

文章编号 1004-924X(2023)15-2287-08

# 模拟初级视觉皮层增强 CNN 神经网络结构的 稳定性

张丽娟<sup>1,2</sup>, 胡梦达<sup>2</sup>, 张紫薇<sup>2</sup>, 姜雨彤<sup>3</sup>, 李东明<sup>1\*</sup>

- (1. 无锡学院 物联网工程学院, 江苏 无锡 214105;  
2. 长春工业大学 计算机科学与工程学院, 吉林 长春 130012;  
3. 中国北方车辆研究所, 北京 100072)

**摘要:**针对卷积网络模型的稳定性能较差,对抗训练方法会使得网络结构过于复杂并占用大量运算资源的问题,提出了一种基于人体视觉神经系统生物特征的卷积神经网络模型改进方法(VVNet)。在卷积神经网络的基础上,融合人体视觉的结构特征,在不增加网络层数或保持准确率不变的情况下,提高神经网络面对噪声干扰的稳定性。在数据集 Cifar10 上对 3 种不同神经网络模型(VVNet, VOneNet 以及原网络模型)进行测试。实验结果表明,使用 VVNet 网络模型、VOneNet 网络模型和原始的网络模型 DenseNet121 对四类图像(噪声图像、模糊图像、遮挡图像和饱和曝光图像)的分类准确率进行对比,验证了提出的 VVNet 网络结构对不同类型图像的分类准确率几乎不变,在使用对抗样本情况下, VVNet 网络结构的图像分类准确率提高了约 10%。与深度学习网络相比,基于人体视觉系统结构的网络能够在保持准确率的同时有效地提高神经网络的稳定性,并具有可移植性。

**关键词:**计算机视觉;机器学习;图像识别;视觉皮层

中图分类号:TP394.1;TH691.9 文献标识码:A doi:10.37188/OPE.20233115.2287

## Simulating primary visual cortex to improve robustness of CNN neural network structures

ZHANG Lijuan<sup>1,2</sup>, HU Mengda<sup>2</sup>, ZHANG Ziwei<sup>2</sup>, JIANG Yutong<sup>3</sup>, LI Dongming<sup>1\*</sup>

- (1. School of Internet of Things Engineering, Wuxi University, Wuxi 214105, China;  
2. Collegel of Computer Science and Technology, Changchun University of Technology,  
Changchun 130012, China;  
3. China North Vehicle Research Institute, Beijing 100072, China)

\* Corresponding author, E-mail: LDM0214@163.com

**Abstract:** The robustness of convolutional neural network (CNN) models is usually improved by deepening the number of network layers to ensure the accuracy of the results. However, increasing the number of network layers will make the network more complex and occupy more space. This paper proposes an im-

收稿日期:2023-02-14;修订日期:2023-03-29.

基金项目:国家自然科学基金资助项目(No. 61801439);吉林省科技发展计划重点研发项目(No. 20210204050YY);吉林省生态环境厅科研项目(吉环科学第 2021-07 号);无锡学院引进人才科研启动专项经费资助项目(No. 2023r004, No. 2023r006)

proved CNN modeling method based on human visual features. Through the CNN, the structural features of human vision are fused to improve the robustness of the network against noise without increasing the number of layers or affecting the original accuracy of the model. The experimental results on the Cifar10 dataset show that the classification accuracy of the image inserted into the proposed VVNet is almost the same as that of the original network, and the classification accuracy is improved by approximately 10% in the case of image destruction. Compared with the original deep learning network, the network based on human visual system structure can effectively enhance the robustness of the network while maintaining the original accuracy.

**Key words:** computer vision; machine learning; image recognition; visual cortex

## 1 引言

自 20 世纪中叶机器学习理论被提出以来,随着相关研究的不断深入和计算机算力的不断提升,机器学习的应用越来越广泛,图像识别就是机器学习在计算机视觉领域上的主要应用之一。1998 年,LeCun<sup>[1]</sup>等提出了卷积神经网络(Convolutional Neural Network, CNN),并通过 LeNet-5 网络结构成功实现了手写数字的图像识别。自此,以 CNN 为基础的深度学习概念受到了相关领域大量研究人员的关注和研究,各种新型的网络结构模型和训练方法被提出。2012 年,Krizhevsky 等<sup>[2]</sup>提出了 AlexNet,在 ImageNet 图像识别竞赛中凭借优秀准确率夺得了分类比赛的冠军。2015 年,He<sup>[3]</sup>等提出的 ResNet 网络结构在 ImageNet 分类任务比赛中获得第一名,ResNet50 和 ResNet101 网络结构成为许多深度学习方法的基础,如 Inception 网络结构<sup>[4]</sup>以及 DenseNet 结构<sup>[5]</sup>等。随着神经网络准确率的不断提高,CNN 在搜索引擎、智能家居、生物医疗等多个领域都得到了广泛的应用,同时随着研究的深入,神经网络稳定性不足的缺陷也逐渐显现。

稳定性通常指神经网络模型接收到异常输入仍能输出正确结果的能力。2014 年,Szegedy 等<sup>[6]</sup>提出的对抗样本表明,所有的神经网络都有着稳定性不足的缺陷。神经网络是探索人脑神经元结构和功能并通过计算机对其模拟、相互连结而构建的,以此实现人工智能。目前,神经网络算法已经能够在一定程度上模拟人类的识别、记忆、运算等功能,但 Szegedy 等<sup>[6]</sup>发现,通过在图像上加入人类无法识别的噪声干扰,神

经网络模型会产生错误的判断。这种缺陷是神经网络自身局限性导致的,依赖于神经网络的图像识别是通过认知图像信息中必要的信息<sup>[7]</sup>,而必要信息的维数往往远小于输入数据的总维数,大量不被人类视觉所认知的冗余信息也会影响神经网络的准确率,即神经网络往往优先学习图像中的高频纹理特征,这种认知方式与人类认知信息的方式间存在分歧。

提高神经网络稳定性的主要方法有两种<sup>[8]</sup>:一是通过图像压缩<sup>[9]</sup>、去噪自编码器<sup>[10]</sup>等降噪手段,或是通过对比降噪前后样本<sup>[11]</sup>等方法直接筛查出对抗样本,以减轻网络受到对抗样本的攻击性,进而降低对抗样本对结果的干扰性;二是通过对抗训练<sup>[12-13]</sup>或生物启发模型等形式,直接增强模型自身的稳定性。前者往往防御能力有限,在会降低网络自身性能的同时也无法有效地抵抗高强度的干扰<sup>[14]</sup>。提高网络稳定性效果最好的方法是数据增强,但训练成本较高,还可能导致模型过拟合和网络层次过深、结构过于臃肿等问题。相比之下,人类视觉往往比神经网络模型具有更高的稳定性,更能分辨局部的图像特征<sup>[15]</sup>。

1962 年,Hubel 和 Wiesel<sup>[16]</sup>提出了视觉皮层的概念,对视觉系统中简单细胞和复杂细胞的反应功能做出描述以来,对于初级视觉皮层(V1)反应的建模始终是研究的主要领域之一。近年来,一些学者重新从类脑视觉的角度出发<sup>[17-21]</sup>,研究人类视觉模型对神经网络的改进。Reddy 等<sup>[22]</sup>在网络结构中引入了模拟人体视网膜和视觉皮层注视点的结构。Kim 等<sup>[23]</sup>建立了一种模拟视网膜的模型。

本文从模拟生物视觉神经结构的角度出发,以 Joel Dapello<sup>[24]</sup>等提出的 VOneNet 结构为基础,设计出 VVNet 混合 CNN 视觉模型。该模型延续了 VOneNet 模型基于经典的神经科学的线性-非线性-泊松(LNP)模型的结构,增加了非线性层的层数,并插入瓶颈层,进一步模仿生物视觉,以增强网络稳定性。在 Cifar10 数据集上进行训练, VVNet 结构在保持原网络模型性能的同时,有效地提高了模型在面对常见图像损坏和噪声时的稳定性。本文提出的网络可以作为输入层扩展模块接入其他 CNN,具有良好的可迁移性和可复现性。

## 2 VVNet 的构成及工作原理

### 2.1 对初级视觉皮层(V1)进行模拟建模

早期研究者们通过调整 Gabor 滤波器实现了简单细胞<sup>[25]</sup>和复杂细胞<sup>[26]</sup>的反应预测,并通过引入两个连续的线性-非线性结构来模拟 V1 的行为<sup>[27]</sup>。依据神经生理学的研究,人体的视觉系统中的初级视觉皮层中存在简单细胞和复杂细胞两种主要的感受器<sup>[16]</sup>。二者能分别对不同特定方向的线条或边缘产生反应,背外侧膝状核(LGN)中两种细胞互相合作,对接收到的图像信息中的轮廓、细节和运动等信息进行识别,并将得到的信息形成高级的视觉信息传递到脑部进行进一步的处理,在面部识别、运动感知中起到了重要的作用。

部分 CNN 在预测 V1 行为时胜过 Gabor 滤波器组<sup>[28]</sup>,这表明需要多个线性-非线性层来实现 V1 的复杂行为。但 CNN 的结构导致其会依赖于一些灵长类视觉系统不使用的一部分视觉特征作为判断基础。Joel 等<sup>[24]</sup>等通过使用基于偏最小平方回归(PLS)<sup>[29-30]</sup>的标准神经预测方法,评估每个模型诱发的单个 V1 神经元反应的解释程度,来对 AlexNet, DenseNet, VGG, ResNet 等神经网络进行训练分析,发现在白盒对抗性攻击下的模型准确率与 V1 解释方差呈现明显的正相关关系。以此为基础,本文提出的 VVNet,其最大特征是对生物视觉系统中视网膜到 V1 前端的部分做了进一步的模拟。每个 VVNet 由 VVBlock 和一个由对应基础 CNN 改编的后端网络组成, VVBlock 具有多个非线性层(见图 1),分别模拟

灵长类动物视觉神经系统 V1 区域中简单细胞和复杂细胞对图像信息的处理。最终,用 VVBlock 和一个特定的过渡层来替换原本 CNN 中的第一个模块(一般为一个卷积层、归一化层、非线性层和池化层的一个堆叠)。VVBlock 与被替换的模块相比虽空间图尺寸相同,但具有更多的通道数,而过渡层中一般有一个  $1 \times 1$  的卷积作为瓶颈,将更多的通道数压缩到原模型的深层。VV-Block 的构成遵循 LNP 模型,包括卷积层、非线性层和随机发生器三个连续的处理阶段,并且在通道都有一定数量的神经单元包含有简单细胞和复杂细胞两种不同的神经元类型。

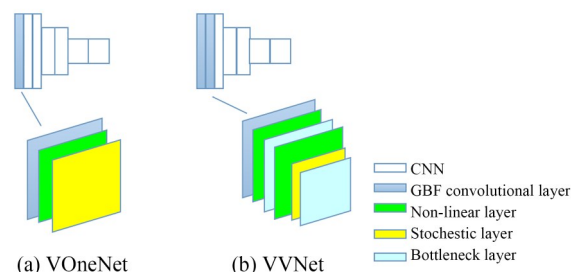


图 1 VVBlock 结构示意图

Fig. 1 Schematic diagram of VVBlock

### 2.2 生物约束的 Gabor 滤波器组的卷积层构建

VVBlock 中第一层是一个数学参数化的 GFB,其参数被调整为近似灵长类 V1 神经反应数据。它用多方向、多尺度及多空间频率的 Gabor 滤波器对 RGB 输入图像进行融合。将 GFB 的跨度设置为 4,产生一个  $56 \times 56$  的激活空间图。由于大多数 CNN 第一次卷积中的通道数量相对较少(在适应的架构中为 64 个),在 VVBlock 中将通道的数量设成较大的数值,以便 GFB 能覆盖更大的范围并更好地接近灵长类的 V1。将主要的 VOneNet 模型设置为 512 个通道,在简单和复杂细胞间平均分配。GFB 中的每个通道都对输入图像中的一个颜色通道进行卷积。

Gabor 函数由一个具有高斯包络的二维光栅组成,定义为:

$$G_{\theta, f, \varphi, n_x, n_y}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp[-0.5(x_{\text{rot}}^2/\sigma_x^2 + y_{\text{rot}}^2/\sigma_y^2)] \cos(2\pi f + \varphi), \quad (1)$$



$$\begin{aligned} x_{\text{rot}} &= x \cos \theta + y \sin \theta \\ y_{\text{rot}} &= -x \sin \theta + y \cos \theta, \end{aligned} \quad (2)$$

$$\begin{aligned} \sigma_x &= \frac{n_x}{f}, \\ \sigma_y &= \frac{n_y}{f}, \end{aligned} \quad (3)$$

其中:  $x_{\text{rot}}$  和  $y_{\text{rot}}$  分别为相对于光栅的正交和平行方向,  $\theta$  是光栅方向的角度,  $f$  是光栅的空间频率,  $\varphi$  是光栅相对于高斯包络的相位,  $\sigma_x$  和  $\sigma_y$  分别是正交和平行于光栅的高斯包络的标准偏差, 可以定义为光栅周期(频率的倒数)的倍数( $n_x$  和  $n_y$ )。

尽管 GFB 大大减少了定义 V1 神经元线性空间成分所需要的参数数量, 但它每个通道仍然有 5 个参数。幸运的是, 在神经生理学方面有大量文献对灵长类动物的 V1 反应特性进行了详细的描述, 可以用来约束这些参数。为了实例化一个带有 CV1 通道的 VVBlock, 这里根据经验上的约束分布对每个参数的 CV1 值进行采样。根据输入图像的分辨率, 限定空间频率( $f < 5.6$  c/d) 和周期数( $n > 0.1$ ) 的范围。

### 2.3 简单细胞和复杂细胞

简单细胞曾被认为是计算复杂细胞反应的一个中间步骤, 它们现在已形成了对 V2 的大部分下游投射<sup>[31]</sup>。基于 Hubel 和 Wiesel<sup>[16]</sup> 提出的视觉功能柱中具有多个独立的简单和复杂细胞层的设想, VVBlock 设计成具有两个不同的非线性层, 根据其对应的细胞类型应用于各自的通道中, 模拟简单细胞的整流线性变换, 以及复杂细胞的正交相位对的频谱功率, 即:

$$\begin{aligned} S_{\theta, f, \varphi, n_x, n_y}^{\text{nl}} &= \begin{cases} S_{\theta, f, \varphi, n_x, n_y}^{\text{l}}, & \text{if } S_{\theta, f, \varphi, n_x, n_y}^{\text{l}} > 0 \\ 0, & \text{otherwise} \end{cases}, \quad (4) \\ S_{\theta, f, \varphi, n_x, n_y}^{\text{nl}} &= \frac{1}{\sqrt{2}} \sqrt{(C_{\theta, f, \varphi, n_x, n_y}^{\text{l}})^2 + (C_{\theta, f, \varphi + \pi/2, n_x, n_y}^{\text{l}})^2}, \end{aligned} \quad (5)$$

其中:  $S^{\text{l}}$  和  $S^{\text{nl}}$  分别为简单神经元的线性和非线性响应,  $C^{\text{l}}$  和  $C^{\text{nl}}$  对于复杂神经元来说也是相同的。

### 2.4 随机性层和瓶颈层

神经元反应的一个决定性属性是它们的随机性对一个神经元进行重复测量, 以应对名义上相同的视觉输入, 结果是不同的尖峰序列。对于清醒的猴子, 平均尖峰计数(多次重复的平均值)

取决于呈现的图像, 在同一组实验中, 尖峰序列近似呈泊松分布, 即尖峰计数方差基本等于平均值<sup>[32]</sup>。为了逼近神经元反应, 本文向 VVBlock 的每个单元添加独立的高斯噪声, 其方差等于激活度。最终, 使用现有的经验分布来约束 GFB 模型, 生成一个近似于灵长类 V1 的神经元空间。同时, 在每个非线性层后加入一个  $1 \times 1$  的卷积作为瓶颈层, 模拟视觉系统中对输出维度的约束<sup>[33]</sup>, 并将更多的通道数压缩到原模型的深层。

## 3 消融实验与结果

实验环境如下: NVIDIA A100 SXM4 40 GB 显卡的运行内存为 80 GB, 采用 Pytorch1.13.1 深度学习框架和 Python3.8 的服务器。本文使用的数据集为 CIFAR-10(图 2), 它由 Alex Krizhevsky, Vinod Nair 和 Geoffrey Hinton 收集完成。CIFAR-10 数据集包含 60 000 张  $32 \times 32$  的彩色图像, 分为 10 类, 每类 6 000 张图片, 其中 50 000 张作为训练图像, 10 000 张作为测试图像。

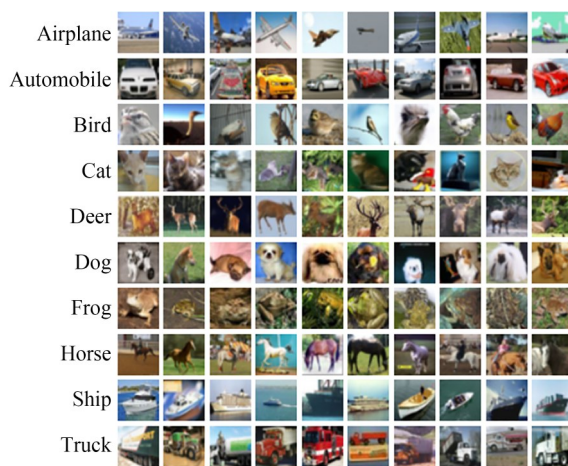


图 2 CIFAR-10 数据集图像示例

Fig. 2 Example images of CIFAR-10 dataset

为验证本文提出网络模型的性能, 实验选择在 CIFAR-10 数据集上使用常用图像识别算法与本文提出的算法在同一数据集上进行训练。为验证本文提出网络结构的稳定性, 使用 imagecorruptions 对图像进行破坏(见图 3), 模拟噪声干扰的情况, 并进行对比实验。

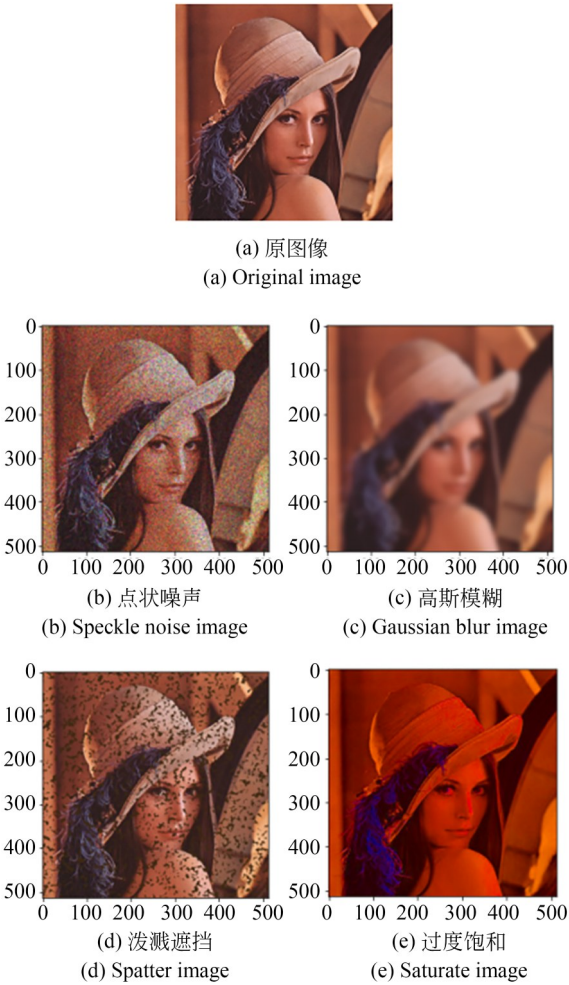


图3 人为加入噪声的图像

Fig. 3 Images with artificial insertion of noise

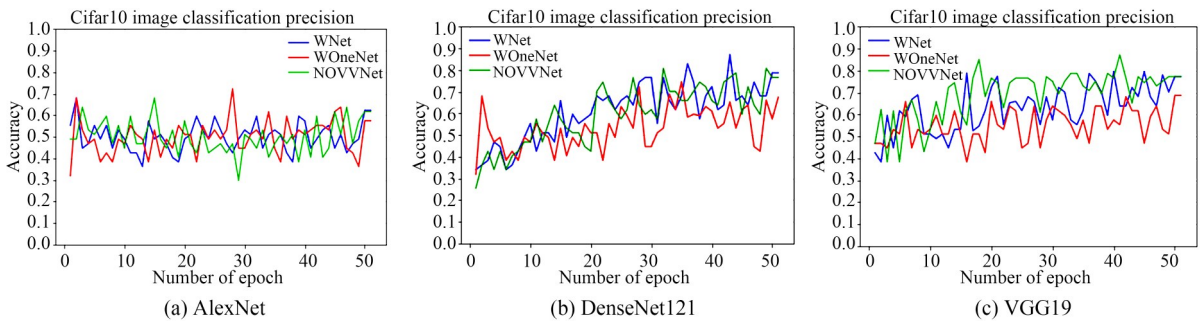


图4 在Cifar10数据集上对3种常见网络架构的原网络和VVNet重构后网络的准确率对比

Fig. 4 Accuracy comparison of original network and reconstructed network of VVNet on dataset Cifar10 for three common networks

3.2 模型稳定性测试

为了评估 VVNet 在噪声干扰下的稳定性,本文对数据集 Cifar10 的图像进行破坏处理,形成点状噪声、模糊、泼溅遮挡和饱和曝光 4 类图像,

3.1 VVNet 的适用性和可移植性测试

为验证 VVNet 网络的可行性,本文以 AlexNet<sup>[2]</sup>,DenseNet121<sup>[5]</sup>和 VGG19<sup>[34]</sup>3 种基本网络架构为基础来构建 VVNet,如图 4 所示。选择标准 Top1 图像分类准确率作为评价指标,通过对比 3 种网络在使用 VVNet,VOneNet 和原本网络结构方法下训练 50 轮的分类准确率,来检验插入 VVBlock 模块对算法准确率的影响。

从图 4 可以看出,与原网络相比,移植了 VVNet 结构的神经网络的图像分类识别准确率几乎没有变化,这证明了该方法对不同 CNN 网络的适应性和通用性。相比之下,套用 VOneNet 结构的网络准确率会下降 5%~10%,VVNet 结构对于不同网络的适用性比 VOneNet 更好,准确率变化波动小,能够在不对原网络产生负面影响的情况下进行优化。

表 1 三种网络不同架构在数据集 Cifar10 上的准确率

Tab.1 Accuracy for three network in different architectures on dataset Cifar10

| Network  | VVNet | VOneNet | None  | Epoch |
|----------|-------|---------|-------|-------|
| AlexNet  | 0.623 | 0.574   | 0.617 | 50    |
| DenseNet | 0.787 | 0.674   | 0.765 |       |
| VGG      | 0.773 | 0.687   | 0.772 |       |

使用的网络模型为在普通的训练集上训练完成的一般模型,在噪音破坏过的测试集上进行分类识别,通过对比准确率验证 VVNet 结构对神经网络模型稳定性的影响。

从图5可以看出,不论哪种噪声,随着破坏程度的加重,网络模型的准确率都有了明显的下降,与原网络准确率骤降的情况相比,不论是具有VVNet结构还是具有VOneNet结构网络都能够保持一个更平缓的下降趋势。从实验结果可

以看出,VVNet相比VOneNet,曲线更平滑,在破坏程度最高等级的情况下,分类准确率也能提高10%左右。特别是在饱和度破坏的情况下,本文提出的方法能够有效地减少图像破坏对网络模型准确率的影响。

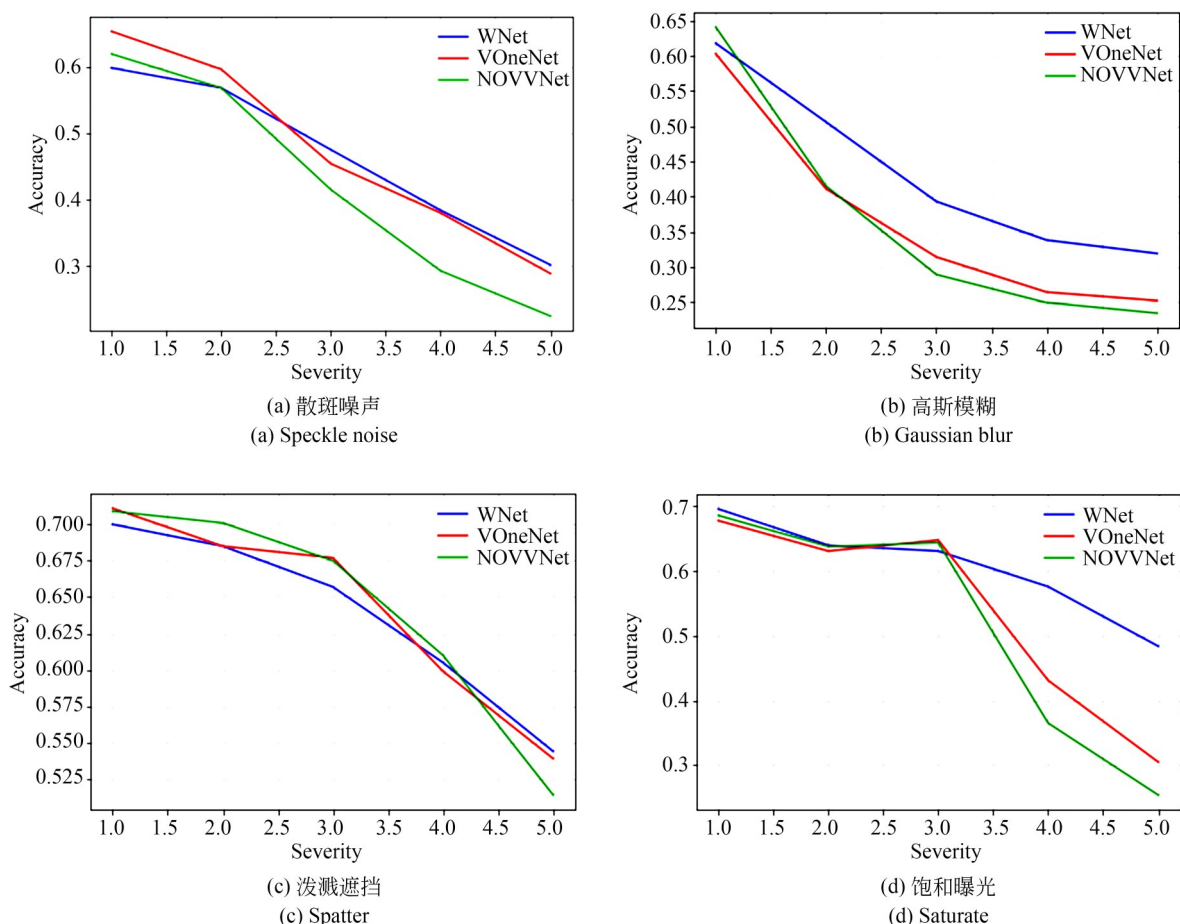


图5 使用VVNet,VOneNet和原始的网络模型DenseNet121对4类噪声图像的分类准确率对比

Fig. 5 Comparison of image classification accuracy for four types of noise images using VVNet network, VOneNet network, and original network model

在CIFAR-10数据集上进行多组对抗性测试,结果表明,使用VVNet结构的神经网络模型在不进行对抗训练的情况下确实有效地提高网络的稳定性。

## 4 结 论

本文提出了一种基于人体视觉神经系统生物特征的CNN结构改进方法,通过将模拟初级视觉皮层结构的VVBlock作为CNN前端输入模

块来优化原本的网络性能,能够在不增加网络层数并保持原准确率的情况下提高深度学习算法的稳定性。实验结果表明,与原网络模型相比,使用VVNet方法的网络模型的准确率提升了5%~10%,对于增强神经网络的稳定性具有积极影响。同时,VVNet具有良好的可迁移性,通过实验验证,与原网络VOneNet相比,VVnet网络在不同的CNN结构(AlexNet<sup>[2]</sup>、DenseNet121<sup>[5]</sup>和VGG19<sup>[34]</sup>)间可以顺利移植,并且不会对原网络模型的准确率造成较大的影响。



但 VVNet 作为一种生物启发模型的探索仍有许多不足,优化效果还不如对抗训练,今后的研究

工作会尝试进一步优化结构和实验方法,以提高图像分类的稳定性。

#### 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 E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [3] HE K M, ZHANG X Y, REN S Q, *et al.* Deep residual learning for image recognition [C]. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 27-30, 2016, Las Vegas, NV, USA. IEEE, 2016: 770-778.
- [4] SZEGEDY C, VANHOUCHE V, IOFFE S, *et al.* Rethinking the inception architecture for computer vision[C]. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 27-30, 2016, Las Vegas, NV, USA. IEEE, 2016: 2818-2826.
- [5] HUANG G, LIU Z, VAN DER MAATEN L, *et al.* Densely connected convolutional networks[C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. July 21-26, 2017, Honolulu, HI, USA. IEEE, 2017: 2261-2269.
- [6] SZEGEDY C, ZAREMBA W, SUTSKEVER I, *et al.* Intriguing properties of neural networks[EB/OL]. 2013: *arXiv*: 1312.6199. <https://arxiv.org/abs/1312.6199>.
- [7] ILYAS A, SANTURKAR S, TSIPRAS D, *et al.* Adversarial examples are not bugs, they are features [EB/OL]. 2019: *arXiv*: 1905.02175. <https://arxiv.org/abs/1905.02175>.
- [8] 林点, 潘理, 易平. 面向图像识别的卷积神经网络稳定性研究进展[J]. *网络与信息安全学报*, 2022, 8(3):111-122.  
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. (in Chinese)
- [9] DZIUGAITE G K, GHAMRANI Z, ROY D M. A study of the effect of JPG compression on adversarial images [EB/OL]. 2016: *arXiv*: 1608.00853. <https://arxiv.org/abs/1608.00853>.
- [10] 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*. July 5-9, 2008, Helsinki, Finland. New York: ACM, 2008: 1096-1103.
- [11] XU W, EVANS D, QI Y. Feature squeezing: detecting adversarial examples in deep neural networks [EB/OL]. 2017: *arXiv*: 1704.01155. <https://arxiv.org/abs/1704.01155>.
- [12] TSIPRAS D, SANTURKAR S, ENGSTROM L, *et al.* Robustness may be at odds with accuracy [EB/OL]. 2018: *arXiv*: 1805.12152. <https://arxiv.org/abs/1805.12152>.
- [13] XIE C H, TAN M X, GONG B Q, *et al.* Adversarial examples improve image recognition [C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 13-19, 2020, Seattle, WA, USA. IEEE, 2020: 816-825.
- [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*. 3 November 2017, Dallas, Texas, USA. New York: ACM, 2017: 3-14.
- [15] ULLMAN S, ASSIF L, FETAYA E, *et al.* Atoms of recognition in human and computer vision [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2016, 113 (10): 2744-2749.
- [16] 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.
- [17] ZADOR A M. A critique of pure learning and what artificial neural networks can learn from animal brains [J]. *Nature Communications*, 2019, 10: 3770.
- [18] MARBLESTONE A H, WAYNE G, KORDING K P. Toward an integration of deep learning and neuroscience [J]. *Frontiers in Computational*

- Neuroscience*, 2016, 10: 94.
- [19] NAYEBI A, GANGULI S. Biologically inspired protection of deep networks from adversarial attacks [EB/OL]. 2017: *arXiv*: 1703.09202. <https://arxiv.org/abs/1703.09202>.
- [20] LINDSAY G W, MILLER K D. How biological attention mechanisms improve task performance in a large-scale visual system model[J]. *eLife*, 2018, 7: 38105.
- [21] HASANI H, SOLEYMANI M, AGHAJAN H. Surround modulation: a bio-inspired connectivity structure for convolutional neural networks [J]. *Advances in Neural Information Processing Systems*, 2019.
- [22] REDDY M V, BANBURSKI A, PANT N, *et al.* Biologically inspired mechanisms for adversarial robustness [EB/OL]. 2020: *arXiv*: 2006.16427. <https://arxiv.org/abs/2006.16427>.
- [23] KIM E, REGO J, WATKINS Y, *et al.* Modeling biological immunity to adversarial examples [C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 13-19, 2020, Seattle, WA, USA. IEEE, 2020: 4665-4674.
- [24] DAPELLO J, MARQUES T, SCHRIMPF M, *et al.* Simulating a primary visual cortex at the front of CNNs improves robustness to image perturbations[J]. *Advances in Neural Information Processing Systems*, 2020, 33: 13073-13087.
- [25] JONES J P, PALMER L A. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex[J]. *Journal of Neurophysiology*, 1987, 58(6): 1233-1258.
- [26] ADELSON E H, BERGEN J R. Spatiotemporal energy models for the perception of motion [J]. *Journal of the Optical Society of America A*, 1985, 2(2): 284.
- [27] VINTCH B, MOVSHON J A, SIMONCELLI E P. A convolutional subunit model for neuronal responses in macaque V1[J]. *The Journal of Neuroscience*, 2015, 35(44): 14829-14841.
- [28] CADENA S A, DENFIELD G H, WALKER E Y, *et al.* Deep convolutional models improve predictions of macaque V1 responses to natural images [J]. *PLoS Computational Biology*, 2019, 15(4): e1006897.
- [29] SCHRIMPF M, KUBILIUS J, HONG H, *et al.* Brain-Score: which Artificial Neural Network for Object Recognition is most Brain-Like?[J]. *bioRxiv*, 2018, DOI: 10.1101/407007.
- [30] KUBILIUS J, SCHRIMPF M, HONG H, *et al.* Brain-like object recognition with high-performing shallow recurrent ANNs[EB/OL]. 2019: *arXiv*: 1909.06161. <https://arxiv.org/abs/1909.06161>.
- [31] EL-SHAMAYLEH Y, KUMBHANI R D, DHURUV N T, *et al.* Visual response properties of V1 neurons projecting to V2 in macaque[J]. *The Journal of Neuroscience*, 2013, 33(42): 16594-16605.
- [32] SOFTKY W R, KOCH C. The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs[J]. *The Journal of Neuroscience*, 1993, 13(1): 334-350.
- [33] LINDSEY J, OCKO S A, GANGULI S, *et al.* A unified theory of early visual representations from retina to cortex through anatomically constrained deep CNNs [EB/OL]. 2019: *arXiv*: 1901.00945. <https://arxiv.org/abs/1901.00945>.
- [34] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. 2014: *arXiv*: 1409.1556. <https://arxiv.org/abs/1409.1556>.

#### 作者简介:



张丽娟(1978—),女,吉林梅河口人,博士,教授,2001年于吉林师范大学获得学士学位,2004年、2015年于长春理工大学分别获得硕士和博士学位,主要从事计算机视觉及光学图像处理等方面的研究。E-mail: zhanglijuan@ccut.edu.cn