

基于光电混合神经网络的单像素快速运动物体分类 (特邀)

郑淑君¹, 姚曼虹², 王晟平¹, 张子邦^{1,3}, 彭军政^{1,3}, 钟金钢^{1,3*}

- (1. 暨南大学光电工程系, 广东 广州 510632;
2. 广东技术师范大学光电工程学院, 广东 广州 510665;
3. 广东省光纤传感与通信技术重点实验室, 广东 广州 510632)

摘要: 对快速运动物体进行持续分类具有重要的应用前景。受限于有限的数据传输带宽和存储空间, 目前基于场景图像的物体分类技术难以实现对运动物体的持续分类。受到单像素成像在时间上累积获取信息这一方式的启发, 结合深度学习, 提出了一种基于光电混合神经网络的单像素快速运动物体分类方法。该方法不需要获取目标物体的图像, 利用对光场的空间调制和单像素测量, 直接获取用于分类的特征信息, 从而避免了在持续分类过程中基于图像分类方法产生的海量图像数据。单像素测量过程作为神经网络的一部分, 将光计算与电子计算无缝衔接起来, 构建了一个光电混合神经网络用于对物体的分类。通过对快速旋转圆盘上的手写数字进行持续分类实验测试, 证明了提出的方法在分类快速运动的手写数字方面的能力, 超过了人眼视觉。

关键词: 物体分类; 深度学习; 单像素测量; 光电混合神经网络

中图分类号: TB133 **文献标志码:** A **DOI:** 10.3788/IRLA20210856

Single-pixel fast-moving object classification based on optical-electronical hybrid neural network (*Invited*)

Zheng Shujun¹, Yao Manhong², Wang Shengping¹, Zhang Zibang^{1,3}, Peng Junzheng^{1,3}, Zhong Jingang^{1,3*}

- (1. Department of Optoelectronic Engineering, Jinan University, Guangzhou 510632, China;
2. School of Optoelectronic Engineering, Guangdong Polytechnic Normal University, Guangzhou 510665, China;
3. Guangdong Provincial Key Laboratory of Optical Fiber Sensing and Communications, Guangzhou 510632, China)

Abstract: Successive classification of fast-moving objects is significant in various fields. However, due to the limited data transmission bandwidth and data storage space, it is challenging to perform fast-moving object classification based on scene photography for a long duration. Inspired by single-pixel imaging and combined with deep learning, a single-pixel fast-moving object classification method based on optical-electronic hybrid neural network was proposed. The proposed method had no need to acquire the images of objects, but obtained the feature information for classification directly by using spatial light modulating and single-pixel detecting. Thus, the massive image data produced by the image-based classification for a long duration was avoided. As part of the neural network, the single-pixel detecting connected optical computing and electronic computing seamlessly, an optical-electronic hybrid neural network for object classification was constructed. The proposed method in classifying fast-moving handwritten digits on a rotating disk was experimentally demonstrated, which

收稿日期: 2021-11-16; 修订日期: 2021-12-07

基金项目: 国家自然科学基金 (61905098, 61875074); 中央高校基础研究基金 (11618307); 广东省基础与应用基础研究基金 (2020 A1515110392); 广东技术师范大学校级科研项目人才专项 (2021 SDKYA049)

作者简介: 郑淑君, 女, 硕士生, 主要从事计算光学成像方面的研究。

导师(通讯作者)简介: 钟金钢, 男, 教授, 博士生导师, 博士, 主要从事计算光学成像方面的研究。

passed through the field of view successively. The experiment confirmed that the classification ability of the proposed method had exceeded human vision.

Key words: object classification; deep learning; single-pixel measurement; optical-electronical hybrid neural network

0 引言

对场景中的物体进行分类,一直被认为是机器视觉最基本的问题之一^[1],也是视觉系统理解场景内容的基础。经过多年的发展,特别是近年来借助人工智能技术,机器视觉的分类能力已有很大的进步,但是仍然面临许多挑战,例如对快速运动物体的持续分类。而对快速运动物体的持续分类在自动驾驶、军事安全等领域具有重要的应用前景。

目前,机器视觉物体分类大多以获取图像为基础,利用图像对目标物体进行分类。先利用相机拍摄目标物体的图像,接着从图像中提取用于分类的特征信息,再将提取的特征信息和已知的物体特征信息进行比较,得出分类结果^[2]。显然,图像的质量将影响分类的准确度。对快速运动物体的持续分类来说,为了尽量降低图像的运动模糊和持续获取进入场景的物体图像,需要利用高速相机对场景进行持续的拍摄。然而,高速相机持续拍摄会产生海量图像数据,将给数据的传输带宽和存储空间带来极大的压力。因此,目前基于图像的物体分类技术难以实现对快速运动物体的持续分类。但是,另一方面,物体的分类并不需要图像中含有的所有信息,图像数据对于物体分类任务来说有较大冗余。因此,大幅度降低用于物体分类的数据量可能是解决这一难题的有效途径。单像素成像技术作为一种典型的计算成像技术,由于其独特的成像机理,在一些特殊领域相对传统成像具有优势,受到越来越多的关注^[3-8]。单像素成像技术利用空间光调制器产生的时变结构光场,将目标物体的空间信息转换为二维时间信息。不同于传统基于透镜成像的一次同时获取所有信息的方式,其获取信息的过程是在时间上不断累积的过程。单像素成像的信息获取方式从成像的角度导致了成像速度的降低;但是可以按需获取信息,从而降低获取信息的带宽和节约信息存储空间。就物体分类来说,这种按需获取信息的方式为获取只用于分类的特征信息和大幅降低数据量提供了可能。受此启发,科研人员开始对单像

素物体分类技术进行了探索研究,目前大多是针对静态物体的分类,利用特定图案的结构光获取目标物体的特征信息并结合深度学习或机器学习进行物体分类,使用的结构光图案包括哈达玛基底图案^[9]、离散余弦变换基底图案^[10]、随机图案^[11]、经过优化的图案^[12]等。

但是,物体分类的许多应用是针对运动物体,并且要求根据分类结果进行迅速响应,例如自动驾驶中的目标识别和军事安全中对飞行器的识别。这对物体分类技术提出了更加苛刻的要求,不仅要求获取信息的速度要快,而且处理信息的计算速度也要足够快。

近年来,光计算在人工智能领域相对电子计算展现出能耗低、计算速度快的优势,已成为研究热点^[13-16]。目前受限于工艺水平,光计算难以实现高效的光缓存和逻辑运算,在相同计算精度下不能达到微电子处理器的集成度。因此,用光计算全面取代电子计算为时尚早,仍处于概念研究阶段^[17]。

在单像素成像中,使用单像素探测器获取结构光照明目标物体后的光强值或目标物体的图像经空间调制后的光强值,实质上是以光的形式实现了物体空间分布函数与空间调制光场函数的内积运算。基于此,笔者课题组提出了一种基于深度学习的结构光照明单像素运动物体分类技术^[18]。利用空间光调制器生成的结构光对物体照明,单像素探测器对结构光照明后的物光强进行积分测量,该过程实现了神经网络的第一层卷积运算,完成了光探测和电子计算神经网络的无缝对接,因此,该技术构建的神经网络本质上属于光电混合的神经网络。在此基础上,为了进一步证明基于光电混合神经网络的单像素快速运动物体分类技术的可行性,文中构建了一种新的光电混合神经网络,即全卷积光电神经网络,全部的网络层由卷积层组成,以证明光电混合神经网络的概念也可以在不同的网络架构下实现;并改用适用于远距离物体分类的结构化探测光场调制方式进行运动物体分类,在一定程度上克服了应用场景的局限。通过对快速旋转盘上的手写数字进行持续分类实验测试,证明了

文中提出的技术在分类快速运动手写数字方面的能力超过了人眼视觉。

1 方法与系统架构

1.1 基于单像素探测的光电混合神经网络的设计

图 1 所示为结构化探测方式的单像素成像光路, 目标物体 (Target object) 被成像到空间光调制器 (Spatial light modulator, SLM) 表面, SLM 产生调制图案对物体的像进行调制, 调制后的出射光由单像素探测器 (Single-pixel detector, SPD) 收集测量。该过程在数学上可以表示为:

$$D = \alpha \int \int P(x,y) \cdot O(x,y) dx dy + \varepsilon \quad (1)$$

式中: $P(x,y)$ 表示 SLM 产生的调制图案; $O(x,y)$ 表示目标物体的空间分布函数; α 表示单像素探测器的光电响应系数; ε 表示背景杂散光强; D 表示单像素探测器的输出电压。在单像素成像过程中, 每产生一个调制图案, 对目标物体的像进行一次调制, 单像素探测器实施一次测量只能获得一部分图像信息, 理论上需要实施和图像像素数一样多的测量数, 才能利用所有的测量值重建一幅信息完全的图像。

而对于物体分类任务来说, 只需要少量的物体特征信息。为了避免获取过多的冗余信息, 可以采取类似单像素成像的信息获取方式, 实施少量的单像素测量, 获取只用于分类的特征信息。

卷积神经网络 (Convolutional Neural Network,

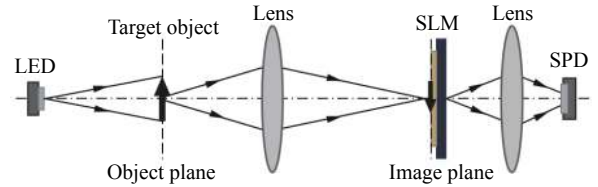


图 1 结构化探测单像素成像光路

Fig.1 Optical configuration of structured detected single-pixel imaging

CNN) 自 20 世纪七八十年代提出以来^[19], 被广泛应用于物体分类领域。对于 CNN 网络, 当卷积核的尺寸与输入图像的尺寸一致时, 卷积运算就是内积运算。从数学描述来说, 内积运算可以表示为:

$$E = \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} f(x,y) \cdot K(x,y) \quad (2)$$

式中: $f(x,y)$ 表示图像; $K(x,y)$ 表示卷积核; E 为输出值。令 SLM 显示神经网络的卷积核图案 $K(x,y)$, 则单像素探测值 D 与卷积层的输出 E 为线性关系, 单像素探测值在数学上是物体的像与卷积核的内积。可以发现, 公式 (2) 的内积运算与单像素成像中探测器获取光强值的物理过程完全一致, 因此可以通过单像素探测器获取光强值的方式实现“光”内积。

为与单像素探测实现“光”内积的形式对应, 设计了一种用于目标物体分类的全卷积神经网络, 如图 2 所示, 该网络结构全部由卷积层组成。例如: 输入图像的尺寸为 28×28 pixel, 经过与 N 个尺寸为 28×28 的卷积核卷积, 得到 $1 \times 1 \times N$ 的一维特征图。然

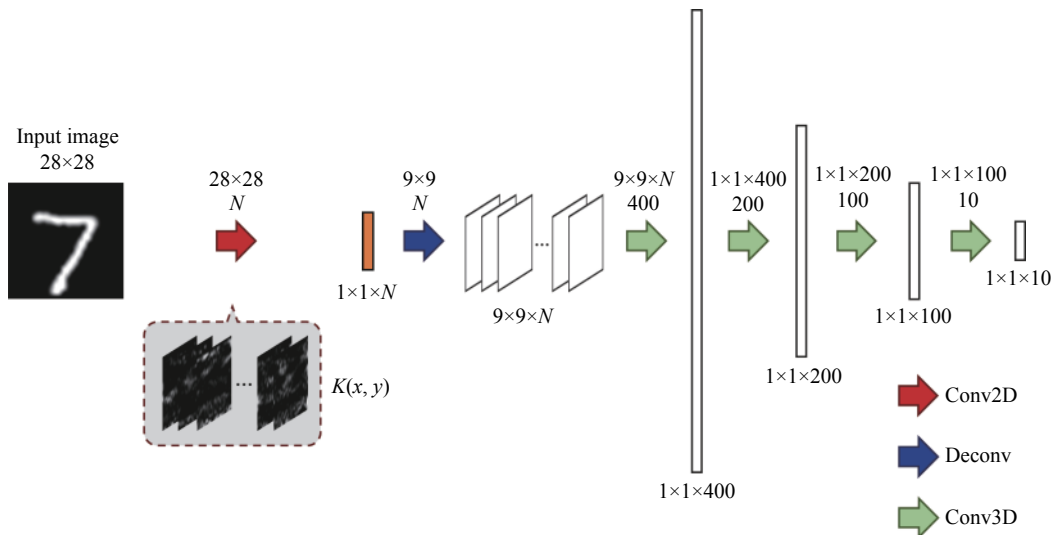


图 2 全卷积神经网络结构

Fig.2 Framework of the fully convolutional neural network

后经过 N 个尺寸为 9×9 的反卷积核, 得到 $9 \times 9 \times N$ 的特征图。再经过 400 个尺寸为 $9 \times 9 \times 15$ 的卷积核, 得到 $1 \times 1 \times 400$ 的一维特征图。此后依次经过 200、100 和 10 个 1×1 卷积核, 最终输出的 10 个神经元代表 10 个类别的概率。网络中所有卷积核参数使用截尾高斯分布 (Truncated Normal) 初始化。除最后一层的激活函数为 Softmax 外, 其余层均使用 Relu 激活函数。训练过程中选择交叉熵损失函数和 ADAM (Adaptive Moment Estimation) 算法来优化参数。

设计的基于单像素探测的光电混合神经网络如图 3 所示, 就是将该全卷积神经网络的第一层卷积运算, 用单像素测量以“光”内积运算代替。这样, 单像素测量过程作为神经网络的一部分, 将光计算与电子计算无缝衔接起来。通过构建的光电混合神经网络学习得到的第一层 N 个卷积核作为空间光调制器的 N 个调制图案, 分别对目标物体像进行调制, 单像素探测器获取的 N 个测量值送入后续的电子神经网络部分, 最后完成目标物体的分类。

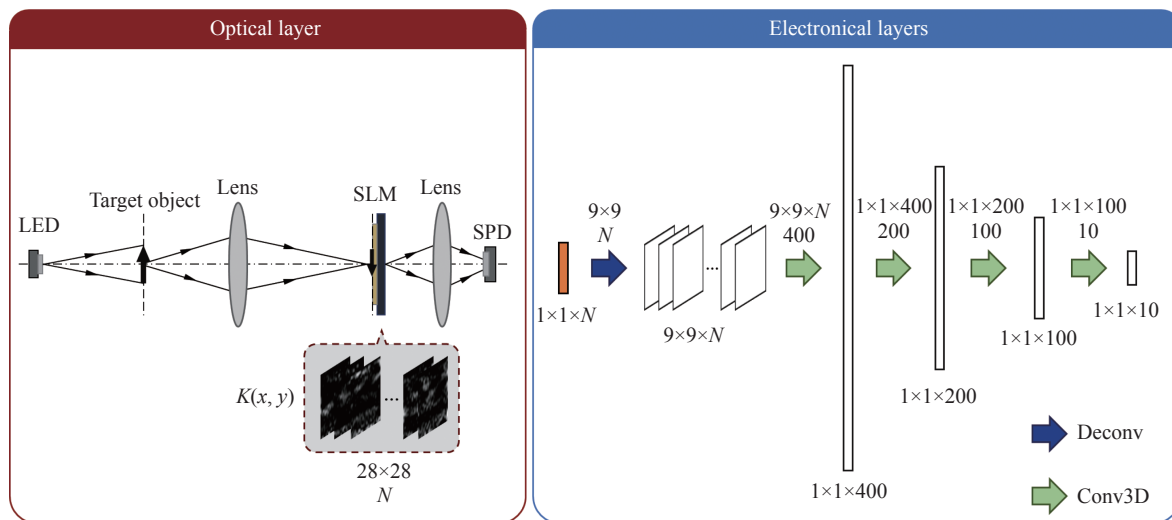


图 3 光电混合神经网络

Fig.3 Optical-electronical hybrid neural network

1.2 光电混合神经网络的训练

以 MNIST 数据集^[20]中的手写数字分类作为示例, 展示提出的光电混合全卷积网络在分类任务上的性能。该手写数据集包含 0~9 十个类别, 有 60 000 张训练集图片和 10 000 张测试集图片, 每张图片的尺寸为 28×28 pixel。为了测试提出的光电混合神经网络对运动物体的分类能力, 将这些手写数字放置在一个快速旋转的圆盘上。为了增强网络的健壮性, 对 MNIST 手写数据集中的图像进行 $-4^\circ \sim 4^\circ$ 的随机旋转和 $-12 \sim 12$ pixel 的随机水平平移, 如图 4 所示, 这些经过随机旋转和平移的图像将用来训练和测试网络。

通常对提出的光电混合神经网络的训练需要将手写数字图片作为被识别的物体, 如图 3 所示, 将用于训练的数字图片成像到 SLM 上, 单像素探测器测得光强值, 将获得的单像素测量值送入后续的电子神

经网络部分进行训练。这种训练方式需要耗费较大的时间和人力, 这也是许多人工智能应用落地的瓶颈之一。为了降低训练难度, 利用单像素测量过程的内积运算模型, 可以直接用数据集中的手写数字代替图 3 中手写数字的像, 用手写数字和卷积核的内积模拟单像素测量值, 也就是利用和光电混合神经网络对应的电子神经网络进行训练 (图 2)。



图 4 部分训练集图片及随机旋转和水平平移后的图片

Fig.4 Example of the original training images and corresponding images with random rotation and lateral shift

使用 60 000 张经过随机旋转和平移的训练集图片来训练网络,网络训练完成后,用 10 000 张经过随机旋转和平移的测试集图片对全卷积神经网络进行性能测试。第一层卷积层卷积核数量 N 为 15 时,测试集准确率达 93.70%。测试集分类结果的混淆矩阵如图 5 所示,大部分预测标签沿对角线分布并与真实标签匹配,表明提出的全卷积神经网络能对大多数测试数字进行正确分类,具有较好的分类能力。

由于应用训练好的光电混合神经网络时,实验中所用的调制图案来自于网络的第一层卷积核,训练结束后需提取第一层卷积核。以第一层卷积核数量 N 为 15 的网络为例,将这 15 个尺寸为 28×28 的二维卷积核提取出来,如图 6 所示。这些卷积核图案作为学习到的调制图案去调制物体的像,获取物体用于分类的特征信息。

Predicted labels	0	1	2	3	4	5	6	7	8	9
0	938	1	8	7	0	9	5	2	3	4
1	4	1115	4	1	3	2	3	9	17	8
2	4	3	980	14	3	0	2	15	3	1
3	4	1	13	911	0	25	1	11	24	7
4	0	3	6	3	958	4	7	6	13	40
5	3	2	1	37	0	809	10	2	9	15
6	12	1	6	1	5	11	924	1	10	4
7	4	0	4	11	2	8	1	947	5	11
8	9	8	6	17	1	12	5	2	881	12
9	2	1	4	8	10	12	0	33	9	907
	0	1	2	3	4	5	6	7	8	9

图 5 手写数字测试集分类结果混淆矩阵 (15 个卷积核)

Fig.5 Confusion matrix of the classification results on handwritten digit test set (15 kernels)

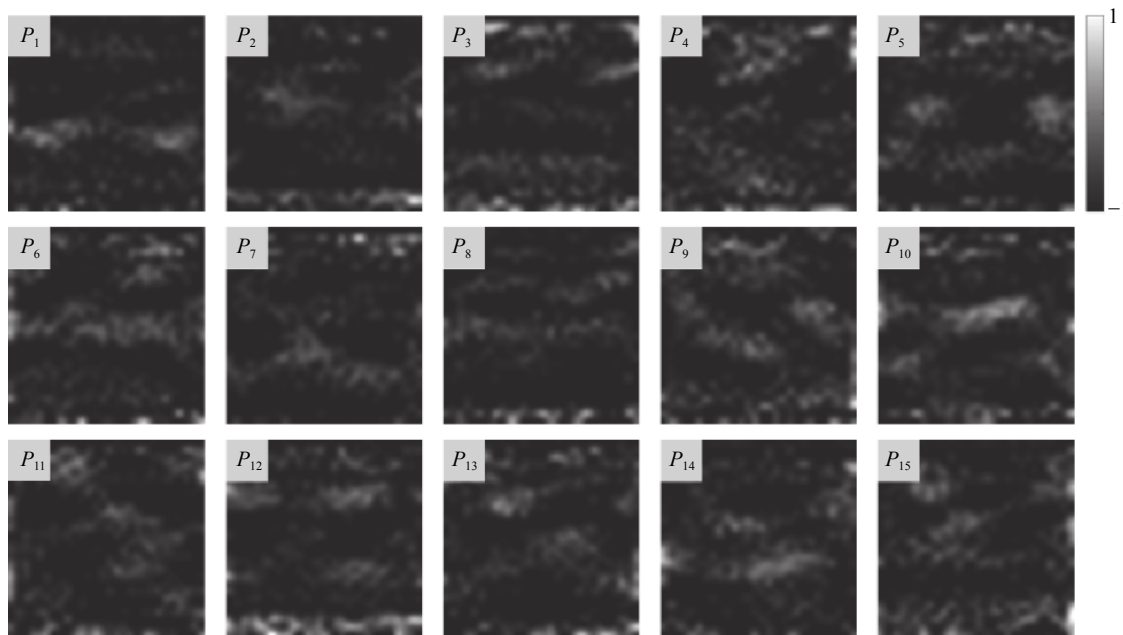


图 6 全卷积神经网络第一层二维卷积核图案

Fig.6 2D convolutional kernel images of the first layer in the fully convolutional neural network

为了考察调制图案的数量对光电混合神经网络分类性能的影响,为第一层卷积层设置不同数量的卷积核进行训练。不同卷积核数量对应的测试集分类准确率如图 7 所示。由图中曲线可知,随着卷积核数量的增加,分类准确率总体呈上升趋势。这是因为卷积核数量越多,网络提取的特征信息也越多。当卷积核数量达到 10 个时,准确率超过 90%;当卷积核数量

超过 19 个时,分类准确率稳定在 95% 左右。因此,根据准确率与效率之间的权衡可以选择合适的卷积核数量。

文中的代码在 TensorFlow 2.1.0 框架上基于 Python 3.7.6 实现。网络一共训练了 50 次,大约花费 5 min。网络的训练与测试都在一台搭载了 AMD Ryzen 7 1700 X CPU、英伟达 RTX 2080 Ti GPU 及

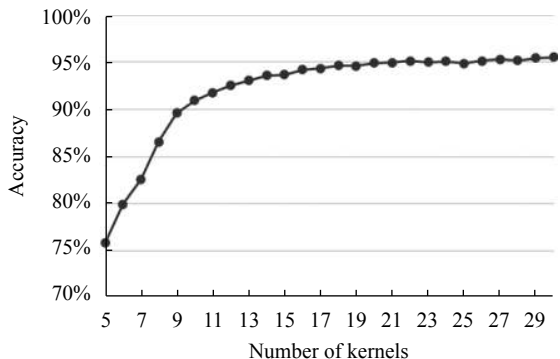


图 7 不同卷积核数量网络在 MNIST 测试集的分类准确率

Fig.7 MNIST test set classification accuracy of networks with different number of convolutional kernels

32 GB RAM 的电脑上运行。

2 实验及结果

对快速旋转盘上手写数字分类的实验装置如图 8 所示, 采用单像素结构化探测方案。用一个 10 W 的白光 LED 光源照明转盘上的一个手写数字, 经过透镜 1 成像到数字微镜阵列 DMD(VIALUX V-7001, 22 727 Hz) 表面。数字微镜阵列上依次显示不同的卷积核图案, 再利用透镜 2 将数字微镜阵列反射的光会聚到单像素探测器 (Thorlabs PDA-100 A2, gain = 0) 上。单像素探测器将光信号转化为电信号, 经数据采集卡 (National Instruments USB-6366 BNC, 2 MHz) 量化后, 输入至后半部分的电子神经网络, 如图 3 所示, 实现基于光电混合神经网络的单像素物体分类。转盘上手写数字的制作是将手写数字镂空地雕刻在黑色亚克力板上。转盘由电机驱动旋转, 并通过脉宽调制器的电压占空比 (Pulse Width Modulation, PWM) 来控制运动速度。

数字微镜阵列 DMD 是一种开、关 (即 0、1) 的二值化调制器件, 而从网络中提取出来的卷积核包含负值。为了能够在 DMD 上显示卷积核图案, 对卷积核进行了如下处理:

(1) 首先调整卷积核 $K(x,y)$ 的数值范围, 使其范围分布在 $-1 \sim 1$ 之间:

$$K'(x,y) = \frac{K(x,y)}{|K(x,y)|_{\max}} \quad (3)$$

式中: $|K(x,y)|_{\max}$ 表示取卷积核 $K(x,y)$ 中所有元素绝对值的最大值。

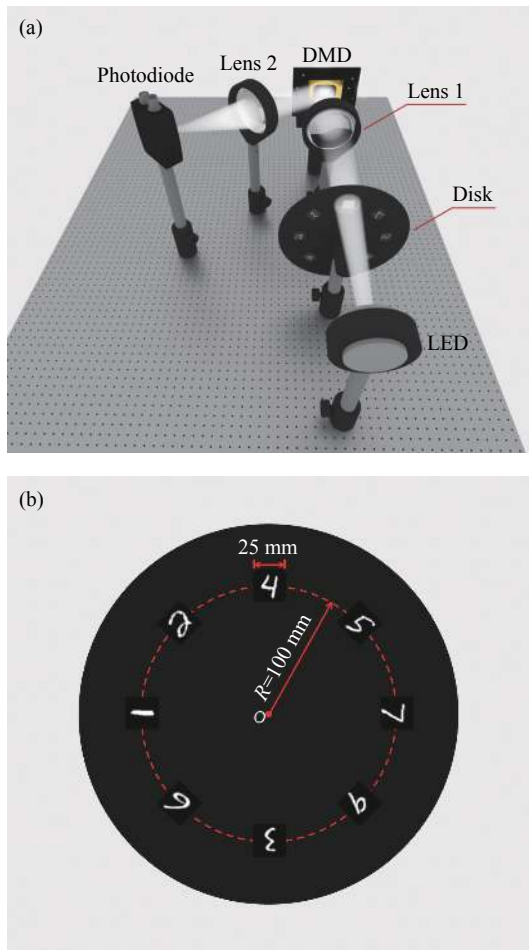


图 8 光学系统。(a) 实验装置; (b) 转盘上手写数字的分布

Fig.8 Optical system. (a) Experimental setup; (b) Layout of the handwritten digits on disk

(2) $K'(x,y)$ 中强度值大于 0 的像素为 P^+ 图案, 小于 0 的像素将强度值取反为 P^- 图案:

$$P^+(x,y) = \begin{cases} K'(x,y) & , K'(x,y) > 0 \\ 0 & , \text{other} \end{cases} \quad (4)$$

$$P^-(x,y) = \begin{cases} -K'(x,y) & , K'(x,y) < 0 \\ 0 & , \text{other} \end{cases}$$

(3) 将生成的一系列卷积核图案插值到符合实际场景的尺寸, 再利用抖动算法^[21] 对其进行二值化操作。

图 9 为对图 5 中第一张卷积核图案进行上述处理后得到的 P_1^+ 和 P_1^- 图案。那么利用数字微镜阵列显示一对 P_1^+ 和 P_1^- 图案, 将单像素探测器分别测量到的光强值相减, 即表示内积运算值 (对应公式 (2))。

利用数字微镜阵列显示二值化的卷积核图案, 单像素探测器采集每对卷积核对应的光强值作差, 送入图 3 所示的全卷积神经网络进行识别, 实现基于光电混合神经网络的单像素物体识别。

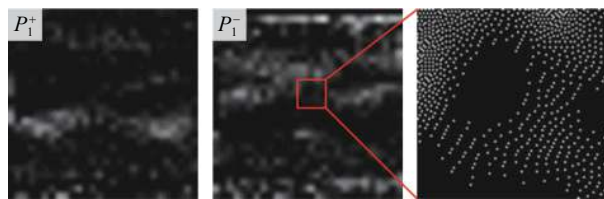


图 9 一对二值化的卷积核图像

Fig.9 A pair of binarized convolutional kernel images

设计了三组不同速度的运动物体分类实验, 通过设置 PWM 为 0%、20% 和 40%, 控制手写数字运动的线速度分别为 1.364、2.450、4.926 m/s, 其对应转速分别为 2.17、3.90、7.84 r/s。随着运动速度的提高, 相同时间内经过视场的手写数字的数量随之增加。为直观地展示快速运动物体的速度, 使用一台 60 fps 的相机 (FLIR, BFS-U3-04 S2 M-CS) 拍摄手写数字在不同

运动速度下的视频。相机的曝光时间为 1/60 s, 帧尺寸为 180×180 pixel。图 10 为从视频中保存下来的数字“5”的帧图像, 即使物体以转盘的最低速度运动, 人眼也很难准确地分辨出该数字。

实验中数字微镜阵列 DMD 以最高的刷新速率 (即 22 727 Hz) 多次循环显示卷积核图案。图 11(a) 给

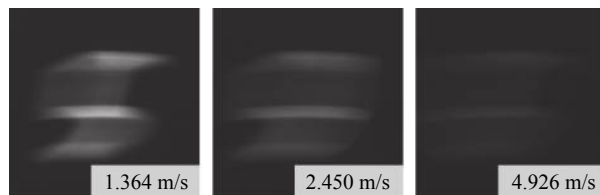


图 10 不同运动速度下用相机拍摄的数字“5”

Fig.10 Snapshots of digit "5" in motion at different speeds captured by using a camera

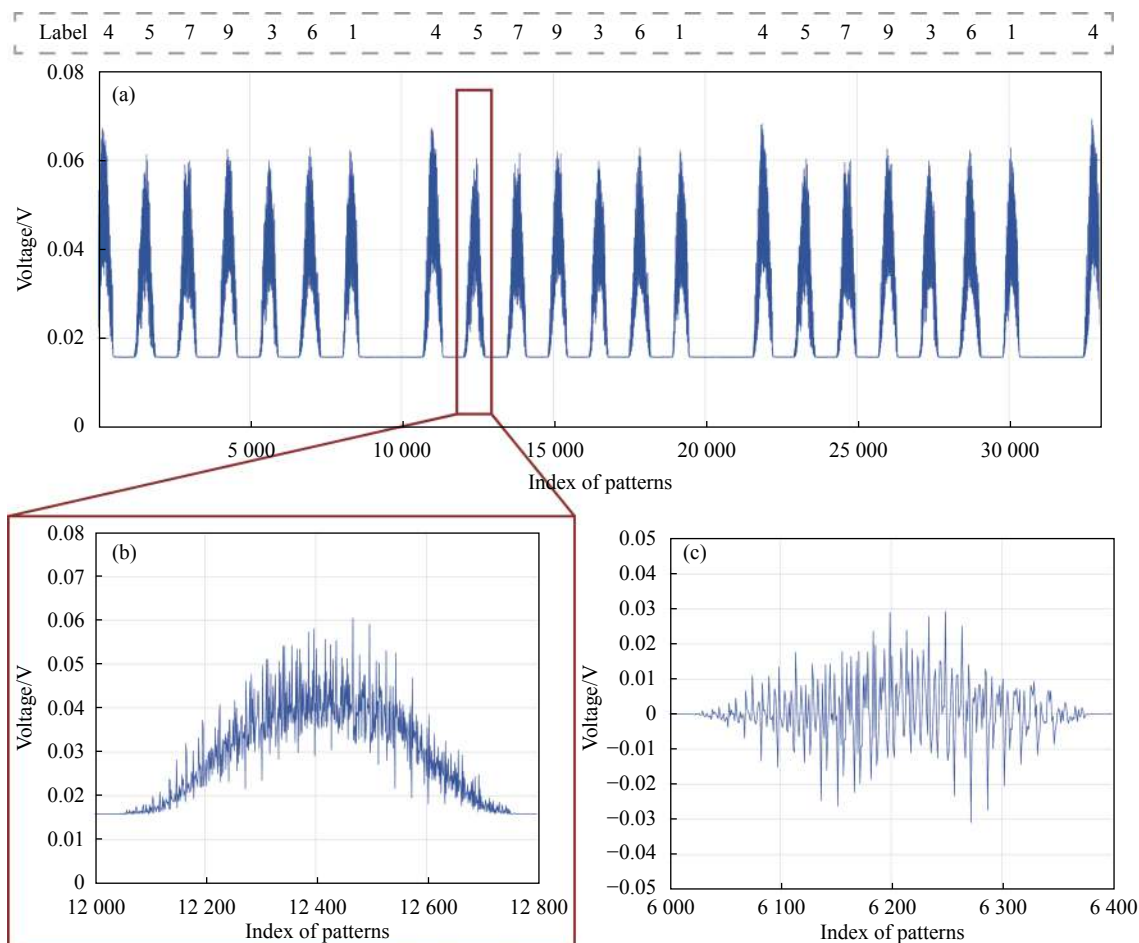


图 11 运动手写数字的单像素探测值。(a) 1.5 s 内手写数字连续经过视场采集到的单像素探测值; (b) (a) 中数字“5”单像素探测值的局部放大图; (c) 对 (b) 两两作差得到的结果

Fig.11 Single-pixel measurements of moving handwritten digits. (a) Single-pixel measurements of handwritten digits passing through the field of view successively in 1.5 s; (b) Partially enlarged view of the single-pixel measurements of the digit "5" in (a); (c) Result of the differential measurement from (b)

出了转盘以 2.17 r/s 转速 (1.364 m/s 线速度) 运动时, 在 1.5 s 内手写数字连续地经过视场的单像素探测值。当视场中有手写数字经过时, 光可以透过转盘, 采集到的单像素探测值较高; 当视场中没有手写数字或不是完整的手写数字时, 光被挡住, 采集到的单像素探测值较低。为了确定某段数据对应的真实标签, 实验时人为地挡住转盘上的其中一个数字, 如数字“2”, 那么转盘上被挡住的部分会造成数值低的单像素探测值更多, 这样就简单方便地确定被挡住数字的位置, 实现了对转盘上手写数字的标记。

由于只有当完整的手写数字出现在视场中被测量, 才有可能被正确分类识别, 因此通过设置阈值把手写数字完整出现在视场中的测量数据筛选出来。阈值的设置方法有多种, 实验中的阈值由以下公式计算:

$$t = \frac{S_{\max} - S_{\min}}{\beta} + S_{\min} \quad (5)$$

式中: S_{\max} 表示单像素探测值信号的最大值; S_{\min} 表示单像素探测值信号的最小值; β 是一个控制阈值的高低的因子, 需根据不同的实验状况选取, 不同转速下 β 可能不同。由于反图案 P^- 对应的单像素探测值通常

较低, 所以只利用正图案 P^+ 对应的单像素探测值来筛选出可以用于分类的有效数据。在 P^+ 对应的单像素探测值中, 找出连续大于阈值的数据, 即为通过阈值筛选出的有效单像素探测值。图 11(b) 为图 11(a) 中数字“5”对应单像素探测值的局部放大图, 图 11(b) 中的数据利用阈值 (选 β 值为 2) 共可筛选出 265 个有效单像素探测值。图 11(c) 为对图 11(b) 中的单像素探测值两两作差的结果, 也是送入后续神经网络的数据。利用 15 对二值化卷积核图案来获取物体的特征信息, 每 30 个单像素探测值进行一次分类, 这样可以从图 11(b) 的 265 个有效单像素探测值中进行八次分类。

光电混合神经网络在不同速度下的分类结果见表 1。不同速度下对每种卷积核数量的网络进行三次重复实验, 每次实验的数据采集时间均为 1.5 s, 三次实验的分类结果汇总呈现在表 1 中。通过阈值筛选出有效的单像素测量值, 这些数据能进行的分类次数作为总的测量次数。统计所有分类次数中正确预测的分类结果, 与总测量次数相除得到该卷积核数量神经网络的分类准确率。

分析表 1 的实验结果: (1) 当手写数字低速运动

表 1 运动手写数字实验分类结果

Tab.1 Experiment classification results of moving handwritten digits

Linear velocity/m·s ⁻¹	Number of kernels	Correct	Total	Correct/Total
1.364	5	785	2181	35.99%
	10	523	681	76.80%
	15	584	607	96.21%
	20	339	339	100.00%
	25	323	346	93.35%
	30	180	195	92.31%
2.450	5	737	2110	34.93%
	10	399	605	65.95%
	15	464	535	86.73%
	20	249	271	91.88%
	25	209	263	79.47%
	30	190	287	66.20%
4.926	5	892	2679	33.30%
	10	543	973	55.81%
	15	420	625	67.20%
	20	190	332	57.23%
	25	145	326	44.48%
	30	114	301	37.87%

时,分类准确率总体上随着卷积核数量的增加而提高,这与图 7 的测试集结果相吻合。在 1.364 m/s 时,卷积核数量为 20 的网络分类准确率达到 100.00%。但需要提醒的是,这是在 339 个样本下的分类统计结果,与图 7 的分类准确率并不矛盾(图 7 为 10 000 个样本的测试统计结果)。(2) 随着运动速度加快,分类准确率不再随着卷积核数量的增加而一直提高。因为卷积核数量越多,获取一次分类需要的数据采集时间越长,物体运动模糊带来的负面影响就越严重。在 1.364 m/s 和 2.450 m/s 时,20 个卷积核数量的网络取得的分类准确率最高,而在 4.926 m/s 时,15 个卷积核数量的网络达到的分类准确率高于 20 个。(3) 卷积核数量越多,分类准确率随运动速度下降得越严重。对比 15 个和 30 个卷积核网络在 1.364 m/s 和 4.926 m/s 的分类准确率,15 个卷积核网络的准确率下降了近 30%,而 30 个卷积核网络的准确率下降了

近 60%。

3 讨论

上述实验都是以手写数字为分类对象,为了说明提出的光电混合神经网络也适用于手写数字以外更复杂的物体的分类识别,用 Fashion-MINST 服饰图片数据集^[2]仿真测试了提出的光电混合全卷积神经网络对服饰复杂物体的分类能力。该手写数据集共包含 10 个类别,如图 12 所示,有 60 000 张训练集图片和 10 000 张测试集图片,每张图片的尺寸为 28×28 pixel。网络的设计和参数与图 2 一致。同样地,为第一层卷积层设置不同数量的卷积核进行训练,对应的测试集分类准确率如图 13 所示。随着卷积核数量的增加,分类准确率总体呈上升趋势。与手写数字相比,提出的全卷积神经网络分类服饰的能力稍弱。当卷积核数量超过 17 个时,分类准确率在 88% 左右稳定。

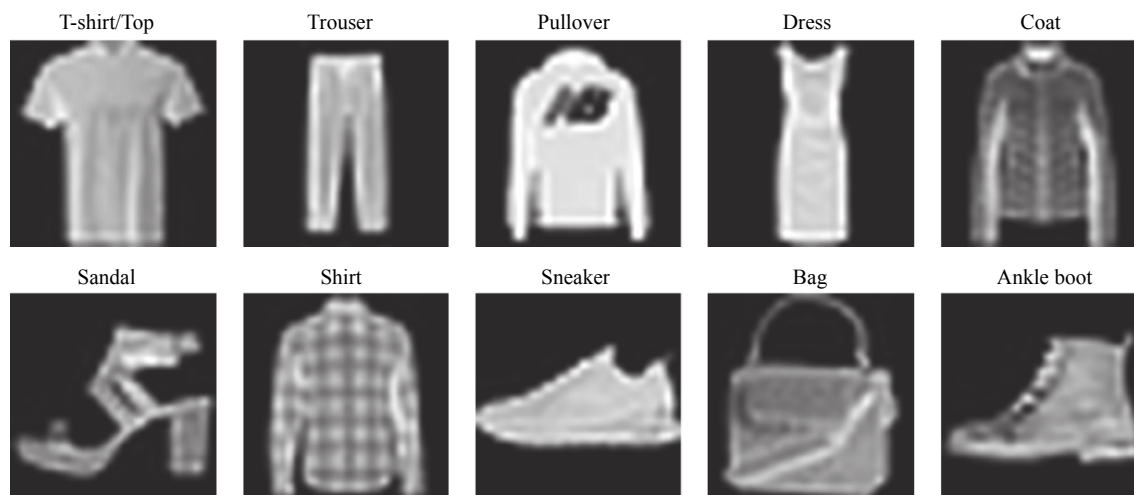


图 12 Fashion-MINST 数据集中的十个类别和示例图片

Fig.12 The ten classes and example images in Fashion-MINST dataset

为了评估所采用的免图像物体分类方法的分类能力,与现有的一些基于手写数字图像的分类方法的分类能力进行了比较。LeCun Y 等人比较了 69 种主要的分类器在 MNIST 手写数字数据集上的准确率^[20],表 2 列出了其中四种经典的分类方法的分类准确率和文中提出的光电混合全卷积神经网络对静态手写数字分类的准确率(为了公平比较,使用的数据集均为未经处理的原始手写数字图片,提出的光电混合全卷积神经网络的卷积核数量为 15 个)。由表 2 给出的

准确率可知,提出的光电混合全卷积神经网络在 MNIST 数据集上取得的准确率稍低,但文中提出的免图像方法产生的数据量少,对动态物体的分类具有优势。

所提出的免图像物体分类方法主要依赖于空间光调制器对目标物体像的调制,以获取用于物体分类的空间特征信息,并且需要多次调制测量才能进行一次分类识别,因此空间光调制器的刷新速率是限制高转速下物体分类准确率的主要原因,增加空间光调制

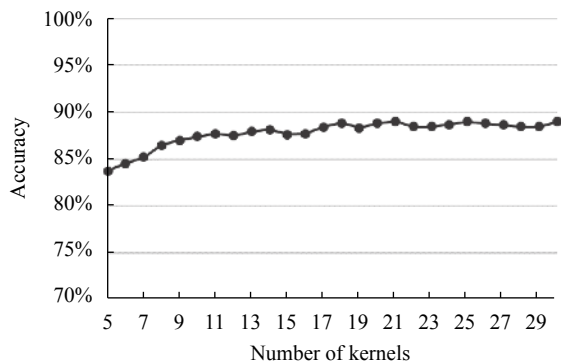


图 13 不同卷积核数量网络在 Fashion-MINST 测试集的分类准确率
Fig.13 Fashion-MINST test set classification accuracy of networks with different number of convolutional kernels

表 2 不同模型的 MNIST 数据集分类结果

Tab.2 Results of different models on MNIST datasets

Classifier name	Accuracy
Linear classifier ^[20]	88.00%
SVM ^[23]	98.60%
6-layer neural network ^[24]	99.65%
Deep convolutional network ^[25]	99.65%
Proposed network	97.99%

器的调制频率可能是提高高速旋转下分类准确率的最有效方法。另外,在相同的空间光调制器的调制频率下,如果采用更复杂、层数更多的网络也许能稍许提高分类准确率,但是可能会增大计算量,拖慢分类识别的速度。

实验中,采用了仿真数字进行网络训练,尽管针对旋转圆盘上的手写数字进行分类实验,仿真时加入了一些旋转和平移,但是在运动物体进入视场后并不能精确仿真模拟目标物体在视场中的姿态。这确实是造成更高运动速度下分类准确率不高的一个原因。因此,在仿真训练时,根据具体的目标物体的运动状态更精细化仿真模拟物体在视场中的姿态,有助于提高运动物体的分类准确率。

4 结 论

文中提出了一种基于光电混合神经网络的单像素快速运动物体分类方法。该方法采用单像素结构化探测光场调制方式,以来自目标物体的透射光作为光电混合神经网络的输入,利用数字微镜阵列调制目

标物体的像,调制图案来自网络学习得到的卷积核。单像素测量过程即为该全卷积神经网络的第一层卷积运算过程,实现了光探测和电子计算神经网络的无缝衔接,单像素探测器获取的测量值送入后续的电子计算神经网络部分,最后完成目标物体的分类。

快速旋转盘上手写数字的持续分类实验证明了提出的方法具有分类快速运动物体的能力。选择合适数量的卷积核,在相对低速运动(1.364 m/s)时,网络的分类准确率可达 90% 以上;在较高速运动(4.926 m/s)时,网络的分类准确率依然超过 60%。实验结果表明,该方法在分类快速运动的手写数字方面的能力超过了人眼视觉,为快速运动物体的持续分类提供了一种新路径。

参 考 文 献:

- [1] Sermanet P, LeCun Y. Traffic sign recognition with multi-scale convolutional networks [C]//The 2011 International Joint Conference on Neural Networks. IEEE, 2011: 2809-2813.
- [2] Andreopoulos A, Tsotsos J K. 50 years of object recognition: Directions forward [J]. *Computer Vision and Image Understanding*, 2013, 117(8): 827-891.
- [3] Edgar M P, Gibson G M, Padgett M J. Principles and prospects for single-pixel imaging [J]. *Nature Photonics*, 2019, 13(1): 13-20.
- [4] Zhang Z, Ma X, Zhong J. Single-pixel imaging by means of Fourier spectrum acquisition [J]. *Nature Communications*, 2015, 6: 6225.
- [5] Gibson G M, Johnson S D, Padgett M J. Single-pixel imaging 12 years on: a review [J]. *Optics Express*, 2020, 28(19): 28190-28208.
- [6] Sun B, Edgar M P, Bowman R, et al. 3 D computational imaging with single-pixel detectors [J]. *Science*, 2013, 340(6134): 844-847.
- [7] Sun M J, Zhang J M. Single-pixel imaging and its application in three-dimensional reconstruction: A brief review [J]. *Sensors*, 2019, 19(3): 732.
- [8] Yao M, Cai Z, Qiu X, et al. Full-color light-field microscopy via single-pixel imaging [J]. *Optics Express*, 2020, 28(5): 6521-6536.
- [9] Latorre-Carmona P, Traver V J, Sánchez J S, et al. Online reconstruction-free single-pixel image classification [J]. *Image and Vision Computing*, 2019, 86: 28-37.
- [10] He X, Zhao S, Wang L. Ghost handwritten digit recognition

- based on deep learning [J]. *arXiv preprint arXiv*, 2020: 2004.02068.
- [11] Rizvi S, Cao J, Hao Q. High-speed image-free target detection and classification in single-pixel imaging [C]//SPIE Future Sensing Technologies. International Society for Optics and Photonics, 2020, 11525: 115250 X.
- [12] Fu H, Bian L, Zhang J. Single-pixel sensing with optimal binarized modulation [J]. *Optics Letters*, 2020, 45(11): 3111-3114.
- [13] Lin X, Rivenson Y, Yardimci N T, et al. All-optical machine learning using diffractive deep neural networks [J]. *Science*, 2018, 361(6406): 1004-1008.
- [14] Zhou T, Lin X, Wu J, et al. Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit [J]. *Nature Photonics*, 2021, 15(5): 367-373.
- [15] Shen Y, Harris N C, Skirlo S, et al. Deep learning with coherent nanophotonic circuits [J]. *Nature Photonics*, 2017, 11(7): 441-446.
- [16] Jiao S, Feng J, Gao Y, et al. Optical machine learning with incoherent light and a single-pixel detector [J]. *Optics Letters*, 2019, 44(21): 5186-5189.
- [17] Zhou Zhiping, Xu Pengfei, Dong Xiaowen. Computing on silicon photonic platform [J]. *Chinese Journal of Lasers*, 2020, 47(6): 0600001. (in Chinese)
- [18] Zhang Z, Li X, Zheng S, et al. Image-free classification of fast-moving objects using “learned” structured illumination and single-pixel detection [J]. *Optics Express*, 2020, 28(9): 13269-13278.
- [19] 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.
- [20] LeCun Y, Cortes C, Burges C J C. THE MNIST DATABASE of handwritten digits [EB/OL]. [2021-11-16] <http://yann.lecun.com/exdb/mnist/>.
- [21] Zhang Z, Wang X, Zheng G, et al. Fast Fourier single-pixel imaging via binary illumination [J]. *Scientific Reports*, 2017, 7(1): 1-9.
- [22] Xiao H, Rasul K, Vollgraf R. Fashion-mnist: A novel image dataset for benchmarking machine learning algorithms [J]. *arXiv preprint arXiv*, 2017: 1708.07747.
- [23] Burges C J C, Schölkopf B. Improving the accuracy and speed of support vector machines [C]//Proceedings of the 9th International Conference on Neural Information Processing Systems, 1996: 375-381.
- [24] Ciresan D C, Meier U, Gambardella L M, et al. Deep big simple neural nets excel on handwritten digit recognition [J]. *arXiv preprint arXiv*, 2010: 1003.0358.
- [25] Ciresan D C, Meier U, Masci J, et al. Flexible, high performance convolutional neural networks for image classification [C]//Twenty-second International Joint Conference on Artificial Intelligence, 2011: 1237-1242.