

具有感存算一体化的新型神经形态视觉传感器*

廖付友^{1,2} 柴扬^{1,2,†}

(1 香港理工大学深圳研究院 深圳 518063)

(2 香港理工大学应用物理系 香港 999077)

Neuromorphic vision sensors based on an in-sensor computing paradigm

LIAO Fu-You^{1,2} CHAI Yang^{1,2,†}

(1 Shenzhen Research Institute, The Hong Kong Polytechnic University, Shenzhen 518063, China)

(2 Department of Applied Physics, The Hong Kong Polytechnic University, Hong Kong 999077, China)

2020-07-29收到

† email: ychai@polyu.edu.hk

DOI: 10.7693/wl20210603

摘要 传统的数字图像处理系统包括图像传感器与图像处理单元，二者在物理空间上分离，图像信息在其间的传输造成了延时与能耗。此外，数字图像传感器基于“帧”的工作原理，可能丢失一些重要信息，或者造成数据冗余。人类视觉系统提供了一种高效并行的信息处理方式。神经形态视觉传感器能够模拟人类视网膜的功能，同时具备感知光信号、存储信号和进行信息预处理的功能。这类感存算一体化的神经形态视觉传感器简化了人工视觉系统的电路复杂性，提升了信息处理效率，节省了系统功耗。文章总结了传统的数字图像传感器存在的问题，介绍了几种重要的人工神经网络，讨论了新型神经形态视觉传感器的研究进展和存在的问题。

关键词 图像传感器，神经形态视觉传感器，人工神经网络，感存算一体化

Abstract Conventional digital image systems include both image sensors and image processing units, which are physically separated. The data transfer between them gives rise to time delay and high-power consumption. In addition, the digital imaging system works through a frame-based operation, which causes the loss of important information or results in data redundancy. Our human visual system offers an efficient and parallel information processing method. Neuromorphic vision sensors can emulate the functions of the human retina for sensing light signals, storing the signals, and performing information preprocessing. This paradigm greatly simplifies the circuit complexity of the artificial vision system, improves the efficiency of information processing, and reduces system power consumption. In this work, we summarize the problems of conventional digital image sensors, introduce a few important artificial neural networks, and discuss the research progress and existing problems of emerging neuromorphic vision sensors.

Keywords image sensor, neuromorphic vision sensor, artificial neural network, in-sensor computing

* 国家自然科学基金(批准号: 61851402)、香港研究资助局(批准号: 15205619)和深圳科技创新委员会(批准号: JCYJ20180507183424383)资助项目

1 引言

在生活中，我们会见到各种图像与视频，目前从真实世界中记录图像与视频的手段主要是通过基于半导体器件的图像传感器。市场对于图像传感器的要求与日俱增，早期的需求主要来自于包括智能手机在内的消费电子，目前新兴的无人驾驶汽车、无人机、机器人以及各种监测进一步加大了对图像传感器的需求。图像传感器的广泛使用要求尽可能地降低其功耗，在一些特别的场景中(比如无人驾驶)，还要求具有实时的响应与分析。

基于传统半导体(硅基或者砷化镓基)的电荷耦合器件(charge coupled device, CCD)或者有源像素图像传感器(active pixel sensors, APS, 有时也称为CMOS图像传感器)可以感知到真实世界,并将光学信息转化为包括电压、电流在内的电子信号。但是将收集到的电子信号转化成可视的图像,需要一个处理器对电子信号进行处理和存储单元来存储得到的数据。这样一个完整的成像系统由图像传感器、图像处理器和存储单元三部分组成。

图像传感器、图像处理器和存储器有着不同的功能与要求,其制备方法也不尽相同,所以一般会通过封装技术把三者集成在电路板上形成一个整体系统进行工作,这样它们在物理空间上有着一定的距离。APS可以和数字信号处理器以及存储器等电路集成在一片芯片上,相对于CCD有着较短的信号传输距离和更快的工作速度。此外,图像传感器需要从嘈杂的真实环境中提取模拟信号,而图像处理器通常是工作在基于冯·诺依曼架构的数字世界中,这样二者之间通常需要模拟数字转换器。随着图像的分辨率、视频的帧数越来越高,海量的数据需要在图像传感

器、图像处理器以及存储器之间进行转换与传输,这样不可避免地造成了很大的功耗以及延时。

与基于半导体器件的数字技术相比,人类的视觉系统工作速度不算快,精度也不算高,但是它的能耗却远远低于目前的数字技术,而且可以更为有效地从复杂环境中提取关键信息。如何理解人类视觉系统的工作原理,并将其原理应用到半导体器件中,从而实现新型的视觉传感器,进而取代或者加强传统的图像传感器,或许是一个提高成像的工作效率并降低功耗的有效方法。然而目前我们对人类视觉系统工作原理的理解实际上还相对有限,可以确定的一点是眼睛的功能类似于视觉传感系统,而大脑则类似于视觉处理系统。例如,物体通过眼睛的玻璃体投影到视网膜上的图像与真实世界是相反的,然后通过大脑的视觉皮层细胞进一步处理将图像反转,形成与真实世界一致的图像,这个两步过程开普勒早在1604年就已经指出,如图1中的蝴蝶图像所示。尽管人类视觉开始于眼睛,但是却发生在人脑中。

在本文中,我们将首先介绍可用于视觉传感器中的神经网络计算模型,随后将介绍可以应用于视觉传感器的传统半导体器件及其缺点,最后将讨论几种可以在传感器端实现部分计算功能的新型器件,从而可以进一步提高视觉传感的能效与时效。

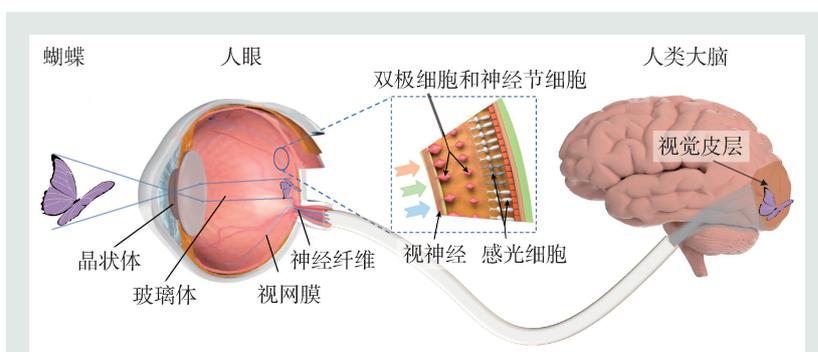


图1 人类视觉系统示意图^[1]。蝴蝶反射的光经过眼部的晶状体和玻璃体聚焦到视网膜,视网膜内的感光细胞将光信号转换成电信号,双极细胞接收感光细胞输出的电信号并传递给神经节细胞。经过视网膜预处理的电信号通过视神经传输给大脑视觉皮层做进一步的信息处理

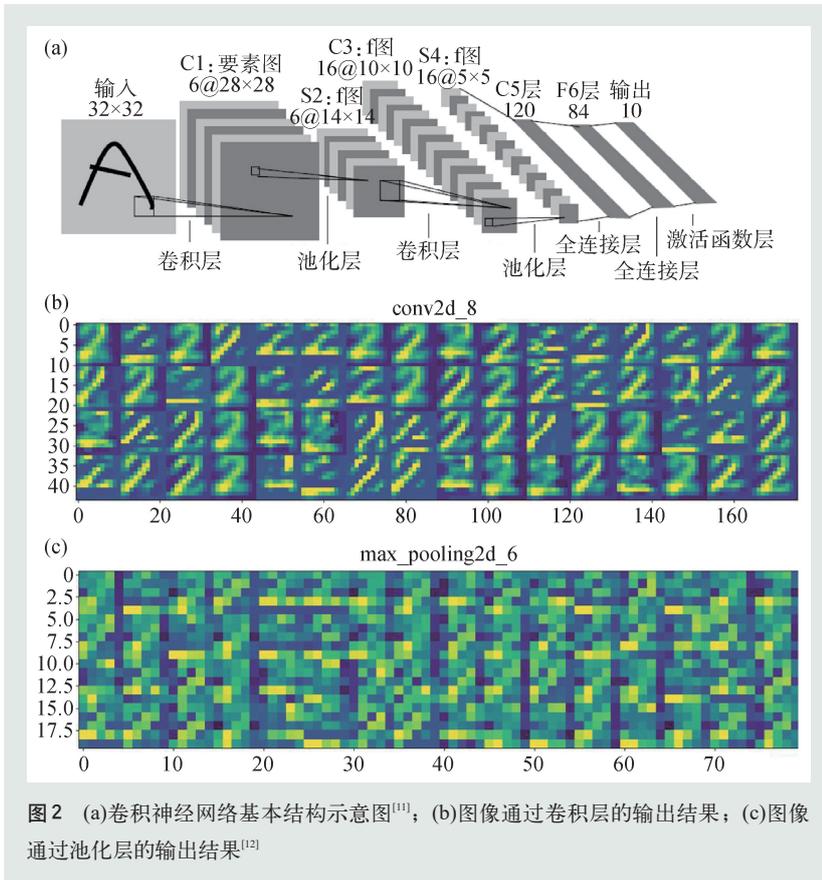


图2 (a)卷积神经网络基本结构示意图^[11]；(b)图像通过卷积层的输出结果；(c)图像通过池化层的输出结果^[12]

2 神经网络简介

人类视觉系统可以高效地处理复杂环境的一个重要原因是通过神经网络对视觉信号进行处理。神经网络是一种模仿生物神经网络结构与功能的计算模型，由大量的神经元高度互连进行计算，神经元之间的连接为权重可调的突触，它可以集体地、并行地展开存算一体的计算，具有分布式、并行化、局部性计算以及适应性的特点。在它的发展过程中，引入了许多统计学与信号处理的方法。在图像传感与处理过程中，可以采用不同的计算模型来处理不同的问题。

对于在神经网络中的单个神经元，从小的局部模型来看，具有线性关系，其数学模型可以表达为

$$z = f\left[\sum_{i=1}^n (w_i x_i + b)\right], \quad (1)$$

其中 z 为神经元输出， f 为传递函数或者激活函数， w_i 为来自不同突触的权重， x_i 为来自不同

突触的输入量， b 为偏置量。对于一个具有多层的神经网络，第 j 层的第 k 个神经元的输出可以表达为

$$z_k^j = f\left[\sum_{i=1}^n (w_{ik} x_{ik} + b_k^j)\right]. \quad (2)$$

从其代数表达式可以看出，人工神经网络中存在大量的向量矩阵乘法。关于人工神经网络的更多知识已经有很好的综述文章^[2-6]，本文只简单介绍可应用在新型视觉传感器的两种人工神经网络模型。

2.1 脉冲神经网络

人类的视网膜是一种精妙的三层结构，具有感光功能的视杆和视锥细胞在最内层，视杆和视锥细胞收到视觉信号，然后传递到中间的双极细胞层，对这些信号

号进行加工处理，最后把经过处理的信号传递到视网膜的表层，表层主要由神经节细胞构成^[7, 8]。神经节细胞根据收到的信号产生一个动作电位，只将有用的信息通过视神经传输到大脑皮层中，神经节细胞的这种行为具有事件驱动的特性，从而将大量的冗余信息过滤掉，可以极大地降低信息处理能耗。

脉冲神经网络(spiking neural networks, SNN)是一个比较接近真实神经网络的计算模型^[9]，一个神经元可以接收从多个神经元来的脉冲输入，当输入累计到一定程度达到阈值时，这个神经元就会输出一个新的脉冲，传输到它之后的神经元。SNN和深度学习神经网络(deep learning neural network, DNN)最大的区别是，脉冲神经元是关于时间的序列，每个神经元会接收多次输入，并且可以被激活多次；而DNN中每个神经元在单次计算中只会用到一次信息。神经编码着眼于理解神经元如何对不同刺激做出的反应，SNN遵从的是“时间编码”，基于单个电峰的确切出现时间，与外界的刺激绑定，非常适合模拟视觉系统

的工作。而DNN遵从的则是“频率编码”，单纯取决于动作神经元的“发放频率”，也就是单位时间内的平均电峰数量。然而在视觉系统中，发放频率会随着刺激强度非线性地增加，因此“频率编码”效率较低，但是信噪比好。SNN具有离散的特性，最大的优势在于其能够充分利用基于时空事件的信息。因为它具有事件驱动的特征(“时间编码”)，并且信息以稀疏方式记录，所以SNN有着极低的功耗来实时处理动态感知数据。

Leaky integrate-and-fire(LIF)模型接近于真实的生物神经元，经常用来构建脉冲神经网络。在该模型中，神经元接收到输入电流后，膜电势会升高，直至达到激活阈值，释放脉冲；脉冲释放后，神经元膜电势立刻恢复至静息电位。该模型可以简单地用一个RC电路来形容，其动力学过程可以用以下微分方程来描述^[10]：

$$I(t) = C_m \frac{dV_m(t)}{dt}, \quad (3)$$

其中 V_m 为神经元的膜电势， C_m 为神经元膜电容， $I(t)$ 为外界输入电流。

然而神经元在没有输入的情况下，神经元的膜电势能够很快恢复到静息电位，为了能反应该特征，通常将电路模型修正为一个电容与一个电阻的并联电路，其动力学过程可以用以下微分方程来表示：

$$I(t) - V_m/R_m = C_m \frac{dV_m(t)}{dt}, \quad (4)$$

其中 R_m 为等效的并联电阻。

2.2 卷积神经网络

卷积神经网络(convolutional neural networks, CNN)是一类包含卷积计算且具有深度结构的前馈神经网络(feedforward neural networks)，其每一层的特征都是由上一层的局部区域通过共享权值的卷积核激励得到的，非常适合于视觉图像特征提取与分类。人类的视觉原理也是从瞳孔摄入原始信号，接着做初步处理(大脑

皮层某些细胞发现边缘和方向)，然后抽象(大脑判定眼前物体的形状)，再进一步抽象。在卷积神经网络中，低级卷积层提取低级特征如边缘、线条、角落，更高层的卷积层提取更高级的特征。

CNN的基本结构主要包括卷积层、池化层、激活函数层和全连接层4个部分，如图2(a)所示^[11]。卷积层是卷积核在上一级输入层上通过逐一滑动窗口计算而得，卷积核中的每一个参数都相当于传统神经网络中的权值参数，与对应的局部像素相连接，将卷积核的各个参数与对应的局部像素值相乘之和，得到卷积层上的结果。一般地，使用卷积核进行特征提取和特征映射(图2(b))^[12]。在池化层中进行池化操作后的图像矩阵大小相比于输入的图像矩阵变小了，也就是说池化操作其实就是一个下采样的过程。图2(c)是一张图像经过池化层之后的输出结果，相比卷积层的输出，池化层的输出更加抽象，通过这样的下采样能够很好的保留主要信息，减少计算量^[12]。激活函数可以实现复杂的映射函数，它可以带来非线性的结果，而非线性可以很好的拟合各种函数。全连接层的作用是将经过多个卷积层和池化层的图像特征图中的特征进行整合，获取图像特征具有的高层含义，之后用于图像分类。

随着图像分辨率的提高，大量的图像数据导致传统图像处理效率很低，CNN可以有效地将图像降维成小数据量，可以把一个高像素的图像压缩为一个低像素的图像，并保持其图像特征。数字

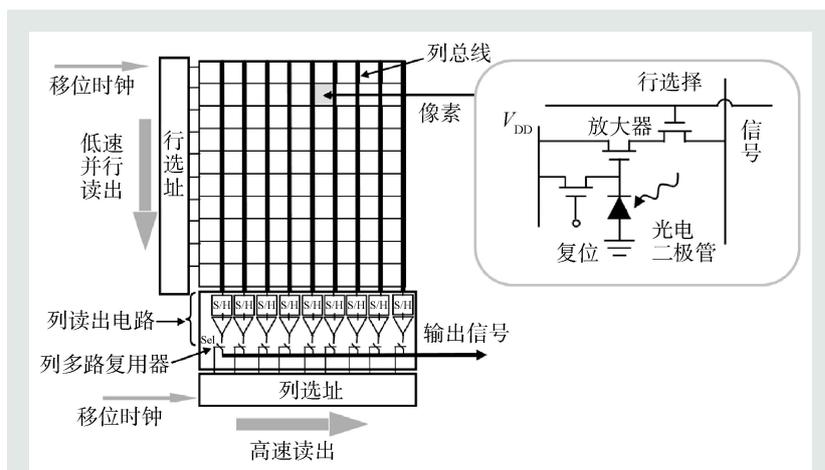


图3 APS结构示意图以及APS单个像素电路^[13]

化图像处理很难保留原有的特征，导致图像处理的准确率不高，而CNN能够有效地保持其图像特征。

3 传统的数字图像传感器

传感的数字图像传感器有电荷耦合器件(CCD)和有源像素图像传感器(APS)两类，目前市面上占主流的是APS，因此本节主要介绍和讨论APS。APS的结构包括像素单元阵列与外围辅助控制电路两大部分。像素单元是APS的核心，每一个像素内除了反偏的pn结光电二极管和选择晶体管外，还包含一个栅极注入光电流放大器和一个复位晶体管(图3右图)。当一束光照射在光电二极管上，反向偏置光电二极管的输出电流随光强增加而线性增加，这样像素单元就可以实现光电信号的转换，并输出电信号。外围辅助控制电路的主要任务是产生像素单元的驱动信号，将模拟信号进行放大和模数转换处理等。APS的工作过程是通过光学镜头将景物聚焦到图像传感器像素

阵列上，此时时序发生器对像素单元阵列复位后开始进行光电转换，产生光生电荷，形成电信号。在行列选择译码器的控制下，依次选通行列总线，将电信号输出到模拟信号处理模块进行降低噪声处理，使电信号获得较好的信噪比，然后经模数转换器转换成数字信号输出，如图3左图所示。

基于APS的传统图像处理系统，需要复杂的电路设计，而且数据存储单元和图像处理单元之间物理上分离，对器件的集成度和功耗都提出了巨大的挑战。这类传统的图像传感器都是由时钟信号和控制信号进行驱动，并且以一连串“帧”的方式获取图像信息，与现实世界的动态信息无关，导致“帧”与“帧”之间可能丢失一些重要的信息。同时每一“帧”都来源于所有的像素，导致数据量庞大。当前最先进的人工视觉系统，能够连续实时地探测图像，但是会产生大量的冗余数据，这些数据需要占用大量的存储空间，并引起巨大的功耗。此外，传统的图像传感器还存在

数据传输延迟、数字化信号造成图像特征的损失等问题，因此数字型器件难以有效实行神经网络模型的计算。

4 新型的神形态视觉传感器

人类视觉系统中的视网膜，不仅能够探测光信号，还能够对图像进行初步的预处理，然后将提取到的精简准确信息传递到大脑的视觉皮层，进行更加复杂的信息处理。受到人类视网膜工作原理的启发，近些年出现了神经形态视觉传感器，这类器件可以根据不同应用场景进行低级或高级的神形态

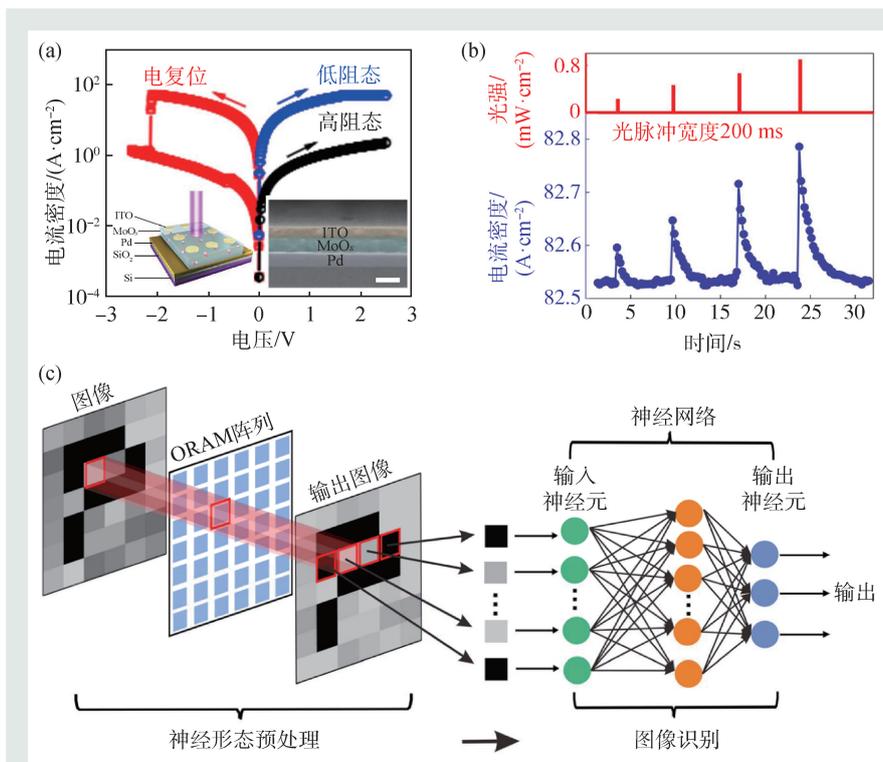


图4 基于MoO₃ ORRAM的神经形态视觉传感器^[29] (a)器件I—V特性曲线，插图是MoO₃ ORRAM的结构示意图和SEM图像；(b)在4个不同光照强度下的器件短程突触可塑性；(c)基于ORRAM的图像预处理和人工神经网络图像识别形成的神经形态视觉系统

运算。基于传统 CMOS 技术的神经形态视觉传感器已有许多相关的报道^[14-18]，例如动态视觉传感器 (DVS)^[19-22]、异步基于时间的图像传感器 (ATIS)^[23] 以及动态和有源像素图像传感器 (DAVIS)^[24, 25] 等等。传统的神经形态视觉传感器具有电路复杂、像素面积大、填充因子低和噪音高等缺点。为了克服这些缺点，近些年来报道了一些基于光子突触^[26, 27]、光控阻变存储器 (ORRAM)^[28, 29] 和二维材料光电器件^[30, 31] 等新型半导体器件的神经形态视觉传感器。单个或者少数器件即可同时具备对光信号的感知、对信号进行初步计算以及存储的功能，有望实现感存算一体化的神经形态视觉传感器。

4.1 具有预处理功能的神经形态视觉传感器

传感器在嘈杂的环境中通常收集到信噪比较大的非结构性数据，为了做进一步的数据处理，需要选择性地对收集到的时空信息进行有效的预处理，从大量的原始数据中提取有用的信息，降低噪声与失真，并提高数据特征，以进行后续更复杂的信息处理。

我们设计了一种两端结构的 Pd/MoO_x/ITO ORRAM 突触器件^[29]，如图 4(a) 所示。该器件能够感知紫外光，在 365 nm 光照下由高阻态变成低阻态，通过施加一个负偏压，可以对其进行复位操作。该器件具有光可调节的突触特性，改变光照时间和光照剂量可以改变器件的突触行为(图 4(b))。基于 MoO_x ORRAM 的 8×8 图像传感器阵列，具有图像存储和图像预处理功能。例如在紫外光下将图像投射到图像传感器，停止光照 3 分钟后，阵列依然能够记住图像，并且光照越强记忆越明显。此外，传感器阵列还能够对图像进行预处理，例如实现对对比度增强和降低噪音。利用

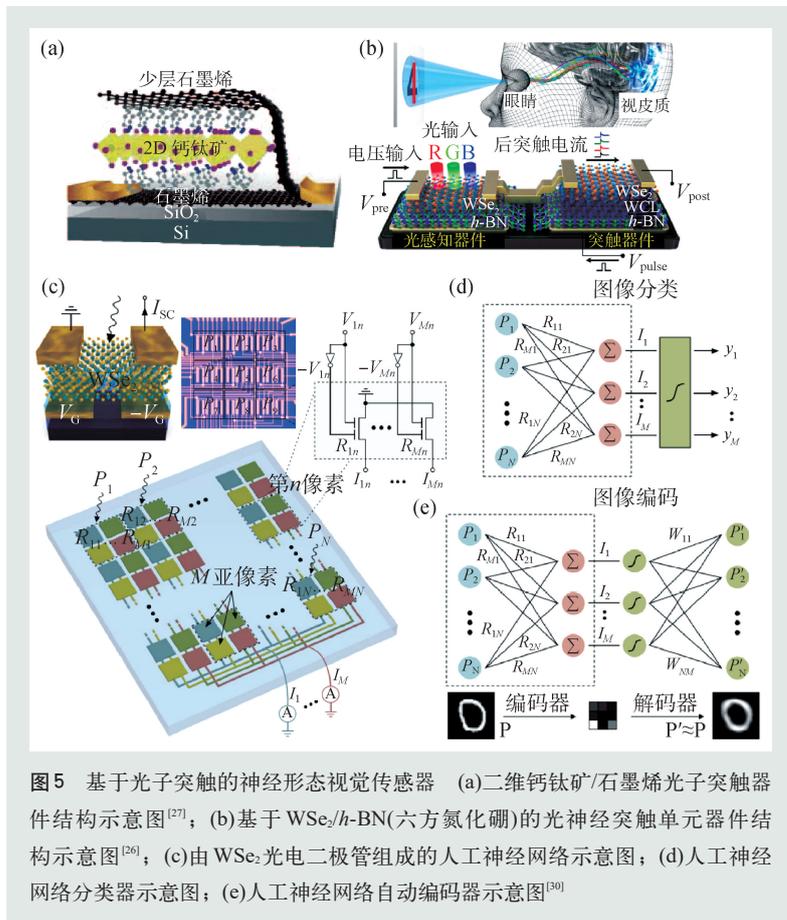


图5 基于光子突触的神经形态视觉传感器 (a)二维钙钛矿/石墨烯光子突触器件结构示意图^[27]; (b)基于 WSe₂/h-BN(六方氮化硼)的光神经突触单元器件结构示意图^[26]; (c)由 WSe₂光电二极管组成的人工神经网络示意图; (d)人工神经网络分类器示意图; (e)人工神经网络自动编码器示意图^[30]

这样的 ORRAM 阵列图像传感器和人工神经网络可以构建一个神经形态视觉系统(图 4(c))。结果表明，基于 ORRAM 的视觉系统只需要 1000 次训练即可达到 0.986 的准确度，在准确度、处理速度和功耗方面明显优于没有经过 ORRAM 预处理的视觉系统。利用 ORRAM 突触器件可以模拟人类视网膜的功能，简化神经形态视觉系统的电路和降低功耗，有助于进一步提升动态视觉信息的处理效率。

4.2 具有高级处理功能的神经形态视觉传感器

高级的视觉信号处理可以分析更为抽象的内容，包括图像的定位、识别与分类，该信号处理过程可以识别出输入传感信号的位置与内容等。神经网络模型可以有效地开展高级的视觉信号处理。光子突触器件的权重可以通过光来调节，这一特性集成了信息传感与处理的功能，可以用来模拟人类视网膜的性能。利用光子突触和其他神

经元器件可以构成视觉神经网络,实现神经形态视觉系统。

Tian 等人^[27]利用二维钙钛矿/石墨烯异质结实现了一种可重构光触发增强/抑制功能的光子突触(图 5(a)),具有较高的光响应度(730 A/W)和很高的稳定性(74 天)。利用这样的光子突触和输出神经元可以组成一个两层的神经网络,该神经网络的图像识别率高达 80%。Seo 等人^[26]在同一片 WSe₂ 材料上制备了一个 WSe₂/h-BN 光感知器件,再串联一个 WSe₂/WCL/h-BN 突触器件(图 5(b)),组成了光神经突触单元。其对不同波长的可见光具有不同的响应特性。利用光神经突触组成的神经网络可以模拟人类视觉系统的彩色和混色模式识别。最近 Mennel 等人^[30]报道了一种直接利用可重构的二维半导体 WSe₂ 光电二极管阵列构成的人工神经网络,它同时具备光信息感知和处理的功能,适合应用在超快的机器视觉领域(图 5(c))。每个光电二极管的光响应特性可以由两个分立的栅极独立控制,每个器件的光响应度(突触权重)可以存储在一个连续可调的光响应矩阵中。结果证明由 27 个器件组成的 3×3 阵列(每个像素有 3 个亚像素)的神经网络图像传感器(图 5(c))

可以执行监督式学习和非监督式学习,通过训练可以对图像进行分类(图 5(d))和编码(图 5(e))。这样的神经网络视觉传感器进一步实现了感存算一体化,在集成度、工作效率和能耗方面都具有巨大的优势。

5 结论与展望

神经形态视觉传感器由感知、存储和运算功能的器件与电路组成,这个交叉科学工程研究包含了材料、器件、电路、算法以及集成技术等,神经形态视觉传感器的成功实现需要以上多方面协调发展。例如在材料方面,一些新型的材料包括光致阻变材料、二维层状材料,对光信号有着独特的响应^[28],可以实现多功能的器件。在器件方面,一些多功能传感器件可以有效集成传感、存储与运算的功能,有效地简化电路结构^[32, 33]。在集成技术方面,Chiplet 封装技术可以借助已有的成熟技术,降低设计与制造成本,有效地提高集成度和性能。神经形态视觉传感器可以有效地降低能耗与时耗,实现高速并且复杂的视觉信号处理功能。

参考文献

- [1] Gu L *et al.* Nature, 2020, 581: 278
- [2] Tang J *et al.* Adv. Mater., 2019, 31: e1902761
- [3] Basheer I A *et al.* J. Microbiol. Methods, 2000, 43: 3
- [4] Jeong D S *et al.* Adv. Mater., 2018, 30: 1704729
- [5] Kim C H *et al.* Nanotechnology, 2018, 30: 032001
- [6] Xia Q *et al.* Nat. Mater., 2019, 18: 309
- [7] Masland R H. Nat. Neurosci., 2001, 4: 877
- [8] Zrenner E. Science, 2002, 295: 1022
- [9] Maass W. Neural Networks, 1997, 10: 1659
- [10] 程龙 等. 控制与决策, 2018, 33: 923
- [11] Lecun Y *et al.* Proc. IEEE, 1998, 86: 2278
- [12] 柯研 等. 电子技术与软件工程, 2018, 22: 72
- [13] Magnan P. Nucl. Instrum. Methods Phys. Res., Sect. A, 2003, 504: 199
- [14] Kyuma K *et al.* Nature, 1994, 372: 197
- [15] Posch C *et al.* Proc. IEEE, 2014, 102: 1470
- [16] Cho D D *et al.* Sensor. Mater., 2015, 27: 447
- [17] Delbruck T. In: 2016 46th European Solid-State Device Research Conference (ESSDERC), IEEE, 2016, p.7
- [18] Steffen L *et al.* Front. Neurorobotics, 2019, 13: 28
- [19] Lichtsteiner P *et al.* In: 2006 IEEE International Solid-State Circuits Conference-Digest of Technical Papers, IEEE, 2006, p.2060
- [20] Lichtsteiner P *et al.* IEEE J. Solid-St. Circ., 2008, 43: 566
- [21] Berner R *et al.* IEEE T. Circuits I: Regular Papers, 2011, 58: 1581
- [22] Lenero-Bardallo J A *et al.* IEEE J. Solid-St. Circ., 2011, 46: 1443
- [23] Posch C. In: 2010 IEEE International Solid-State Circuits Conference (ISSCC), IEEE, 2010, p.400
- [24] Berner R *et al.* In: 2013 Symposium on VLSI Circuits, IEEE, 2013, p.C186
- [25] Brandli C *et al.* IEEE J. Solid-St. Circ., 2014, 49: 2333
- [26] Seo S *et al.* Nat. Commun., 2018, 9: 5106
- [27] Tian H *et al.* In: 2018 IEEE International Electron Devices Meeting (IEDM), IEEE, 2018, p.38.6.1
- [28] Zhou F *et al.* Research, 2019, 2019: 9490413
- [29] Zhou F *et al.* Nat. Nanotechnol., 2019, 14: 776
- [30] Mennel L *et al.* Nature, 2020, 579: 62
- [31] Wang C Y *et al.* 2020, arXiv: 2003.02423
- [32] Chai Y. Nature, 2020, 579: 32
- [33] Zhou F *et al.* Adv. Funct. Mater., 2018, 28: 1800080