

基于多通道全光卷积神经网络的语义分割

谢小东^{1,2}, 张雨超^{1,2*}

¹上海理工大学光子芯片研究院, 上海

²上海理工大学智能科技学院, 人工智能纳米光子学中心, 上海

收稿日期: 2025年4月13日; 录用日期: 2025年5月5日; 发布日期: 2025年5月13日

摘 要

目前, 光学神经网络作为一种全新的人工智能技术, 以其高算力和低能耗得到了广泛的研究和应用。但是目前常规的光学神经网络结构简单, 缺乏对复杂任务的处理能力。文章提出了一种全新的多通道光学卷积神经网络(MC-OCNN)架构, 来处理机器视觉中的语义分割问题。MC-OCNN由透镜相位、随机相位, 以及光栅相位构建。透镜相位用来实现光学卷积操作, 随机相位用来组成光学卷积核, 以及光栅相位用来实现多通道架构。MC-OCNN具备多层次结构, 由光学卷积层以及光学全连接层级联而成, 以光速并行进行卷积计算, 并且实现了多维度提取物体不同特征, 以及多层次提取物体特征的功能。我们将MC-OCNN应用在人像分割任务以及飞机检测任务中, 取得了良好的结果, 为其进一步在机器视觉、自动驾驶、智慧医疗等领域的应用奠定了基础。

关键词

全光卷积神经网络, 光学卷积, 语义分割, 多通道

Semantic Segmentation Based on Multi-Channel All-Optical Convolutional Neural Networks

Xiaodong Xie^{1,2}, Yuchao Zhang^{1,2*}

¹Institute of Photonic Chips, University of Shanghai for Science and Technology, Shanghai

²Center of Artificial Intelligence Nanophotonics, School of Artificial Intelligence Science and Technology, University of Shanghai for Science and Technology, Shanghai

Received: Apr. 13th, 2025; accepted: May 5th, 2025; published: May 13th, 2025

*通讯作者。

文章引用: 谢小东, 张雨超. 基于多通道全光卷积神经网络的语义分割[J]. 建模与仿真, 2025, 14(5): 282-292.
DOI: 10.12677/mos.2025.145393

Abstract

Currently, optical neural networks (ONNs) have attracted wide research and application due to their high computational power and low energy consumption. However, the structure of conventional ONNs is simple, and their ability to handle complex tasks is very limited. Here, we propose a novel multi-channel optical convolutional neural network (MC-OCNN) architecture to address the semantic segmentation problem in machine vision. The MC-OCNN is composed of the lens phase, random phase, and grating phase. The lens phase is used to implement the optical convolution, the random phase is used to form the optical convolutional kernel, and the grating is used to construct the multi-channel architecture. MC-OCNN has a multi-layer structure formed by cascading optical convolutional layers and fully connected optical layers, and it also could parallelly perform convolution computations at the speed of light. Besides, it achieves the extraction of different features of the objects, as well as the multi-level extraction of object features. The MC-OCNN is applied to portrait parsing tasks and aircraft segmentation tasks, achieving promising results that demonstrate great potential for application in fields of machine vision, autonomous driving, smart healthcare, and so on.

Keywords

All-Optical Convolutional Neural Network, Optical Convolution, Semantic Segmentation, Multi-Channel

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来, 人工神经网络(ANNs)的发展推动了机器学习应用的革命性突破, 包括大型语言模型 ChatGPT [1]、医学影像诊断[2]、语音识别[3]、机器人技术[4]和自动驾驶[5]等领域。然而, 随着人工智能算法日益复杂, 传统电子硬件在处理技术、计算架构、功耗与散热等方面面临多重困难。受限于冯·诺依曼瓶颈, 现有硬件平台在处理时间和能耗上面临严峻挑战。

为克服电子平台的缺陷, 光学神经网络(ONNs)被提出, 该框架利用光学的高并行性、低延迟和显著能效优势, 在光速分类与高分辨率成像任务中展现出卓越性能[6]。由于在实空间处理计算机视觉任务的局限性, ONNs 已被扩展至傅里叶空间, 并在 MNIST 手写数字识别等任务中取得成功[7]。由于光学非线性可显著提升光学神经网络的分类精度与鲁棒性, 非线性研究已成为光学神经网络发展的重要方向, 包括采用光折变晶体(SBN: 60) [7]、可重构电光硬件[8]、电磁诱导透明(EIT) [9]和光学非线性调制[10]等技术实现非线性激活函数。基于硅基和自由空间架构的多种光学神经网络也被提出[11]-[13], 利用不同光学器件完成调制过程。在这些架构中, 光学神经网络已展现出强大的通用功能性与任务适应性。类似地, 光学神经网络在计算机图像处理中显示出巨大潜力, 涵盖图像识别[14]-[16]、图像去雾[17]、图像放大与缩小[18]等任务。片上集成光学神经网络的研发也快速推进, 基于衰减器与探测器[19]、衍射[20]、逆向设计[21]和高维方法[22]的实现方案, 均为光学神经网络的实际应用提供了可能。然而, 当前 ONNs 仍存在仅支持单通道执行单一任务的问题, 且处理更复杂计算机视觉任务的能力有限。

图像分割通常指对自然场景中显著目标的检测与分割, 被视为解释图像并从中提取有效信息的方法

或技术。作为图像分析与模式识别的第一步,传统图像分割方法基于边缘检测[23][24]与阈值分割[25][26]。为区分复杂图像结构,基于图论[27]的图像分割技术持续发展。随着深度学习的兴起,多种面向图像分割的卷积神经网络架构被提出,包括 FCN [28]、U-Net [2]和 R-CNN [29], 这些方法将图像分割性能提升至新高度。图像分割已在医学图像分割[30]-[32]、道路标线检测[33]、交通标志识别[34][35]和遥感成像[36]等领域展现出重要应用价值。

在本研究中,我们提出一种多通道光学卷积神经网络(MC-OCNN)架构,通过透镜相位构建了光学卷积操作,并基于孔径函数的随机相位组成了卷积核,以及基于二维光栅(2D Grating)实现了多通道卷积。MC-OCNN 以光速并行进行光学卷积计算,具备极高的算力,并且通过多个光学卷积层和光学全连接层的级联结构实现了物体特征的多层次提取能力,以及基于多通道架构实现了提取物体不同特征的能力。我们将其应用在语义分割的像素级分类任务中,发现在人像解析以及飞机目标分割任务中取得了较好的效果,证明了 MC-OCNN 处理复杂视觉任务的能力,为其进一步的应用奠定了基础。

2. 基本原理

2.1. 多通道光学卷积神经网络(MC-OCNN)的架构与原理

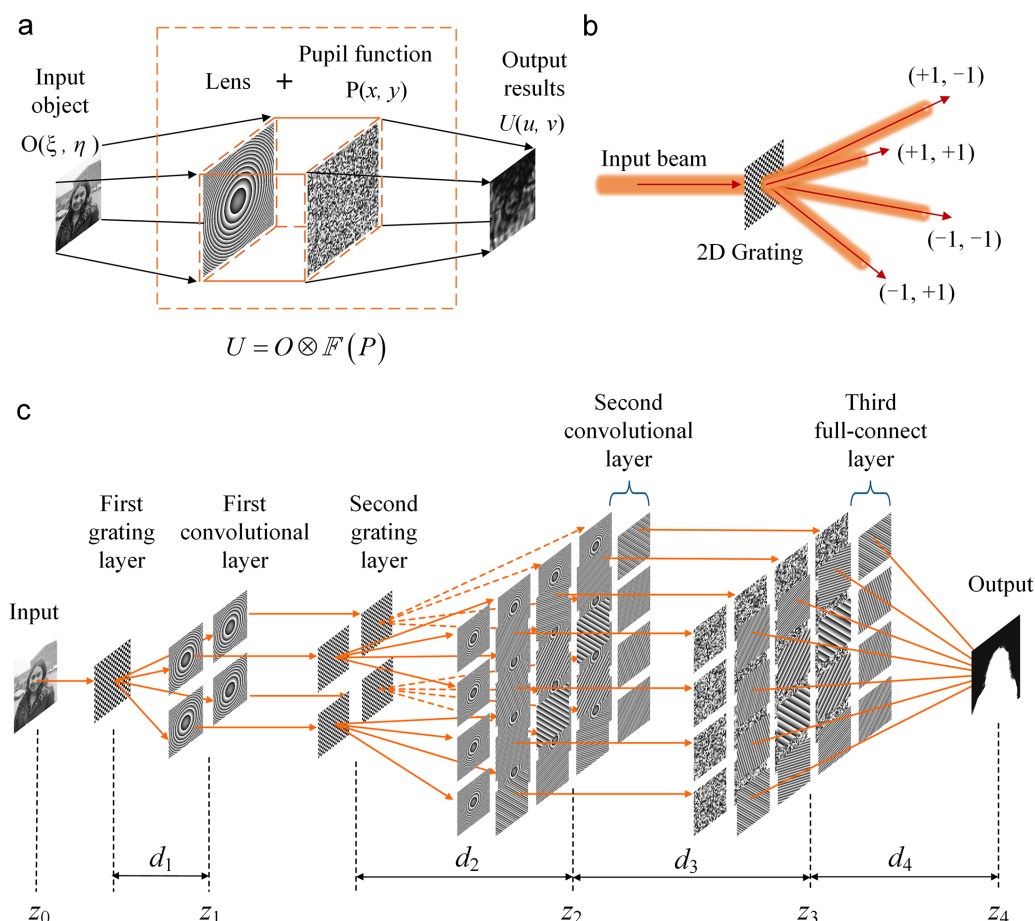


Figure 1. The architecture of a multi-channel optical convolutional neural network: (a) Optical convolution operation based on lens and aperture function construction; (b) Realization of beam splitting operations based on 2D gratings; (c) Overall architecture of MC-OCNN

图 1. 多通道光学卷积神经网络的架构: (a) 基于透镜以及孔径函数构建的光学卷积操作; (b) 基于二维光栅实现分束操作; (c) MC-OCNN 的整体架构

MC-OCNN 的架构与原理如图 1 所示, 该网络基于透镜相位以及透镜的孔径函数相位来形成光学卷积操作, 并基于二维光栅来构建多通道卷积架构。如图 1(a)所示, 透镜由一个二次相位以及孔径函数 $P(x, y)$ 组成, 其可以表示为:

$$L(x, y) = P(x, y) \cdot \exp\left[-i \frac{k}{2f} (x^2 + y^2)\right] \quad (1)$$

根据透镜的成像定律, 当物像满足成像条件时, 所得到的结果是物体与成像系统点扩散函数的卷积: $U = O \otimes H$, 其中点扩散函数 H 是透镜孔径函数的傅里叶变换: $H = \mathbb{F}[P(x, y)]$ 。因此, 基于这个特点, 通过将卷积核集成在透镜的孔径函数中, 我们构建了光学卷积操作。接着, 我们基于二维光栅(2D Gratings)构建了多通道光学卷积网络架构, 如图 1(b)所示。二维光栅通过周期性结构可以将入射光束分成多个不同的衍射级次, 每个衍射级次在空间中沿不同方向传播, 进而可以构建多通道的光学神经网络。二维光栅的表达式可表示为:

$$G_{2D}(x, y) = \sum_{m, n} \exp[i2\pi f_D(mx + ny)] \quad (2)$$

其中, f_D 代表是二维光栅沿 x 轴和 y 轴的空间频率, 根据对称性我们将其设置为同样的数值, $(m, n) = [0, \pm 1, \dots]$ 表示二维光栅在 x 方向以及 y 方向上的衍射级次。在本研究中, 我们第一层的二维光栅 G_{2D}^1 选择 $(m, n) = \{(+1, +1); (-1, +1); (+1, -1); (-1, -1)\}$, 此时二维光栅包含 4 个衍射级次。如图 1(b)所示, 光束通过二维光栅后, 将被分成四个衍射级次并向四个不同方向传播, 从而将不同的卷积通道在空间中分离出来, 实现对图像不同特征的提取功能。整体的多通道光学卷积神经网络(MC-OCNN)的架构如图 1(c)所示。物体位于 $z = z_0$ 的输入平面, 第一个二维光栅位于输入平面与第一个卷积层之间, 其与第一个卷积层的距离为 d_1 。根据光栅的分光定律, 对于衍射级次为 (m, n) 的分量, 其对应的空间频率为 (mf_D, nf_D) , 对应的传播方向为 $(\cos\theta_x, \cos\theta_y) = \lambda f_D(m, n)$, 因此经过 d_1 的传播距离后, (m, n) 衍射级次在空间中的偏移为 $(m\lambda f_D d_1, n\lambda f_D d_1)$ 。因此, 在经过第一个光栅层后, 光束分成了四份, 在第一个卷积层的位置形成了四个光斑。于是, 第一个卷积层被分成了四个区域对应于四个光斑。经过第一个卷积层后, 在第二个光栅层, 每个光栅又会将每个光束分为四个衍射级次。因此, 经过第二个光栅层后会形成 16 个光斑, 在空间中呈 4×4 阵列分布。第二个光学卷积层被分为 4×4 个区域, 每个区域会输出一个卷积结果。最后经过第三个光学全连接层, 会将这 16 个结果通过聚焦的方式, 合并后形成最后的输出结果。

接下来是 MC-OCNN 每个层的具体结构分布。如图 2(a)所示, 第二个光学卷积层位于平面 $z = z_1$ 处, 其被分为四个区域, 每个区域包含一个透镜相位用来实现光学卷积, 以及一个由随机相位构成的孔径函数: $P_k^1 = \exp(iM_k^1)$ ($k = 1, 2, 3, 4$), 其中 M_k^1 代表随机分布的相位数值。对于第一个卷积层来说, 其会生成 4 个卷积结果, 对应于 4 个特征图: $U_k^1 = O \otimes \mathbb{F}(P_k^1)$ 。其四个特征图分别由二维光栅 G_{2D}^1 的四个衍射级在空间上分离出来, 每个衍射级次对应的特征图如图 2(b)所示, $(+1, +1)$ 衍射级次对应的特征图位于左上角, 其余的衍射级次也在图中进行了标注。因此, 输入光束通过第一卷积层后, 被分为四个通道, 在空间中分别传播。根据其对应的衍射级次, 我们记通道 $(m, n) = \{(+1, +1); (-1, +1); (+1, -1); (-1, -1)\}$ 分别为 $C_1^1, C_2^1, C_3^1, C_4^1$ 。第二个光栅层位于第一个卷积层与第二个卷积层之间, 其被划分为 2×2 个区域, 对应于第一层生成的 4 个通道, 每个区域包含一个二维光栅 $G_{2D, l}^2$, 其中 l 代表四个不同的区域。每个区域的二维光栅的衍射级次如图 2(c)所示, 例如, 对于区域 1 (Zone 1)来说, 其衍射级次为 $\{(+1, +1), (-3, +1), (+1, -3), (-3, -3)\}$, 以及其对应的空间频率记为 f_D^2 。光束通过第二层光栅后, 每个通道 C_k^1 将被分为 4 个子通道: $C_{k, l}^2$ (对应 $k = 1 \sim 4, l = 1 \sim 4$), 使得第二层光栅共生成 16 个通道。

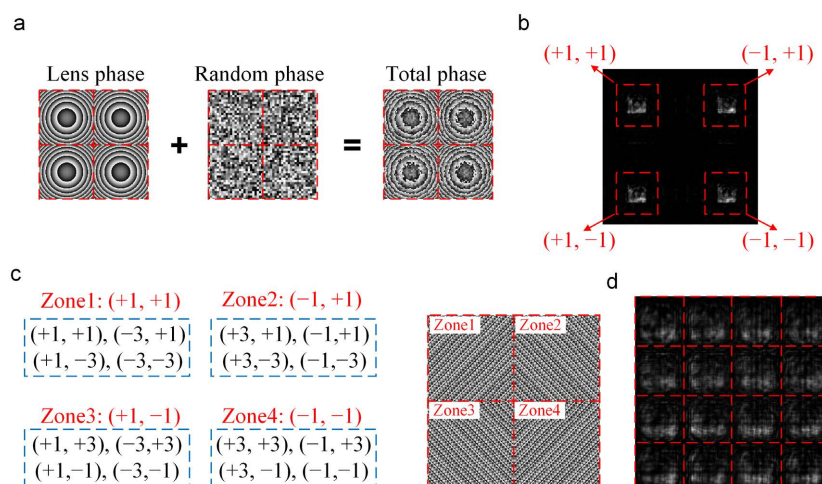


Figure 2. Phase distribution of each layer of multi-channel optical convolutional neural network: (a) Phase distribution of the first optical convolutional layer; (b) The feature map produced by the first convolutional layer; (c) Phase distribution of the second grating layer; (d) Distribution of the 16 light spots produced by the second grating layer

图 2. 多通道光学卷积神经网络每层的相位分布: (a) 第一个光学卷积层的相位分布; (b) 第一个卷积层产生的特征图; (c) 第二个光栅层的相位分布; (d) 第二个光栅层产生的 16 个光斑的分布

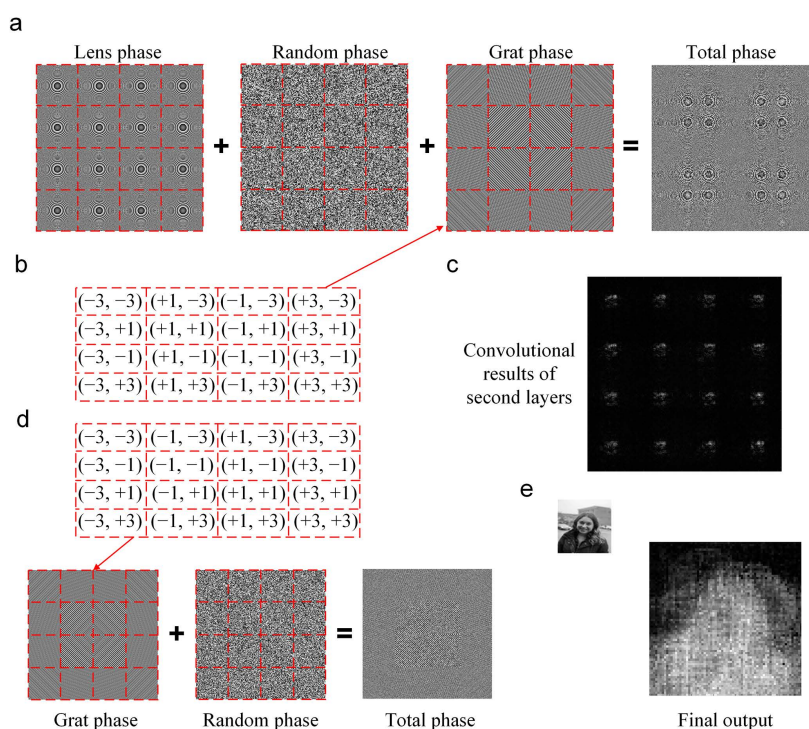


Figure 3. Construction of the second convolutional layer and the third fully connected layer of a multi-channel optical convolutional neural network: (a) Phase configuration of the second convolutional layer, consisting of the lens array phase, the randomly modulated phase, and the grating phase; (b) Distribution of grating diffraction levels per region in the second convolutional layer; (c) The result of the 16 convolutions formed by the second convolutional layer; (d) Construction of the third fully connected layer, consisting of the grating phase as well as the randomly modulated phase; (e) Final output of multi-channel optical convolutional neural network

图 3. 多通道光学卷积神经网络第二卷积层以及第三全连接层的构造: (a) 第二卷积层的相位构造, 由透镜阵列相位、随机调制相位, 以及光栅相位构成; (b) 第二卷积层中每个区域光栅衍射级次的分布; (c) 第二卷积层所形成的 16 个卷积结果; (d) 第三全连接层的构造, 由光栅相位以及随机调制相位构成; (e) 多通道全光卷积神经网络最终的输出结果

图 2(d)展示了第二个卷积层所处平面 $z = z_2$ ，由第二光栅层形成的 16 个通道，以 4×4 阵列在空间中的分布图。据此，第二个卷积层被分为 4×4 个区域，第二层的相位结构如图 3(a)所示。每个区域包含一个透镜相位，一个随机函数构成的孔径函数相位，以及一个光栅相位。增加光栅相位的原因主要是抵消在前面两层传播中光栅衍射所带来的光束传播的发散性，当引入光栅相位后，第二个卷积层所产生的卷积结果将会沿直线传播，每个区域对应的光栅衍射级次如图 3(b)所示。我们记第二个卷积层每个区域的孔径函数为 $P_{k,l}^2 = \exp(iM_{k,l}^2)$ ，其中 $(k=1 \sim 4, l=1 \sim 4)$ ，第二个卷积层所产生的 16 个卷积结果可以表示为 $U_{k,l}^2 = U_k^1 \otimes F(P_{k,l}^2)$ ，如图 3(c)所示，我们展示了第二个卷积层的结果，可以看到 16 个结果呈 4×4 阵列在空间中进行排列。第二层所产生的卷积结果位于 $z = z_3$ 平面，一共有 16 个卷积特征图。为了将这 16 个特征图合并后进行输出，我们引入了第三光学全连接层。如图 3(d)所示，第三个光学全连接层被分为 16 个区域，每个区域包含一个随机分布的调制相位 $M_{k,l}^3$ ，以及一个光栅相位 $G_{2D,k,l}^3$ ($k=1 \sim 4, l=1 \sim 4$)。光栅相位的作用是改变光束的传播方向，将第二层卷积层产生的 16 个特征图全部聚焦到屏幕中心进行合并输出结果。每个区域对应的光栅的衍射级次如图所示，将光栅相位和随机调制相位合并后就会得到第三全连接层的总相位分布。经过第三全连接层后，最后在输出平面 $z = z_4$ 的中心位置，会得到多通道全光卷积神经网络的最后输出结果，如图 3(e)所示。

2.2. 多通道光学卷积神经网络的训练

MC-OCNN 的训练过程基于误差反向传播方法，与常规衍射神经网络(D2NN)的训练方法相类似[6]。训练过程包含以下四个步骤：前向传播、损失计算、反向传播，以及梯度更新。具体训练过程为：首先，第一步是前向传播。对于第一光学卷积层来说，将光场从输入平面经过第一层的卷积后，形成四个卷积特征图，其可以表示为： $U_k^1 = O \otimes F(P_k^1)$ ($k=1 \sim 4$)。对于第二光学卷积层来说，其继续对第一层的四个特征图进行卷积，得到 16 个卷积特征图： $U_{k,l}^2 = U_k^1 \otimes F(P_{k,l}^2)$ ($k=1 \sim 4, l=1 \sim 4$)。对于第三个光学全连接层来说，其通过菲涅尔衍射[37]传播到输出层，因此 $U_{k,l}^3 = W_{k,l}^3 \otimes [U_{k,l}^2 \exp(iM_{k,l}^3) G_{2D,k,l}^3]_{\text{vec}}$ ，上式是用矩阵乘法来计算菲涅尔衍射传播， $W_{k,l}^3$ 代表菲涅尔衍射矩阵， $\otimes$ 代表矩阵相乘，下标 'vec' 代表将复振幅光场进行向量化，然后与菲涅尔衍射矩阵相乘得到传播后的光场 $U_{k,l}^3$ 。在输出平面 $z = z_4$ ，得到的光场是 16 个 $U_{k,l}^3$ 的叠加： $U^4 = \sum_{k,l} U_{k,l}^3$ ($k=1 \sim 4, l=1 \sim 4$)。因为最后光电探测器 CMOS 相机探测结果是光强，所以最终的输出结果是： $I = |U^4|^2$ ，代表着光场 U^4 的强度。

第二步为计算 MC-OCNN 的损失函数(Loss Function)，常规的方法是基于均方误差来计算损失函数： $L = |I - T|^2$ ，其中 I 为 MC-OCNN 输出， T 为真实标签值。然而，在复杂图形处理中，皮尔逊相关系数(PCC)函数已被证明更适合于复杂图像的重建，PCC 代表恢复图像与目标图像的相关值， $PCC = 1$ 代表相关性为 1，此时图像与目标值完全一样， $PCC = 0$ 代表相关度为 0，恢复结果与目标值相差较大。损失函数可以选用负的皮尔逊相关系数(NPCC)来表示，NPCC 函数定义为：

$$NPCC = \frac{-\sum_{i,j} [I(i,j) - \tilde{I}][T(i,j) - \tilde{T}]}{\sqrt{\sum_{i,j} [I(i,j) - \tilde{I}]^2} \cdot \sqrt{\sum_{i,j} [T(i,j) - \tilde{T}]^2}} \quad (3)$$

上式中， $\tilde{I}$ 与 $\tilde{T}$ 分别为 I 和 T 的均值。选定该损失函数后，误差光场 E^4 可根据文献[38]中的求导方式计算为： $E^4 = (\partial L / \partial I) \odot (U^4)^*$ ，其中 $\partial L / \partial I$ 表示损失函数 L 对输出 I 的梯度， $(U^4)^*$ 为光场 U^4 的共轭， $\odot$ 表示逐元素相乘。

第三步为误差光场 E^4 的反向传播，根据文献[38]中的计算方法，第三层的误差光场可计算为 $E_{k,l}^3 = \exp(iM_{k,l}^3) G_{2D,k,l}^3 \odot [(W_{k,l}^3)^T \otimes E^4]$ ，其中 $\odot$ 表示逐元素相乘， $\otimes$ 代表矩阵相乘。第二层是光学卷积

层, 其卷积核位于透镜的孔径函数中, 因此我们需要借助傅里叶变换来计算卷积核误差函数, 第二卷积层的误差光场为 $E_{k,l}^2 = M_{k,l}^2 \mathcal{F}(E_{k,l}^1)$ 。对于第一光学卷积层来说, 由于其一个通道对应着第二层的四个通道, 因此在计算误差光场时需要将前面四个通道的误差光场合并后得到: $E_k^1 = \sum_l M_k^1 \mathcal{F}[\mathcal{F}(E_{k,l}^2)]$ ($l=1 \sim 4$)。

最后一步是计算各层的相位梯度。根据文献[38]中的方法, 第一卷积层的相位梯度为 $\partial L / \partial M_k^1 = 2 \operatorname{Re}\{i \mathcal{F}(U_k^1) E_k^1\}$, 其中 $\operatorname{Re}\{\}$ 表示取复数的实部, 一共有四个梯度。第二卷积层的相位梯度为: $\partial L / \partial M_{k,l}^2 = 2 \operatorname{Re}\{i \mathcal{F}(U_{k,l}^2) E_{k,l}^2\}$, 一共有 16 个。第三全连接层的相位提取为 $\partial L / \partial M_{k,l}^3 = 2 \operatorname{Re}\{i U_{k,l}^2 E_{k,l}^3\}$ 。获得相位梯度后, 调制相位值可通过动量法(MOM)进行更新: $\mathbf{v} \leftarrow \alpha \mathbf{v} - \eta (\partial L / \partial M^N)$, $M^N \leftarrow M^N + \mathbf{v}$ ($N=1,2,3$)。其中 η 为学习率, $\alpha=0.9$ 为动量系数, $\mathbf{v}$ 为动量累积变量, η 需调整以确保相位更新步长在训练过程中保持在约 0.01 的范围内。

3. MC-OCNN 仿真实验

接下来, 我们将 MC-OCNN 框架应用于像素级语义分割任务, 包括人像解析和飞机目标的检测任务。对于人像解析任务, 图像的每个像素被标注为两类: 人物或背景, 需将人物与背景进行分离。如图 4 所示, 每张图像包含位于不同背景中的人物(Images), 其对应标签为二值掩膜(Labels), 其中背景像素用黑色表示, 人像像素用白色表示。训练与测试数据集选用 AISegment Matting Human Datasets [39] (一种专门用于人像解析与肖像分割的数据集, 包含 800×600 像素的彩色图像), 以及 Microsoft COCO [40] 中的飞机类别语义分割数据集。由于 MC-OCNN 是仅处理灰度图像的光学框架, 因此图像首先被转换为灰度图并调整尺寸至 70×70 像素, 调整后的灰度图及分割真实标签如图 4(b)所示。训练数据包含 1000 张图像, 测试数据为 100 张未见图像。

MC-OCNN 各层的仿真参数如下所示。每层像素尺寸设为 $10 \mu\text{m} \times 10 \mu\text{m}$ 。MC-OCNN 的输入尺寸为 70×70 像素, 位于平面 $z = z_0 = 0 \text{ cm}$ 处。第一光栅层为 256×256 像素, 其与第一卷积层的距离 $d_1 = 8 \text{ cm}$, 其空间频率为 $f_D^1 = \Delta_1 / \lambda d_1 = 0.0127 \mu\text{m}^{-1}$, $\Delta_1 = 64 \mu\text{m}$ 为光斑在第一个卷积层偏移的距离, 波长 $\lambda = 0.632 \mu\text{m}$ 。接着是第一个光学卷积层, 其位于平面 $z_1 = 16 \text{ cm}$ 处, 包含 256×256 像素, 分为四个区域, 每个区域包含 128×128 像素。第二个光栅层位于第一卷积层与第二卷积层之间, 其与第二卷积层的间距为: $d_2 = 16 \text{ cm}$, 分为四个区域, 其空间频率 $f_D^2 = f_D^1 / 2 = 0.0063 \mu\text{m}^{-1}$ 。第二卷积层位于平面 $z_2 = 48 \text{ cm}$ 处, 包含 16 个区域, 每个区域含有 256×256 像素, 总共是 1024×1024 个像素。第二个卷积层光栅的空间频率 $f_D^3 = f_D^2 = 0.0063 \mu\text{m}^{-1}$ 。第三个全连接层位于平面 $z_3 = 64 \text{ cm}$ 处, 间距 $d_3 = 16 \text{ cm}$, 对应光栅的空间频率为 $f_D^4 = f_D^3 = 0.0063 \mu\text{m}^{-1}$ 。最后的输出层位于平面 $z_4 = 96 \text{ cm}$ 处, 间距 $d_4 = 32 \text{ cm}$ 。由于输入是 70×70 像素, 因此输出也被设置为与输入相同的像素, 因此在输出平面我们选取 70×70 个像素, 由于传播过程中光束展宽, 因此在输出平面我们将每个像素的间距设置为 $90 \mu\text{m}$ 。

人像分割(Portrait Segmentation)的训练过程如图 4(a)所示, 数据集为 AISegment Matting Human Datasets, 可见经过 500 个 Epochs 后, 人像分割最终训练集的 PCC 值为 0.87, 测试集的 PCC 值为 0.82。重建结果显示在图 4(b)中, 第一行为原始的图像(Images), 第二行为训练的标签(Labels), 第三行是 MC-OCNN 的最终输出结果。由图可见, MC-OCNN 的输出几乎恢复了图像中的人像轮廓, 且属于人像的像素强度值显著高于背景像素。接下来, 我们将 MC-OCNN 用于对飞机目标的识别, 采用 Microsoft COCO 数据集, 训练过程如图 4(c)所示, 经过 500 个 Epochs 后, 训练集的 PCC 为 0.82, 测试集的 PCC 为 0.76。输出结果如图 4(d)所示, 可以看到, MC-OCNN 较好地分离出了图中的飞机目标。此外, 我们还对比了多通道与单通道训练的结果, 采用相同的人像数据集和网络参数, 结果如图 4(e)所示, 可以看到多通道的结果(PCC = 0.82)要好于单通道(PCC = 0.75), 因此, 相对于单通道网络, 多通道卷积神经网络具有更大的

优势。这些结果证明了 MC-OCNN 在复杂的视觉任务中比传统的单通道 ONN 更具优势。

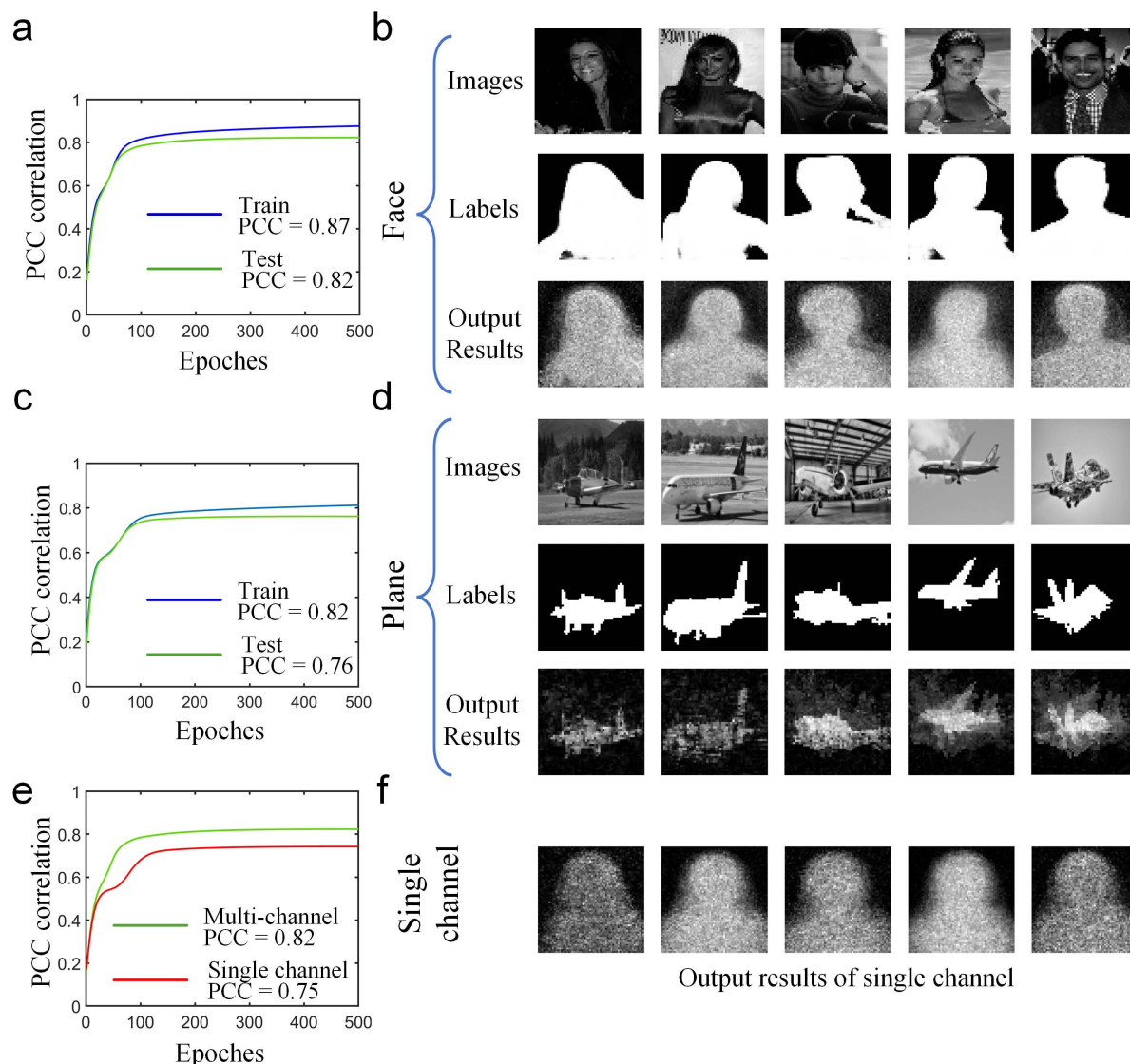


Figure 4. Semantic segmentation results for MC-OCNN: (a) MC-OCNN training process for portrait segmentation, blue line is the training set (PCC = 0.87) and green line is the test set (PCC = 0.82); (b) MC-OCNN training results for portrait segmentation, the first to third rows are the original images, training labels, and segmentation results; (c) Training process of MC-OCNN for aircraft graphic segmentation, the blue line is the training set (PCC = 0.82) and the green line is the test set (PCC = 0.76); (d) MC-OCNN training results for aircraft graphic segmentation; (e) Comparison of results for single channel (PCC = 0.75) versus multi-channel (PCC = 0.82); (f) Single channel output for portrait segmentation

图 4. MC-OCNN 的语义分割结果: (a) MC-OCNN 对人像分割的训练过程, 蓝色线为训练集(PCC = 0.87), 绿色线为测试集(PCC = 0.82); (b) MC-OCNN 对人像分割的训练结果, 第一至第三行分别为原图(Images)、训练标签(Labels), 以及分割结果(Results); (c) MC-OCNN 对飞机图形分割的训练过程, 蓝色线为训练集(PCC = 0.82), 绿色线为测试集(PCC = 0.76); (d) MC-OCNN 对飞机图形分割的训练结果; (e) 单通道(PCC = 0.75)与多通道(PCC = 0.82)的结果对比; (f) 单通道对于人像分割的输出结果

为了更全面地展示 MC-OCNN 的性能, 我们进行了消融实验, 对比了在相同数据集以及相同网络下不同参数的 MC-OCNN 的语义分割表现, 结果如图 5 所示。首先, 我们对比了不同层数的 MC-OCNN 的语义分割结果, 如图 5(a)所示, 可以看到随着层数的增加, 分割结果变得越来越好, 理论上随着层数的增

加, 分割指标可以不断上升, 但是随着层数的增加训练时间也在不断增加, 而且指标的上升幅度也会变小。其次是通道数的对比, 类似 CNNs 的通道数, 通过增加通道数, 我们可以增加特征的提取能力, 调整合适的通道数可以增强网络的表达能力, 从图中可以看出, 当网络从单通道的网络变成多通道的网络时, PCC 指标得到显著提高, 但是在多通道的基础上再进行通道数的增加提升效果将慢慢减少, 如图 5(b) 所示。接下来, 我们对比了不同像素数目对 MC-OCNN 性能的影响, 通过将网络层数设置在 3 层并且通道数设置为 16 进行对比, 如图 5(c) 所示, 经过对比发现随着每个通道像素数目的增加, 分割结果的 PCC 指标也在增加。因此, 通过增加层数、通道数, 以及像素数目, 均可以提升 MC-OCNN 的性能。最后, MC-OCNN 计算速度可以达到光速, 计算得出做一次网络计算需要的时间仅为 $t = z4/c = 3.2 \times 10^{-9}$ s, 而所需的功率仅为几毫瓦, 因此具有极高的算力以及能效比, 适用于实时探测系统中。但是由于 MC-OCNN 层数较少, 以及缺乏非线性激活层, 因此其分割结果不如成熟的传统卷积神经网络架构(U-Net 等), 但是通过引入非线性光学材料, 以及增加层数等办法可以改善其性能, 使其应用在更广泛的领域。

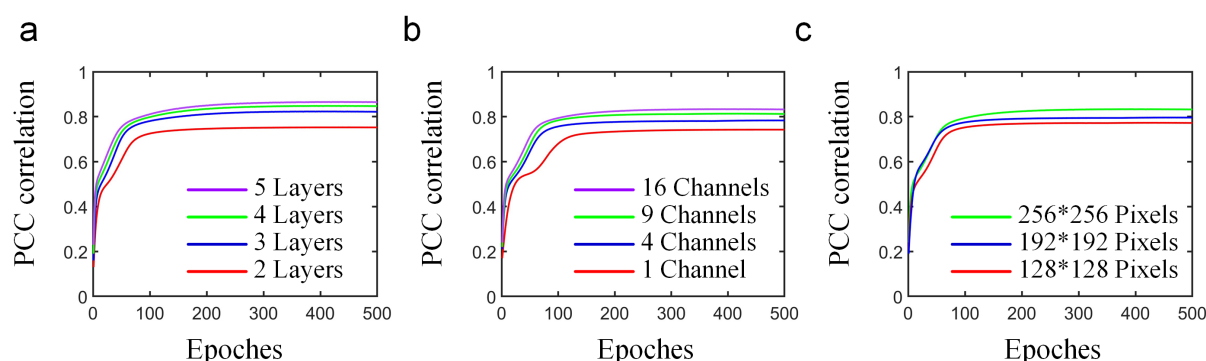


Figure 5. Comparison the results of MC-OCNN with different parameters: (a) The segment results of MC-OCNN with different number of layers; (b) The segment results of MC-OCNN with different channels; (c) The segment results of MC-OCNN with different pixel numbers

图 5. 不同参数 MC-OCNN 的性能对比结果: (a) 不同层数 MC-OCNN 的分割结果; (b) 不同通道数 MC-OCNN 的分割结果; (c) 不同像素数目 MC-OCNN 的分割结果

4. 结论

综上所述, 本文提出了一种全光的高级图像处理与计算机视觉任务的多通道光学卷积神经网络(MC-OCNN)框架。MC-OCNN 由透镜相位、随机相位, 以及二维光栅相位构建, 并且具备光学卷积层和光学全连接层。由于多通道架构, MC-OCNN 包含了多个不同的卷积核可以多维度地提取物体的不同特征, 同时基于多层次级联结构, MC-OCNN 能提取物体不同层次的特征。此外, MC-OCNN 还可以实现光速并行计算, 具备极高的算力和极低的能耗, 为光学人工智能提供了全新的方法, 文章验证了该方法在复杂像素级语义分割任务(包括人像解析与飞机目标检测)中的有效性, 证明了其在语义分割任务中可以实现良好的分割效果, 为其在机器视觉、自动驾驶、智慧医疗等领域的应用提供了基础, 展现出极具前景的应用价值与发展空间。

参考文献

- [1] Brown, T.B., Mann, B., Ryder, N., et al. (2020) Language Models Are Few-Shot Learners. Vancouver, 6-12 December 2020, *Proceedings of the 34th International Conference on Neural Information Processing Systems*, 1877-1901.
- [2] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, Munich, 5-9 October 2015, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

- [3] Yakubovskiy, R. and Morozov, Y. (2023) Speech Models Training Technologies Comparison Using Word Error Rate. *Advances in Cyber-Physical Systems*, **8**, 74-80. <https://doi.org/10.23939/acps2023.01.074>
- [4] 陈仁祥, 邱天然, 杨黎霞, 等. 改进 YOLOv7 的服务机器人密集遮挡目标识别[J]. 光学精密工程, 2024, 32(10): 1595-1605.
- [5] 程焕新, 徐皓天, 骆晓玲. 基于改进 YOLOv7 的自动驾驶目标检测方法[J]. 激光杂志, 2024, 45(7): 91-96.
- [6] Lin, X., Rivenson, Y., Yardimci, N.T., Veli, M., Luo, Y., Jarrahi, M., *et al.* (2018) All-Optical Machine Learning Using Diffractive Deep Neural Networks. *Science*, **361**, 1004-1008. <https://doi.org/10.1126/science.aat8084>
- [7] Yan, T., Wu, J., Zhou, T., Xie, H., Xu, F., Fan, J., *et al.* (2019) Fourier-Space Diffractive Deep Neural Network. *Physical Review Letters*, **123**, Article 23901. <https://doi.org/10.1103/physrevlett.123.023901>
- [8] Williamson, I.A.D., Hughes, T.W., Minkov, M., Bartlett, B., Pai, S. and Fan, S. (2020) Reprogrammable Electro-Optic Nonlinear Activation Functions for Optical Neural Networks. *IEEE Journal of Selected Topics in Quantum Electronics*, **26**, 1-12. <https://doi.org/10.1109/jstqe.2019.2930455>
- [9] Zuo, Y., Li, B., Zhao, Y., Jiang, Y., Chen, Y., Chen, P., *et al.* (2019) All-Optical Neural Network with Nonlinear Activation Functions. *Optica*, **6**, Article 1132. <https://doi.org/10.1364/optica.6.001132>
- [10] Miscuglio, M., Mehrabian, A., Hu, Z., Azzam, S.I., George, J., Kildishev, A.V., *et al.* (2018) All-Optical Nonlinear Activation Function for Photonic Neural Networks [Invited]. *Optical Materials Express*, **8**, 3851-3863. <https://doi.org/10.1364/ome.8.003851>
- [11] Chang, J., Sitzmann, V., Dun, X., Heidrich, W. and Wetzstein, G. (2018) Hybrid Optical-Electronic Convolutional Neural Networks with Optimized Diffractive Optics for Image Classification. *Scientific Reports*, **8**, Article No. 12324. <https://doi.org/10.1038/s41598-018-30619-y>
- [12] Luo, Y., Mengü, D., Yardimci, N.T., Rivenson, Y., Veli, M., Jarrahi, M., *et al.* (2019) Design of Task-Specific Optical Systems Using Broadband Diffractive Neural Networks. *Light: Science & Applications*, **8**, Article No. 112. <https://doi.org/10.1038/s41377-019-0223-1>
- [13] Wang, T., Sohoni, M.M., Wright, L.G., Onodera, T., Ma, S., Anderson, M., *et al.* (2023) Image Sensing with Multi-Layer Nonlinear Optical Neural Networks. *AI and Optical Data Sciences IV*, San Francisco, 28 January-3 February 2023. PC124380M. <https://doi.org/10.1117/12.2650289>
- [14] Li, J., Mengü, D., Luo, Y., Rivenson, Y. and Ozcan, A. (2020) Class-Specific Differential Detection in Diffractive Optical Neural Networks (Conference Presentation). *AI and Optical Data Sciences*, San Francisco, 1-6 February 2020, 112990R. <https://doi.org/10.1117/12.2544014>
- [15] Yu, Y., Cao, Y., Wang, G., Pang, Y. and Lang, L. (2023) Optical Diffractive Convolutional Neural Networks Implemented in an All-Optical Way. *Sensors*, **23**, Article 5749. <https://doi.org/10.3390/s23125749>
- [16] Sun, Y., Dong, M., Yu, M., Xia, J., Zhang, X., Bai, Y., *et al.* (2021) Nonlinear All-Optical Diffractive Deep Neural Network with 10.6 μm Wavelength for Image Classification. *International Journal of Optics*, **2021**, Article ID: 6667495. <https://doi.org/10.1155/2021/6667495>
- [17] Song, M., Li, R., Guo, R., Ding, G., Wang, Y. and Wang, J. (2022) Single Image Dehazing Algorithm Based on Optical Diffraction Deep Neural Networks. *Optics Express*, **30**, 24394-24406. <https://doi.org/10.1364/oe.458610>
- [18] Bai, B., Yang, X., Gan, T., Li, J., Mengü, D., Jarrahi, M., *et al.* (2024) Pyramid Diffractive Optical Networks for Unidirectional Image Magnification and Demagnification. *Light: Science & Applications*, **13**, Article No. 178. <https://doi.org/10.1038/s41377-024-01543-w>
- [19] Ashtiani, F., Geers, A.J. and Aflatouni, F. (2022) An On-Chip Photonic Deep Neural Network for Image Classification. *Nature*, **606**, 501-506. <https://doi.org/10.1038/s41586-022-04714-0>
- [20] Zhu, H.H., Zou, J., Zhang, H., Shi, Y.Z., Luo, S.B., Wang, N., *et al.* (2022) Space-Efficient Optical Computing with an Integrated Chip Diffractive Neural Network. *Nature Communications*, **13**, Article No. 1044. <https://doi.org/10.1038/s41467-022-28702-0>
- [21] Qu, Y., Zhu, H., Shen, Y., Zhang, J., Tao, C., Ghosh, P., *et al.* (2020) Inverse Design of an Integrated-Nanophotonics Optical Neural Network. *Science Bulletin*, **65**, 1177-1183. <https://doi.org/10.1016/j.scib.2020.03.042>
- [22] Dong, B., Aggarwal, S., Zhou, W., Ali, U.E., Farmakidis, N., Lee, J.S., *et al.* (2023) Higher-Dimensional Processing Using a Photonic Tensor Core with Continuous-Time Data. *Nature Photonics*, **17**, 1080-1088. <https://doi.org/10.1038/s41566-023-01313-x>
- [23] Al-Amri, S.S., Kalyankar, N.V. and Khamitkar, S.D. (2010) Image Segmentation by Using Edge Detection. *International Journal of Advanced Trends in Computer Science and Engineering*, **2**, 804-807.
- [24] Kaganami, H.G. and Bei, Z. (2009) Region-Based Segmentation versus Edge Detection. 2009 *Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Kyoto, 12-14 September 2009, 1217-1221. <https://doi.org/10.1109/iuh-msp.2009.13>

- [25] Zhu, S., Xia, X., Zhang, Q. and Belloulata, K. (2007) An Image Segmentation Algorithm in Image Processing Based on Threshold Segmentation. 2007 *Third International IEEE Conference on Signal-Image Technologies and Internet-Based System*, Shanghai, 16-18 December 2007, 673-678. <https://doi.org/10.1109/sitis.2007.116>
- [26] Al-Amri, S.S., Kalyankar, N.V. and Khamitkar, S.D. (2010) Image Segmentation by Using Threshold Technique. arXiv: 1005.4020. <https://doi.org/10.48550/arXiv.1005.4020>
- [27] Felzenszwalb, P.F. and Huttenlocher, D.P. (2004) Efficient Graph-Based Image Segmentation. *International Journal of Computer Vision*, **59**, 167-181. <https://doi.org/10.1023/b:visi.0000022288.19776.77>
- [28] Long, J., Shelhamer, E. and Darrell, T. (2015) Fully Convolutional Networks for Semantic Segmentation. 2015 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, 7-12 June 2015, 3431-3440. <https://doi.org/10.1109/cvpr.2015.7298965>
- [29] Girshick, R., Donahue, J., Darrell, T. and Malik, J. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 580-587. <https://doi.org/10.1109/cvpr.2014.81>
- [30] Sharma, N. and Aggarwal, L. (2010) Automated Medical Image Segmentation Techniques. *Journal of Medical Physics*, **35**, 3-14. <https://doi.org/10.4103/0971-6203.58777>
- [31] Rahman, M.M. and Marculescu, R. (2023) Medical Image Segmentation via Cascaded Attention Decoding. 2023 *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, 2-7 January 2023, 6211-6220. <https://doi.org/10.1109/wacv56688.2023.00616>
- [32] Sun, J., Chen, K., He, Z., Ren, S., He, X., Liu, X., et al. (2024) Medical Image Analysis Using Improved Sam-Med2d: Segmentation and Classification Perspectives. *BMC Medical Imaging*, **24**, Article No. 241. <https://doi.org/10.1186/s12880-024-01401-6>
- [33] Soilán, M., Riveiro, B., Martínez-Sánchez, J. and Arias, P. (2017) Segmentation and Classification of Road Markings Using MLS Data. *ISPRS Journal of Photogrammetry and Remote Sensing*, **123**, 94-103. <https://doi.org/10.1016/j.isprsjprs.2016.11.011>
- [34] Zhu, Y., Zhang, C., Zhou, D., Wang, X., Bai, X. and Liu, W. (2016) Traffic Sign Detection and Recognition Using Fully Convolutional Network Guided Proposals. *Neurocomputing*, **214**, 758-766. <https://doi.org/10.1016/j.neucom.2016.07.009>
- [35] Qin, F., Fang, B. and Zhao, H. (2010) Traffic Sign Segmentation and Recognition in Scene Images. 2010 *Chinese Conference on Pattern Recognition (CCPR)*, Chongqing, 21-23 October 2010, 1-5. <https://doi.org/10.1109/ccpr.2010.5659271>
- [36] Yuan, X., Shi, J. and Gu, L. (2021) A Review of Deep Learning Methods for Semantic Segmentation of Remote Sensing Imagery. *Expert Systems with Applications*, **169**, Article 114417. <https://doi.org/10.1016/j.eswa.2020.114417>
- [37] Basdevant, J. (2009) Le Mémoire de Fresnel sur la diffraction de la lumière. *BibNum*. <https://doi.org/10.4000/bibnum.749>
- [38] Zhou, T., Fang, L., Yan, T., Wu, J., Li, Y., Fan, J., et al. (2020) In situ Optical Backpropagation Training of Diffractive Optical Neural Networks. *Photonics Research*, **8**, 940-953. <https://doi.org/10.1364/prj.389553>
- [39] AISegment Team (2019) Human Matting Dataset (v1.2.1). AISegment. https://github.com/aisegmentcn/matting_human_datasets
- [40] Lin, T., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al. (2014) Microsoft COCO: Common Objects in Context. *Computer Vision—ECCV 2014*, Zurich, 6-12 September 2014, 740-755. https://doi.org/10.1007/978-3-319-10602-1_48