出不变的能力^[20]。卷积神经网络具有平移不变性, 因为网络中卷积层与池化层的特点决定了卷积神经网络可以保证在识别图像中不同位置上的同一特征的一致性; 同时, 这一结构特性使其无法对旋转、缩放等变换保持不变性^[21]。

胶囊网络由 Hinton 等^[22]提出, 使用向量代替标量作为一层的输出, 使模型具有更强的空间不变性。然而, 由于胶囊层结构本质上属于一种全连接结构, 当神经元数量增加时, 参数数量会显著增大, 难以应用于复杂数据集^[23]。经过若干版本的迭代, 胶囊网络的性能有了显著提高^[24-25], 考虑到其结构本身具有的几何不变性, 它依旧是十分具有研究价值的网络结构。

像素扰动失真一般不改变像素的几何分布, 而是直接改变像素的数值。Hendrycks 等^[26]建立了一个失真图像数据集, 包含若干种常见的像素扰动失真, 包括噪声、模糊、天气影响、数字化处理等, 并提出了应对这些失真的有效方法。

1.3 神经网络模型缺乏鲁棒性的原因

对抗样本的广泛存在和迁移性说明, 目前几乎所有的深度神经网络不具备足够的鲁棒性, 且深度神经网络算法本身的不完善, 很有可能是这

种脆弱性的源头。

理想中的分类模型往往是连续而平滑的——充分训练的模型会将相近的样本判别为同一个分类。然而, Szegedy 等^[3]认为, 深度神经网络并不平滑, 对抗样本是输入样本的低维流形中的“不连续口袋”, 且泛化良好的模型中不存在对抗样本。但 Goodfellow 等^[4]认为, 线性才是对抗样本存在的原因, 一个有着巨大输入维数的神经网络会放大输入的微小改变, 最终导致输出的巨大变化。这两种观点从模型的数学性质角度讨论其鲁棒性, 成为后来许多梯度掩蔽^[27]防御方法的理论基础, 但简单地将鲁棒性归结于模型的线性与非线性, 难以反映复杂的深度神经网络的鲁棒性本质。因此, 梯度掩蔽方法被认为不可靠^[28]。

McDaniel 等^[29]认为对抗样本源自机器模型分类边界与真实数据边界之间的差别, 其原因是训练数据集与真实数据的偏差。这个观点是文献[3]中观点的延伸, 认为模型鲁棒性与泛化能力是相关的。

Stutz 等^[30]认为, 正常的输入样本存在于高维输入空间的低维流形, 而对抗样本则处在流形之外, 泛化问题则讨论的是模型对于流形上样本的判别能力, 对抗样本问题则相反。这个观点区分了泛化能力和对抗鲁棒性的概念。

Ilyas 等^[31]认为, 对抗样本并不是一种错误, 而是一种非鲁棒性的特征, 如图 1 所示。这些特征可以完全脱离原图而存在, 并以高置信度造成神经网络的错误识别。对抗扰动被视为一种可利用的特征, 而神经网络模型似乎总是倾向于优先提取对抗扰动特征作为判别的基础。

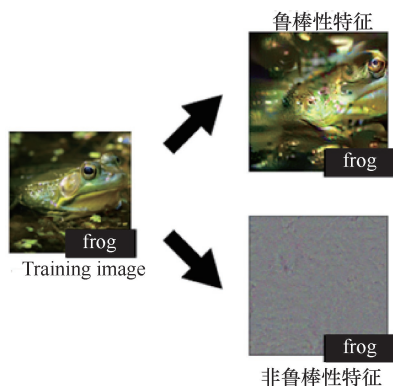


图1 鲁棒性特征与非鲁棒性特征
Figure 1 Robust feature and non-robust feature

将文献[30]中的流形观点与文献[31]中的非鲁棒性特征观点结合, 可以更加接近对抗样本的本质。深度神经网络提供端到端的识别模式, 简化了特征工程的工作, 这使人们无法解析海量参数的含义, 也就无法准确地得知深度神经网络究竟是以什么特征为依据进行推理和判断。

图像识别任务的特点之一是输入数据维数高, 其中包含的必要信息量的维数小于输入数据的维数, 多出来的维数, 存储了不为人类所认知的冗余信息, 即文献[31]中的非鲁棒性特征。考虑到总维数可能远大于流形维数, 冗余信息也会对神经网络有重大影响。事实上, 如果单纯地讨论神经网络模型对自然样本的分类任务, 这些冗余信息未必完全有害, 甚至可能提高分类准确率^[32], 毕竟自然样本中并不存在对抗样本。但这些冗余信息是难以被人类所感知的, 这就必然造成神经网络模型与人类认知之间的分歧。神经网络的鲁棒性或许体现在这种对于不同类型特征的认知偏好中, 即接近人类认知偏好的模型更具有鲁棒性。

1.4 神经网络鲁棒性的研究困境

对抗样本自提出以来就受到了广泛的关注, 以至于“鲁棒性”的概念常常被用来特指“对抗鲁棒性”。但是这一方向上的研究愈发陷入一种“军备竞赛”^[33]。对抗训练和数据增强^[15,34]方法依旧是最有效的能够稳定提高模型鲁棒性的方法, 但这类方法意味着大量时间和算力的消耗, 也可能导致模型在原始数据上欠拟合^[35], 在许多场合下并不是一个优先的选择。

而其他方法则像是在走“捷径”, 以低成本提高模型鲁棒性, 缺点是防御能力十分有限。以

对抗样本的降噪和检测方法为例, 这类方法总体可以分为两类: 基于特征工程的方法和基于深度学习的方法。前者尝试用传统的方法高效地寻找和排除对抗样本的影响, 但由于深度神经网络的复杂性和糟糕的可解释性, 这些方法很容易受到人类认知局限性的影响而不能有效解决问题, 即非“深度”方法很难解决“深度”神经网络的问题。后者使用深度学习模型的同时想要限制时间和算力上的开销, 但简单模型根本无法处理复杂的对抗样本, 而使用复杂模型来处理和识别对抗样本, 又成了另一种形式的对抗训练, 此时讨论回到了鲁棒性和训练成本的权衡上。因此, 鲁棒性的研究需要探索一条新的道路。

2 视觉鲁棒性

目前最有效的鲁棒性提升的方法是数据增强, 这和研究神经网络的出发点并不一致——神经网络本质是为了模拟人类的大脑, 从而使机器模型具备人类的认知能力, 但是人类并不需要通过数据增强来提高自己的认知能力, 因为人类的学习天然具有鲁棒性。讨论视觉鲁棒性时, 并不是讨论模型处理某类特殊输入的能力, 而是讨论其与人类视觉的相似性: 模型判断与人类认知相一致时体现出高鲁棒性, 反之则体现出低鲁棒性。研究模型鲁棒性的目的, 是使模型具有和人类相近的判别能力。

视觉鲁棒性与前文提到的鲁棒性概念并不冲突, 而是更进一步的讨论。例如, 当讨论对抗鲁棒性时, 常常有一个前提条件, 即对抗样本中包含的扰动是微小的, 人类难以察觉这些扰动。但是如果对抗扰动幅度增大到人类可认知的程度, 甚至能够改变图像中一部分特征时, 应当如何讨论模型的鲁棒性? 例如, Hosseini 等^[36]在图像的颜色空间进行扰动, 生成了具有语义特征的对抗样本, 这样的扰动可以轻易被人类感知, 但又没有显著改变图像的分类特征。此时, 如果脱离人类视觉系统本身的特点, 只是从数学或计算机的角度讨论和优化模型的鲁棒性, 往往不能得出令人信服的结果。

这里给出视觉鲁棒性的形式化定义。给定图像分类模型 F , 人类标注员 H , 输入数据全体为

D , 则 $\forall x \in D$, 有

$$F(x) = H(x)$$

显然, 对于任意变换 A , 当 $A(x) \in D$, 则

$$F(A(x)) = H(A(x))$$

即对于任意可被标注的输入数据, 具有视觉鲁棒性的模型输出应当与人类严格保持一致。视觉鲁棒性的定义中不再限制变换的效果, 且一旦数据的变化导致人类对数据的标注发生改变时, 模型相应做出改变, 如图 2 所示。

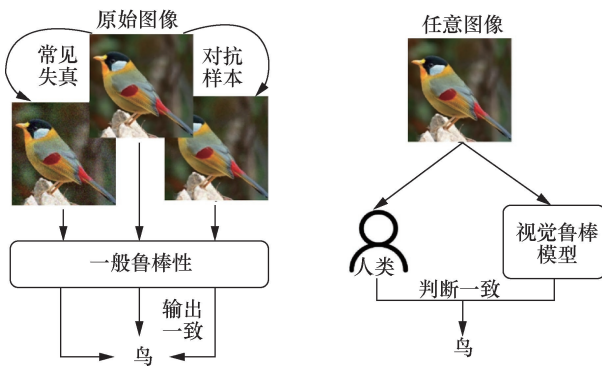


图 2 一般鲁棒模型 (左) 与视觉鲁棒模型 (右) 的差异
Figure 2 Difference between common robust model (left) and visual robust model (right)

进一步地, 考虑到视觉鲁棒性的定义涉及人类的主观判断, 可能在人类标注员之间存在差异。例如, 对于一张介于两个分类之间的 (非自然) 图像, 不同的人可能做出不同的判断。可以进一步优化上述定义, 使其符合现实情况。给定图像分类模型 F , 人类标注员 H , 输入数据全体为 D , 则 $\forall x \in D$, y 为任意有效分类, E 表示数学期望, 对于给定的鲁棒性阈值 $\varepsilon > 0$, 有

$$|F(y|x) - E(H(y|x))| < \varepsilon$$

其中, $F(y|x)$ 表示分类模型在输入图像 x 时在分类 y 上得出的置信度, $H(y|x)$ 表示人类标注员将图像 x 判断为分类 y 的概率。在该定义下, 一个具有鲁棒性的模型对于任意输入所给出的在有效分类下的置信度应当与人类的判断结果在统计学上相近。

2.1 基于神经科学的人类视觉系统研究进展

1962 年, Hubel 等^[37]提出了视觉皮层概念, 并系统阐述了其作用机制。1992 年, Goodale 等^[38]提出, 大脑中存在“两条通路”: 视觉信号进入大

脑后, 经由初级视皮层处理, 一部分会沿着大脑背侧的通道传到顶叶皮层, 称为背侧通路, 又叫“where 通路”, 主要负责运动、空间位置识别等; 另一部分则沿着腹侧的通道传导到下颞叶皮层 (IT cortex, inferior temporal cortex), 称为腹侧通路, 又叫“what 通路”, 主要负责物体识别、记忆等。

“what 通路”大致经过初级视皮层 V1、次级视皮层 V2、高级视皮层 V4 等脑区^[38], 视觉信号传输过程中, 大脑会逐层地由低到高地提取其中的各种特征, 以供更高级的脑区进行处理, 如图 3 所示。虽然上述描述非常简单, 但大脑视觉皮层的复杂度远超常人想象, 核磁共振的实验结果表明, 视觉皮层中至少包含 30 个功能区域, 每个区域都能够独立执行特定的视觉处理任务^[39]。目前, 相关的神经科学研究主要集中在少数较为初级的视皮层中, 尚待进一步发掘。

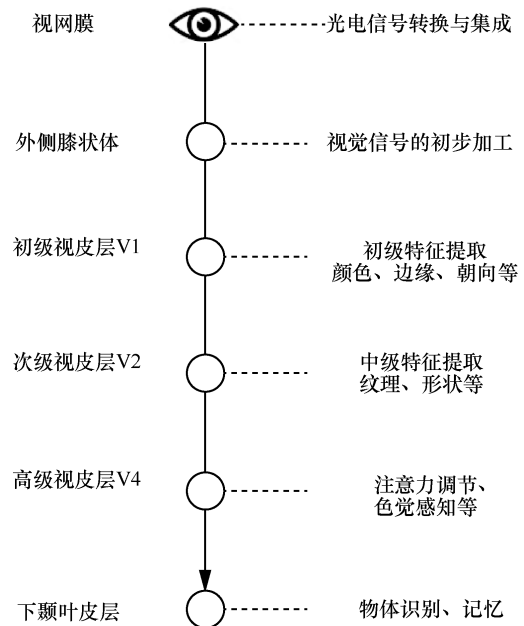


图 3 “what 通路”涉及的主要区域及其大致功能
Figure 3 Areas in “what pathway” and their general functions

下颞叶皮层是大脑中识别行为的核心区域, 视觉信号经过各个视觉皮层的处理之后会汇总到此处, 进行识别和记忆。其在微观上可以分解为若干毫米级的区域, 每一个小区域都会对物体的识别产生特定影响^[40]。大脑对于视觉信号具有相当强的高级特征提取能力。例如, 人脑对于特定刺激有额外的强化, 如人脸、身体、文字等^[41],

因此人类可以快速地识别出面部图像。Konkle 等^[42]发现, 大脑会根据被识别物体在现实世界中的正常尺寸对物体进行聚类分析, 大小相近的物体会激活的脑区也较近。Kriegeskorte 等^[43]发现, 下颞叶皮层对人脸和动物具有相似的响应模式, 而对人造物和植物等则有另一种相似的响应模式, 说明其可以分辨被识别物体的“活力程度”(animate or inanimate)。Proklova 等^[44]对此进行了进一步的研究, 发现轮廓和纹理等初级视觉特征并不能完全解释这一机制。

2.2 人脑视觉皮层与神经网络的结构相似性

从线性判别模型^[45], 到神经网络模型^[46]再到卷积神经网络^[1], 机器学习算法在发展过程中不断地提高自己与大脑视觉结构的相似性。脑科学研究者也看到了这一相似性, 开始尝试从神经网络中获取科研的灵感, 来进一步研究人脑视觉皮层。

Eickenberg 等^[47]与 Horikawa 等^[48]分别尝试使用卷积神经网络对人类大脑的核磁共振信号进行编码, 从而证明人脑的视觉神经信号与卷积神经网络具有相关性。St-Yves 等^[49]研究了人脑视觉系统的感受野特征, 提出人脑中也可能存在类似于卷积神经网络中不同尺寸的特征图机制。Wen 等^[50]研究了人类动态视觉的核磁共振影像与卷积神经网络的相关性。Cadieu 等^[51]的研究表明, 即使不是生物启发模型, 深度神经网络在视觉对象识别任务中的表现与 IT 皮层相当。Bashivan 等^[52]构造了卷积神经网络模型到人脑视觉皮层的映射, 并成功通过神经网络模型构造样本控制了受试猴大脑中特定神经元的激活。这些基于脑科学的研究说明, 人脑的视觉机制与卷积神经网络在结构、功能等各个方面确实具有一定的相似性, 这为深度神经网络算法的进一步发展提供了理论基础。

2.3 对抗样本与人类视觉

人类视觉比神经网络更具有鲁棒性。例如, 人类可以识别更加局部的图像特征^[53], 面对失真图像也能保持比 CNN 更高的识别正确率^[34]。对抗样本是衡量神经网络模型鲁棒性的重要工具, 但在定义上与人类视觉不相容。Elsayed 等^[54]提出的观点为对抗样本的研究提供了一种新思路, 他

们拓展了传统对于对抗样本的定义, 使之适用于人类视觉: 对抗扰动不再是人类无法察觉的, 而是有可能对人类的认知过程产生影响, 此时必须考虑和研究对抗扰动与人类视觉之间的互动。他们根据这一定义, 生成了多模型大扰动的对抗样本, 发现有目标的攻击算法可能会导致人类对图像的判断受到影响^[54]。

Zhou 等^[55]在人类与神经网络模型的认知一致性上做了广泛的实验。他们生成了无实际意义但会被神经网络模型以高置信度分类的“愚弄”图像, 并研究人类与机器模型的判断, 发现人类以显著的高概率给出与机器模型相同的判断; 在纯扰动图像和对抗样本图像的识别实验中, 人类表现出与机器模型极高的一致性^[55]。这一系列的实验说明, 人类视觉系统相比卷积神经网络模型, 除了表现出极高的鲁棒性之外, 还能够以一种近乎直觉的方式“理解”卷积神经网络模型的非鲁棒性。

Santurkar 等^[56]基于经过对抗训练的鲁棒性模型进行了图像生成实验, 他们使用 PGD 算法^[6]分别对正常样本、随机初始化图像、区域损坏图像和草图图像等进行有目标的攻击、优化或转换, 成功在原始图像上添加和生成了人类可识别的语义特征。这说明, 鲁棒的神经网络模型能够学习到更多接近于人类认知的特征, 鲁棒模型的损失梯度与人类的感知一致^[15]。

对抗样本最初是专为神经网络模型设计的攻击手段, 但当允许其被人类所感知时, 它能够反映神经网络模型与人类视觉机制在某种限度上的一致性。这样的研究突破了传统上对于对抗样本的研究仅限于对神经网络模型进行形式化和数学化分析的范畴, 而开始触及基于神经网络的人工智能技术的本质。神经网络模型和人类视觉都在图像中学习可用的特征用于图像识别, 但生理机能的限制使人类与机器对特征的感知和选择存在差异。神经网络模型鲁棒性的提升意味着其对特征的选择更加接近于人类, 即从非鲁棒性的特征转向鲁棒性的特征^[31]。

2.4 神经网络的特征偏好

Ritter 等^[57]使用认知心理学的研究方法研究人类与神经网络模型对于颜色特征与形状特征的

感知,实验结果说明,在这两个特征上,人类视觉与神经网络表现出一致性,都偏好于形状特征。Hosseini 等^[58]则进一步评估了 CNN 在颜色和形状上的偏好,他们在保持形状的前提下对输入图像进行不同的颜色处理,发现卷积神经网络并不是天然对形状鲁棒的,而是需要合适的数据集和训练的支撑。

相比颜色,图像的纹理特征包含了更多更复杂的视觉信息。Geirhos 等^[59]对图像进行了去除颜色、去除纹理等变换,并交给人类和机器模型进行判断,发现颜色信息对二者的判断并不造成太大影响,但在去除形状信息的图像上,人类的判断准确率远远高于机器模型,这说明相比人类,机器模型极大地依赖于图像的纹理信息。他们使用风格迁移算法^[60],将不同图像的纹理信息与形状信息相融合,进一步论证了人类在图像识别中偏向于形状特征,而机器模型更依赖纹理特征,并证明使用去纹理化的数据集可以改善机器模型的这一偏向,使机器模型更具有鲁棒性^[59]。

Wang 等^[61]的研究发现, CNN 在图像识别任务中注重高频分量,而人类注重低频分量和鲁棒性, CNN 对高频特征的重视提升了其训练准确率,但代价是鲁棒性和泛化能力的降低,放弃高频特征会导致准确率下降,这和文献[31]的观点相通;通过研究普通模型和对抗鲁棒性模型特征层的差异,发现对抗鲁棒性模型的卷积核比普通模型更加平滑,说明对抗鲁棒的卷积模型不容易受到对抗扰动的影响。

CNN 对高频纹理特征的偏好可能是其鲁棒性较差的原因之一,研究者也尝试改善这一特性。Shi 等^[62]认为纹理特征会在小区域内以高概率重复出现从而降低自信息率,提出信息丢弃方法来降低模型对纹理信息的获取,实验证明,该方法可以提高模型对加性噪声失真和对抗样本的鲁棒性。Li 等^[63]考虑到高频纹理特征和低频的形状特征都在 CNN 的图像识别中起到重要作用,任何偏见都会导致网络的性能下降,所以提出形状-纹理去偏见化的训练方法。他们使用风格迁移算法^[60]将两个不同类型的图像进行混合,得到混合图像作为训练数据;标签方面,不使用硬标签,而使用软标签,同时标识图像的纹理信息和形状

信息的来源;该方法在 ResNet 模型和 ImageNet 数据集上同时提高了模型的分类正确率和对抗鲁棒性^[63]。

2.5 人类视觉系统的相对优势

相比人工神经网络,人类的视觉系统具有独特的优势。

一是其精细的模块划分。初级视觉皮层提取出的特定视觉特征会以生物电的形式传送到高级视觉皮层的特定位置,且仅激活特定脑区而不影响其他部位。这使大脑能够对信息进行针对性扬弃,提高处理效率,降低能耗。大脑对于一些较为低级的特征也有精细的处理,如颜色会在 V1 皮层进行初步处理,而直到 V4 皮层才会生成对颜色色相的感知^[64]。

二是人脑具有极强的抽象能力。大脑在进行物体识别之前,会根据一定的特征对视觉信号进行聚类分析,如前文提到的“现实世界尺寸^[42]”或“活力程度^[43]”,而这类特征是如何抽象出来的还尚待研究。人类对于脸的识别则是大脑抽象能力的集中体现,一个刚学会画画的孩子画人脸时,往往会用抽象的线条来呈现,而其他人类不需要任何额外的辨认即可知道这是一张脸,这是现在的神经网络所做不到的。

视觉皮层进化出这些特征可能是对视觉系统本身的补偿。人类视网膜中有两类接收光信号的细胞(视锥细胞和视杆细胞),前者决定了在正常光线情况下的视觉分辨率。人类约有 460 万个视锥细胞,其中绝大多数集中在视网膜的中央凹处,其他部位则分布稀疏^[65],无法提供足够的分辨能力。这一生理机能限制导致人类只能清晰地识别一小部分视野,而大脑则用强大的特征提取算法来优化视觉。

2.6 关于神经网络鲁棒性的进一步讨论

可以得出这样一个结论:在图像识别的应用场景下,卷积神经网络倾向于学习高频的纹理特征,而人类的生理特性决定了其对高频特征不敏感,这使神经网络在认知层面与人类存在偏差。而脆弱的高频特征更容易被扰动破坏,这是神经网络缺乏鲁棒性的根本原因。

跳出图像识别这一领域,再讨论神经网络与人类的差异,会发现神经网络应用场景有一个特

点：它擅长完成对于人类来说十分困难的计算和记忆任务，如下围棋、语言翻译等；而对于模式识别类问题，如图像识别、语音识别等，人类可以轻松完成这些任务，但神经网络的表现却与人类有很大差距。考察这两类问题的特点，会发现前者问题虽然复杂，但是在有限和离散的输入维度之内进行计算，如下围棋只需要考虑每一步361个点的状态，语言翻译的场景下单词的数量也是有限的，现有机器的算力足以应对这些场景。而后的输入维度，以224×224像素的RGB图像识别为例，假设每一个像素点的单个颜色深度是8位，那么整个输入空间大小为 $256^{150\ 528}$ ，近乎是无限的。人类由于本身生理机能的限制，只能获取精度有限的视觉信息，但大脑的抽象能力使人类能轻易地掌握图像中物体所具有的高层次特征，而不至于被细节所误导。但神经网络的结构决定了它必须接受和处理图像中的每一个像素点，优点是可以使用细节来提高自己的准确度，缺点在于算力不足以支持细粒度分析的情况下，这些细节会导致其有效特征提取不足，对高级、抽象的特征缺乏感知，最终体现为鲁棒性的缺失。

另外，神经网络的鲁棒性与具体的应用场景和目标任务也有关。对于下围棋这一任务而言，一方面其评价标准是客观的，另一方面数据中几乎所有的信息是有效的，此时神经网络强大的记忆和计算能力使其比人类具有更大的优势，所以基于神经网络的人工智能比人类更具有鲁棒性。而复杂模式识别任务的评价标准较为主观，神经网络在抽象能力方面的不足使其难以学习足够的有效信息，从而表现为鲁棒性的缺失。神经网络模型与人类视觉系统比较如表2所示。换言之，目前的神经网络结构足以胜任类似于下围棋这样的任务，但尚不能完美地解决复杂模式识别问题，仍需要优化和改进。综上所述，对于复杂模式识别问题，要提高神经网络模型鲁棒性，应当从抽象能力出发，让神经网络能够有效地学习鲁棒的高级语义特征。

3 鲁棒性研究展望

鲁棒性的研究正处在一个交叉路口，其中一条路是继续分析模型的数学性质，建立新的更加

复杂的数学模型来描述模型的鲁棒性，这也是大多数人所走的路；另一条路则是从人类视觉系统中寻找灵感，这条路更加困难，但无可否认的是，以往人工智能算法的重大革新正是设计出了与人类神经系统更相似的模型^[1,45-46]。在这一方向上，有互相交叉的两条路径：一是神经科学的研究，直接从大脑结构出发，讨论人脑与神经网络的关系；另一条是以机器学习模型为基础，先提出新的算法或模型结构，再设计人类参与的心理实验来评估模型或者算法的效果。事实上，由于人类的大脑很难被直接研究，目前神经科学领域的研究要借助动物实验。在脑科学的深入探索方面，人工智能学家受到的限制更少，可能在某些方面比脑科学家更有优势，甚至给予脑科学意想不到的启发。

表2 神经网络模型与人类视觉系统比较
Table 2 Comparison of neural network and human visual system

比较项目	神经网络模型	人类视觉系统
信号载体	数据流	生物电
信号操作	卷积操作	兴奋、抑制
抽象能力	弱	强
学习速度	慢	快
计算速度	快	慢
识别速度	快	慢
记忆能力	强	弱

在这一方向上的研究，可以思考以下两个问题。

(1) 先验还是后验

对于目前的神经网络模型，其训练的超参数和网络结构本身是先验的，训练得到的参数则是后验的。人类大脑具有复杂的结构，每一个脑区又能够分解成若干子功能分区，一个重要的课题就是研究在这些分区中，哪些结构是由基因先天决定的，哪些能力是经过后天学习获得的。在这些问题上，神经科学在初级的特征提取方面有了初步的成果，但在高级特征的分类和识别上仍有待探索^[49]。如果能用一定的方法将相关的结论运用在神经网络的结构中，甚至做出进一步优化，必然能够有效地提高神经网络的效率和鲁棒性。He等^[66]提出的ResNet网络是一个典型的例子，

其中的残差结构能够让网络通过学习来降低不合理的网络结构带来的负面影响。

(2) 模块化还是高耦合

目前的神经网络模型的一大特点是端到端的训练：输入端无须进行任何的特征提取，直接输入经过简单处理的原始数据，然后在输出端获取结果。这样的好处是节省了大量人力成本，缺陷则是人们对训练得到的参数缺乏认识，难以找到模型中的缺点并进行改良。所有的数据和参数混在一起无法区分，这是高耦合的典型特征。如果能够参考人脑视觉皮层中结构特性，对神经网络模型结构进行模块化的改进，或许能够提升模型的性能和可解释性。Szegedy 等^[67]提出的 Inception 网络结构中使用不同大小的卷积核提取不同尺度的特征，具有类似的效果。

4 结束语

对于机器学习的研究，大致上可以分成 3 个方向：提升通用任务的性能、对特定任务进行优化、提升鲁棒性。前两个方向的研究，已经在许多方面达成“超越人类”的效果，而鲁棒性研究则一直是短板。大多数鲁棒性研究的关注点在对抗鲁棒性上，体现为各种攻击方法和防御方法的“矛盾”之争。对抗训练效果显著，但由于其成本太高，无法推广。

模型的视觉鲁棒性是本文在传统的鲁棒性研究的基础上提出的概念，研究人类视觉系统与神经网络模型的一致性，是对鲁棒性更深入的讨论。学术界对于鲁棒性的研究尚未得出一个公认的结论，但就目前的趋势而言，相关研究最终要“以人为本”，回归人类与模型的差异，而根本上的解决方法是提出更接近人类视觉机制的机器学习模型和算法。总而言之，对于鲁棒性的研究是接下来机器学习算法研究的重中之重，需要研究者在这一领域更有创造性地工作。

参考文献：

- [1] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.
- [2] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60:84-90.
- [3] SZEGEDY C, ZAREMBA W, SUTSKEVER I, et al. Intriguing properties of neural networks[C]//2nd International Conference on Learning Representations, ICLR 2014.
- [4] GOODFELLOW I J, SHLENS J, SZEGEDY C. Explaining and harnessing adversarial examples[C]//3rd International Conference on Learning Representations, ICLR 2015.
- [5] CARLINI N, WAGNER D. Towards evaluating the robustness of neural networks[C]//2017 IEEE Symposium on Security and Privacy (SP). IEEE, 2017: 39-57.
- [6] MADRY A, MAKELOV A, SCHMIDT L, et al. Towards deep learning models resistant to adversarial attacks[C]//International Conference on Learning Representations. 2018.
- [7] PAPERNOT N, MCDANIEL P, GOODFELLOW I, et al. Practical black-box attacks against machine learning[C]//Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security. 2017: 506-519.
- [8] LIU Y, CHEN X, LIU C, et al. Delving into transferable adversarial examples and black-box attacks[C]//International Conference on Learning Representations. 2017.
- [9] BRENDDEL W, RAUBER J, BETHGE M. Decision-based adversarial attacks: reliable attacks against black-box machine learning models[C]//International Conference on Learning Representations. 2018.
- [10] DZIUGAITE G K, GHAMRANI Z, ROY D M. A study of the effect of jpg compression on adversarial images[J]. arXiv preprint arXiv:1608.00853, 2016.
- [11] VINCENT P, LAROCHELLE H, BENGIO Y, et al. Extracting and composing robust features with denoising autoencoders[C]//Proceedings of the 25th International Conference on Machine Learning. 2008: 1096-1103.
- [12] XU W, EVANS D, QI Y. Feature squeezing: Detecting adversarial examples in deep neural networks[J]. arXiv preprint arXiv:1704.01155, 2017.
- [13] GROSSE K, MANOHARAN P, PAPERNOT N, et al. On the (statistical) detection of adversarial examples[J]. arXiv pre-print arXiv:1702.06280, 2017.
- [14] CARLINI N, WAGNER D. Adversarial examples are not easily detected: Bypassing ten detection methods[C]//Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security. 2017: 3-14.
- [15] SHORTEN C, KHOSHGOFTAAR T M. A survey on image data augmentation for deep learning[J]. Journal of Big Data, 2019, 6(1): 1-48.
- [16] TSIPRAS D, SANTURKAR S, ENGSTROM L, et al. Robustness May Be at Odds with Accuracy[C]//International Conference on Learning Representations. 2019.
- [17] XIE C, TAN M, GONG B, et al. Adversarial examples improve image recognition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 819-828.
- [18] REDDY M V, BANBURSKI A, PANT N, et al. Biologically Inspired Mechanisms for Adversarial Robustness[J]. Advances in Neural Information Processing Systems, 2020, 33.
- [19] KIM E, REGO J, WATKINS Y, et al. Modeling biological immuni-

- ty to adversarial examples[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 4666-4675.
- [20] FAWZI A, FROSSARD P. Manitest: Are classifiers really invariant?[C]//British Machine Vision Conference (BMVC). 2015: 106.1-106.13.
- [21] LENC K, VEDALDI A. Understanding image representations by measuring their equivariance and equivalence[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 991-999.
- [22] HINTON G E, KRIZHEVSKY A, WANG S D. Transforming auto-encoders[C]//International Conference on Artificial Neural Networks. Springer, Berlin, Heidelberg, 2011: 44-51.
- [23] PATRICK M K, ADEKOYA A F, MIGHTY A A, et al. Capsule networks—a survey[J]. *Journal of King Saud University-Computer and Information Sciences*, 2022, 34(1): 1295-1310.
- [24] PHONG N H, RIBEIRO B. Advanced capsule networks via context awareness[J]. *Lecture Notes in Computer Science*. 2019: 166-177.
- [25] RAJASEGARAN J, JAYASUNDARA V, JAYASEKARA S, et al. Deepcaps: Going deeper with capsule networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 10725-10733.
- [26] HENDRYCKS D, DIETTERICH T. Benchmarking Neural Network Robustness to Common Corruptions and Perturbations[C]//International Conference on Learning Representations. 2018.
- [27] PAPERNOT N, MCDANIEL P, WU X, et al. Distillation as a defense to adversarial perturbations against deep neural networks[C]//2016 IEEE Symposium on Security and Privacy (SP). IEEE, 2016: 582-597.
- [28] ATHALYE A, CARLINI N, WAGNER D. Obfuscated gradients give a false sense of security: Circumventing defenses to adversarial examples[C]//International Conference on Machine Learning. PMLR, 2018: 274-283.
- [29] MCDANIEL P, PAPERNOT N, CELIK Z B. Machine learning in adversarial settings[J]. *IEEE Security & Privacy*, 2016, 14(3): 68-72.
- [30] STUTZ D, HEIN M, SCHIELE B. Disentangling adversarial robustness and generalization[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 6976-6987.
- [31] ILYAS A, SANTURKAR S, ENGSTROM L, et al. Adversarial examples are not bugs, they are features[J]. *Advances in Neural Information Processing Systems*, 2019, 32.
- [32] NGUYEN A, YOSINSKI J, CLUNE J. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 427-436.
- [33] TRAMER F, CARLINI N, BRENDEN W, et al. On adaptive attacks to adversarial example defenses[J]. *Advances in Neural Information Processing Systems*, 2020, 33: 1633-1645.
- [34] GEIRHOS R, MEDINA TEMME C R, RAUBER J, et al. Generalisation in humans and deep neural networks[C]//Thirty-second Annual Conference on Neural Information Processing Systems 2018 (NeurIPS 2018). Curran, 2019: 7549-7561.
- [35] ZHENG S, SONG Y, LEUNG T, et al. Improving the robustness of deep neural networks via stability training[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 4480-4488.
- [36] HOSSEINI H, POOVENDRAN R. Semantic adversarial examples[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2018: 1614-1619.
- [37] HUBEL D H, WIESEL T N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex[J]. *The Journal of physiology*, 1962, 160(1): 106-154.
- [38] GOODALE M A, MILNER A D. Separate visual pathways for perception and action[J]. *Trends in Neurosciences*, 1992, 15(1): 20-25.
- [39] KATZNER S, WEIGELT S. Visual cortical networks: of mice and men[J]. *Current Opinion in Neurobiology*, 2013, 23(2): 202-206.
- [40] RAJALINGHAM R, DICARLO J J. Reversible inactivation of different millimeter-scale regions of primate IT results in different patterns of core object recognition deficits[J]. *Neuron*, 2019, 102(2): 493-505. e5.
- [41] KANWISHER N. Functional specificity in the human brain: a window into the functional architecture of the mind[J]. *Proceedings of the National Academy of Sciences*, 2010, 107(25): 11163-11170.
- [42] KONKLE T, OLIVA A. A real-world size organization of object responses in occipitotemporal cortex[J]. *Neuron*, 2012, 74(6): 1114-1124.
- [43] KRIEGESKORTE N, MUR M, RUFF D A, et al. Matching categorical object representations in inferior temporal cortex of man and monkey[J]. *Neuron*, 2008, 60(6): 1126-1141.
- [44] PROKLOVA D, KAISER D, PEELEN M V. Disentangling representations of object shape and object category in human visual cortex: The animate-inanimate distinction[J]. *Journal of Cognitive Neuroscience*, 2016, 28(5): 680-692.
- [45] FISHER R A. The use of multiple measurements in taxonomic problems[J]. *Annals of Eugenics*, 1936, 7(2): 179-188.
- [46] MCCULLOCH W S, PITTS W. A logical calculus of the ideas immanent in nervous activity[J]. *The Bulletin of Mathematical Biophysics*, 1943, 5(4): 115-133.
- [47] EICKENBERG M, GRAMFORT A, VAROQUAUX G, et al. Seeing it all: Convolutional network layers map the function of the human visual system[J]. *Neuro Image*, 2017, 152: 184-194.
- [48] HORIKAWA T, KAMITANI Y. Generic decoding of seen and imagined objects using hierarchical visual features[J]. *Nature Communications*, 2017, 8(1): 1-15.
- [49] ST-YVES G, NASELARIS T. The feature-weighted receptive field: an interpretable encoding model for complex feature spaces[J]. *Neuro Image*, 2018, 180: 188-202.
- [50] WEN H, SHI J, ZHANG Y, et al. Neural encoding and decoding with deep learning for dynamic natural vision[J]. *Cerebral Cortex*, 2018, 28(12): 4136-4160.
- [51] CADIEU C F, HONG H, YAMINS D L K, et al. Deep neural networks rival the representation of primate IT cortex for core visual object recognition[J]. *PLoS Comput Biol*, 2014, 10(12): e1003963.

- [52] BASHIVAN P, KAR K, DICARLO J J. Neural population control via deep image synthesis[J]. Science, 2019, 364(6439).
- [53] ULLMAN S, ASSIF L, FETAYA E, et al. Atoms of recognition in human and computer vision[J]. Proceedings of the National Academy of Sciences, 2016, 113(10): 2744-2749.
- [54] ELSAYED G, SHANKAR S, CHEUNG B, et al. Adversarial examples that fool both computer vision and time-limited humans[C]//Advances in Neural Information Processing Systems. 2018: 3910-3920.
- [55] ZHOU Z, FIRESTONE C. Humans can decipher adversarial images[J]. Nature Communications, 2019, 10(1).
- [56] SANTURKAR S, TSIPRAS D, TRAN B, et al. Image synthesis with a single (robust) classifier[J]. Advances in Neural Information Processing Systems, 2019, 32.
- [57] RITTER S, BARRETT D G T, SANTORO A, et al. Cognitive psychology for deep neural networks: A shape bias case study[C]//International Conference on Machine Learning. PMLR, 2017: 2940-2949.
- [58] HOSSEINI H, XIAO B, JAISWAL M, et al. Assessing shape bias property of convolutional neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2018: 1923-1931.
- [59] GEIRHOS R, RUBISCH P, MICHAELIS C, et al. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness [C]//International Conference on Learning Representations, ICLR 2019.

[作者简介]



林点（1996- ），男，福建福州人，上海交通大学硕士生，主要研究方向为人工智能安全。



潘理（1974- ），男，江苏苏州人，博士，上海交通大学教授、博士生导师，主要研究方向为网络大数据分析、云计算与大数据安全、网络安全管理。



易平（1969- ），男，河南洛阳人，博士，上海交通大学副研究员，主要研究方向为人工智能安全、无线网络安全、软件与系统安全。