

# 基于光学神经网络的语义分割

葛泽宇<sup>1,2</sup>, 张雨超<sup>1,2\*</sup>

<sup>1</sup>上海理工大学光子芯片研究院, 上海

<sup>2</sup>上海理工大学人工智能纳米光子学中心光电信息与计算机工程学院, 上海

收稿日期: 2024年4月25日; 录用日期: 2024年5月23日; 发布日期: 2024年5月31日

## 摘要

语义分割是计算机视觉领域的关键挑战之一。目前, 语义分割主要是基于卷积神经网络(Convolutional neural network, CNN)的各种深度学习算法来实现。然而, 受限于传统的冯诺依曼架构, 随着深度学习算法复杂度的不断增加, 基于电子计算机的神经网络面临着算力, 数据吞吐率和能耗的诸多限制, 由于加工工艺已经接近了后摩尔时代的极限, 进一步提升性能变得非常困难。本文提出了一种基于光学神经网络(Optical neural network, ONN)的图像语义分割方法。光学技术具有信息计算传输速度快、信息携带能力强、抗干扰性好等特点, 在光神经网络中, 矩阵乘法以光速进行并行计算, 并且仅需极低的能量消耗, 有效解决人工神经网络中面临的复杂的矩阵乘法问题。此外, 基于存算一体架构, 光神经网络的调制层既可以存储学习的参数, 又参与了光学计算, 从而避免了计算机面临的数据读取瓶颈。我们所设计的光学神经网络在Portrait 2000数据集上进行训练和测试, 并在Person Correlation Coefficient (PCC)指标中达到0.8, 有着良好的语义分割的效果。

## 关键词

光学神经网络, 语义分割, 光学技术

# Semantic Segmentation Based on Optical Neural Network

Zeyu Ge<sup>1,2</sup>, Yuchao Zhang<sup>1,2\*</sup>

<sup>1</sup>Institute of Photonic Chips, University of Shanghai for Science and Technology, Shanghai

<sup>2</sup>Center of Artificial Intelligence Nanophotonics, School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Apr. 25<sup>th</sup>, 2024; accepted: May. 23<sup>rd</sup>, 2024; published: May. 31<sup>st</sup>, 2024

\*通讯作者。

文章引用: 葛泽宇, 张雨超. 基于光学神经网络的语义分割[J]. 建模与仿真, 2024, 13(3): 3842-3850.

DOI: 10.12677/mos.2024.133350

## Abstract

Semantic segmentation is one of the key challenges in the field of computer vision. At present, various deep learning algorithms based on convolutional neural network (CNN) are employed in this task. However, owing to the serial nature of von Neumann architectures, the electronic hardware platforms now confront significant challenges of computing speed, data throughput, and energy consumption, and the performance growth has become unsustainable due to the manufacturing process of electronic transistors approaches its physical limit. In this paper, an image semantic segmentation method based on optical neural network (ONN) is proposed. ONN has the characteristics of fast computing speed, low computing latency, and strong information carrying capacity. For the optical neural networks, matrix multiplication can be parallelly calculated at the speed of light with more energy-efficient, which effectively solves the complex matrix multiplication problem in artificial neural networks. Furthermore, as a computing in memory architecture, the modulation layers of ONN not only store the learned parameters but also participate in the optical calculation. The network is trained and tested on the Portrait 2000 dataset and achieved 0.8 in the Person Correlation Coefficient (PCC) index, showing good segmentation effect.

## Keywords

Optical Neural Network, Semantic Segmentation, Optical Technology

Copyright © 2024 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

最近, 人工神经网络(Artificial neural networks, ANN)的发展在各种机器学习应用中取得了令人瞩目的成功[1], 如医学图像诊断[2]、视频监控[3]、机器人[4]和自动驾驶[5]。然而, 随着人工智能算法越来越复杂, 传统电子硬件的性能面临着计算机体系结构、加工工艺、功耗、散热等方面的限制, 传统计算机芯片的速度提升变得越来越困难, 摩尔定律的突破也变得更加挑战性。

为了克服电子平台的缺点, 引入光学神经网络(ONN)来进行基于光子而不是电子的神经形态计算。得益于光的自然特性, ONN 具有高度并行、高速和多维度的计算能力, 同时具有大数据吞吐量和极低的能耗。据报道, 国内外已经提出了多种 ONN 架构, 例如基于马赫 - 曾德尔波导干涉仪的光学神经网络[6], 基于频率梳的集成光学神经网络[7], 以及衍射深度神经网络[8] (Diffraction Deep Neural Network,  $D^2NN$ )。其中  $D^2NN$  利用了光衍射的原理进行神经元之间的互联与计算, 可以更大规模的并行处理数据, 得到了广泛的研究和应用。 $D^2NN$  的衍射面由多层物理层组成, 一旦网络训练完成, 便可以以光速获得结果。然而, 目前的  $D^2NN$  更多的还是应用于图像分类任务, 对于处理更高级的计算机视觉任务还没有过多的研究, 在此, 我们基于  $D^2NN$  算法架构实现了其在语义分割任务中的应用。

语义分割(Image Segmentation)是一种像素级别的分类任务, 它可以实现对图片中的每一个像素分配到不同的种类, 是将标签或类别与图片中的每个像素关联的一种深度学习算法, 它用来识别构成可区分类别的像素集合。语义分割可用于自动驾驶、医学影像和工业检测等多种应用场合。早期的语义分割方法有阈值法[9]、基于直方图的捆绑[10]、区域扩展[11]、k-means 聚类[12]、分水岭[13]等, 这些方法分割

精度较低且容易受到噪声影响。随着深度学习的出现, 语义分割技术开始基于计算机架构进行, 比如卷积神经网络[14], 虽然能够高精度实现语义分割, 但其受限于电子计算机的速度和能耗。在这篇文章中, 我们基于光学神经网络实现了语义分割, 它以光速进行推理计算, 完成一次语义分割仅需 1.1 ns, 同时能耗极低, 在需要实时动态的图像处理场景中有着潜在的应用。

## 2. 网络设计

根据卷积神经网络的算法, 我们针对语义分割任务设计了光学神经网络框架, 如图 1 所示。每个衍射层包含一个可调的相位板, 相位板之间通过光的衍射实现连接。每层的相位板包含的像素为  $N \times M$ , 每个像素可以看作是一个光学神经元。对于人脑来说, 其神经元之间的联系是通过轴突来实现的, 对于神经网络来说, 后一层神经元的输入是前一层多个神经元的输出, 因此每个神经元必须有多个突触用来接收上一层的输出, 同时又必须有多个树突来将处理后的信号输入到下一层的不同神经元。在光学中, 光的衍射很好的实现神经元的这种功能, 根据惠更斯原理, 在光的传播过程中, 波面上每一个点可以看作是一个子波源, 下一个波面是前一个波面所有子波源发出的球面波的相干叠加。因此, 超表面上每一个基元可以看作由前面一层所有基元所发出球面波的相干叠加, 而这个基元又会通过发出球面子波传播到下一层的所有基元处, 从而实现类似于神经元的连接功能。

光学神经网络的连接如下图 2 所示。由相位板构成的不同层的神经元通过衍射联系在一起, 前一层的神经元通过衍射建立与下一层的神经元之间的联系。设计过程分为四个步骤: (1) 光波的前向衍射传播, (2) 损失函数的计算, (3) 光波的反向衍射传播, (4) 梯度更新。

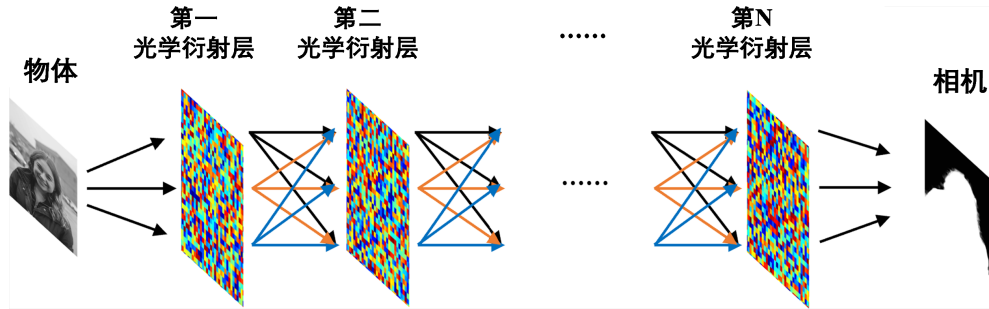


Figure 1. Propagation process of optical neural network  
图 1. 光学神经网络传播过程

步骤 1: 光的前向传播。光的前向传播基于菲涅尔基尔霍夫衍射积分公式进行计算, 建立光从第  $k-1$  层到第  $k$  层的衍射传播矩阵。光在第  $k-1$  层的复振幅分布为  $U_{k-1}(x_{k-1}, y_{k-1})$ , 在第  $k$  层的复振幅分布为  $U_k(x_k, y_k)$ , 光从第  $k$  层传播到  $k-1$  层的的衍射公式如所(1)示:

$$U_k(x_k, y_k) = \frac{\exp(j2\pi/\lambda)}{j\lambda d_k} \iint_{\Sigma} U_{k-1}(x_{k-1}, y_{k-1}) \exp\left\{j \frac{2\pi}{\lambda d_k} \left[ (x_k - x_{k-1})^2 + (y_k - y_{k-1})^2 \right]\right\} dx_{k-1} dy_{k-1} \quad (1)$$

$$W_k = \frac{\exp(j2\pi/\lambda)}{j\lambda d_k} \exp\left\{j \frac{2\pi}{\lambda d_k} \left[ (x_k - x_{k-1})^2 + (y_k - y_{k-1})^2 \right]\right\} \quad (2)$$

通过将二维光场分布  $U_{k-1}$  以及  $U_k$  向量化,  $W_k$  表示为光的前向传播矩阵, 则第  $k-1$  层神经网络与第  $k$  层神经网络所对应的关系为式:  $U_k = M_k W_k U_{k-1}$ 。上式中  $M_k$  代表第  $k$  层神经元对应的相位版的调制相位  $\exp(j\phi_k)$ 。经过多层传播后, 神经网络的最终输出的光场分布  $U_{N+1}$  如公式(3)所示, 由于探测器所探测的是光场的振幅, 所以最终的探测结果如公式(4)所示:

$$U_{N+1} = W_{N+1} \left( \prod_{k=N}^1 M_k W_k \right) U_{k-1} \quad (3)$$

$$O = |U_{N+1}|^2 \quad (4)$$

上式中,  $N$  代表神经网络的层数,  $O$  代表系统的最终输出。

**步骤 2:** 损失函数(Loss function)的计算。根据上一步所得到的系统的最终的输出函数  $O$ , 我们需要建立系统的损失函数  $L = L(O, T)$ , 其中  $T$  为输入对应的输出标签, 其中输出标签为 CCD 所对应的某一区间的光强值。其中 ONN 的损失函数可以选择不同的形式, 针对语义分割任务, 常规的均方差损失函数和交叉熵损失函数在光学神经网络中都不能达到相应的效果, 因此我们选用了皮尔逊相关系数的负数 (Negative Pearson Correlation Coefficient, NPCC) 函数作为损失函数, 它的表达式为公式(5):

$$L = \text{NPCC} = \frac{-1 \times \sum_{i=1}^w \sum_{j=1}^h (O(i, j) - \bar{O})(T(i, j) - \bar{T})}{\sqrt{\sum_{i=1}^w \sum_{j=1}^h (O(i, j) - \bar{O})^2} \sqrt{\sum_{i=1}^w \sum_{j=1}^h (T(i, j) - \bar{T})^2}} \quad (5)$$

上式中, NPCC 代表了负的 PCC, PCC 是相关系数, 它代表了神经网络的输出值与标签值之间的相关性, PCC 为 0 代表完全不相干, PCC 为 1 或者 -1, 代表完全相关或者负相关。

**步骤 3:** 误差反向传播。根据上一步所得到的系统的损失函数  $L$ , 计算光学神经网络在最后一层的误差函数为公式(6):

$$E_{N+1} = \frac{\partial L}{\partial O} \odot U_{N+1}^* \quad (6)$$

上式中,  $U_{N+1}^*$  代表  $U_{N+1}$  的共轭分布, 然后根据第  $(N+1)$  层误差函数, 通过反向传播的方法[15], 计算第  $k$  层的误差函数  $E_k$  如公式(7)所示, 第  $k$  层的相位的梯度可以通过误差函数来求得, 其表达式如(8)所示:

$$E_k = \left( \prod_{i=k+1}^N W_i^T M_i \right) W_{N+1}^T E_{N+1} \quad (7)$$

$$\frac{\partial L}{\partial \phi_k} = 2 \text{Re} \left\{ (U_k \odot E_k) \right\}^T \quad (8)$$

上式中,  $\frac{\partial L}{\partial \phi_k}$  表示相位梯度,  $W_i^T$  表示反向传播算子, 它表示  $W_i$  的转置矩阵。

**步骤 4:** 梯度更新。根据上一步得到相位梯度更新相位板的相位, 更新算法可以采用梯度下降法[16]来更新网络参数, 其公式为公式(9)所示:

$$\phi_k = \phi_k - \eta \times \frac{\partial L}{\partial \phi_k} \quad (9)$$

其中  $\eta$  代表学习率, 通常可以设置为 0.01。梯度下降算法可以在一个凸函数中寻找目标函数的最优解法, 这个最优解法也就是全局的最优解法, 对于非凸函数, 则会找到局部的最优解法。

### 3. ONN 仿真结果

我们使用香港中文大学在 2016 年发布的 Portrait Dataset 2000 数据集[17]作为训练和测试, 我们选取了训练图片包含 1000 张, 测试图片包含了 100 张, 我们将使用 PCC 作为语义分割的性能指标。数据集的部分展示如图 2 所示, 整个数据集一共有两个类别, 人物和背景。本次实验, 计算机 CPU 的型号为 12th Gen

Intel(R) Core(TM) i5-12490F 3.00 GHz, 我们基于 Numpy 工具包作为基础开展网络的设计, 使用 Python3.7 的版本。网络结构为四层衍射层, 由于光学神经网络所处理的数据图为灰度图, 因此输入彩色图片首先进行灰度化, 然后将其尺寸重新采样为  $128 \times 128$  像素。光的波长为 500 nm, 光学神经网络的相位调制层通过空间光调制器(Spatial Light Modulator, SLM)来实现, 每层的像素个数为  $256 \times 256$ , 其像素尺寸为  $8 \mu\text{m} \times 8 \mu\text{m}$ , 衍射层之间的距离为  $D$ , 我们根据在传统最大半锥衍射角理论[18]应满足以下公式(10):

$$D \geq \sqrt{N} \cdot d_f \cdot \sqrt{4d_f^2/\lambda^2 - 1} \tag{10}$$

其中  $N$  为输入图片尺寸,  $d_f$  为 SLM 单元尺寸,  $\lambda$  为波长, 在本文中  $D$  取值为 6.5 cm, 因为总共包含 4 层结构, 因此从输入到输出的总距离为  $l = 32.5\text{cm}$ , 光学计算一次所花的时间为  $t = l/c = 1.1 \text{ ns}$ 。



Figure 2. Portrait dataset  
图 2. 人像数据集

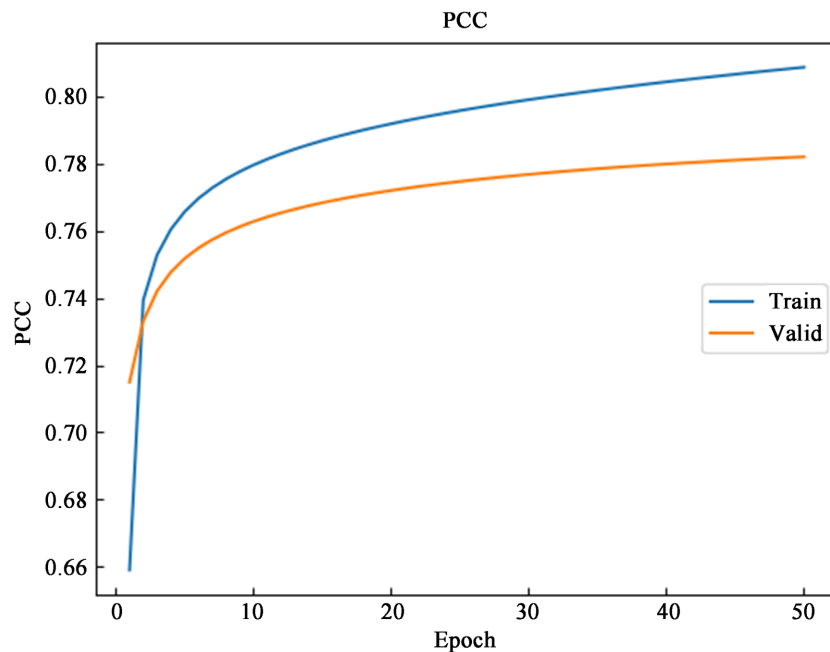


Figure 3. Training results of semantic segmentation of portrait dataset  
图 3. 人像数据集语义分割训练结果

我们针对 1000 个数据集, 训练 50 个 Epochs 后的损失函数变化如图 3 所示, 蓝色的是训练集的损失



函数随着训练过程的变化示意图, 红色的是测试集的损失函数随着训练过程的变化示意图。可以看到, 对于训练集来说, 人脸识别数据集的精度可以达到 0.8 以上。对于测试集来说, 人脸识别数据集的精度在 0.78 左右。在图 4 中给出了神经网络的输出结果, 以及其本身的标签值, 可以看到光学神经网络的输出结果与标签值类似, 准确的恢复了人脸的轮廓, 对于位于人脸的像素, 其光强值高于背景的像素, 说明了光学神经网络可以较好地实现图片中人脸信息的分割。



Figure 4. Display of the semantic segmentation effect of portrait dataset

图 4. 人像数据集语义分割效果展示图

## 4. ONN 性能分析

接下来我们对该网络进行性能分析, 通过对不同层数对网络性能的影响, 不同填充尺寸对网络性能的影响, 以及不同批次量(batchsize)对网络性能的影响进行分析。

### 4.1. 层数的变化对网络性能的影响分析

我们设置了从两层到六层的相位板进行测试, 其余条件不变, 最终结果如图 5 所示, 我们不难发现随着层数的增多, 网络的性能也在提升, 但是当网络层数提升到四层后, 从 PCC 指标上来看, 随着层数的增加, 提升的效果逐渐变小。因此我们可以猜测层数的增加确实可以提高网络的性能, 但是考虑到层数增加, 网络复杂度也在增加, 并且会增加训练时间, 并且超过四层网络, 性能提升不大, 所以我们综合考虑选用四层网络结构为最佳。

### 4.2. 输入图片填充尺寸对网络性能的影响分析

图片信息在进行传输过程中边缘信息会有所缺失, 因此我们会在图片四周填充 0 保留图片的边缘信息。我们使用的原始图片尺寸大小为  $128 * 128$ , 我们通过对分析探究填充尺寸对分割性能的影响。我们设置的填充后的尺寸有  $160 * 160$ ,  $192 * 192$ ,  $256 * 256$ ,  $320 * 320$ , 具体结果如图 6 所示。我们发现在进行少量填充如尺寸  $160 * 160$  和  $192 * 192$  的图片, 其测试效果明显偏低, 其对应的 PCC 值分别为 0.757 和 0.775。当图片填充至  $256 * 256$  时效果为最佳, PCC 的值为 0.782, 当图片填充量继续变大, 如填充至  $320 * 320$  时, PCC 的值变为 0.778, 测试效果开始降低。由此可见, 对于该光学神经网络  $128 * 128$  的输入图片, 当其填充至  $256 * 256$  时, 可以获得最佳结果, 继续填充效果会逐渐变差。

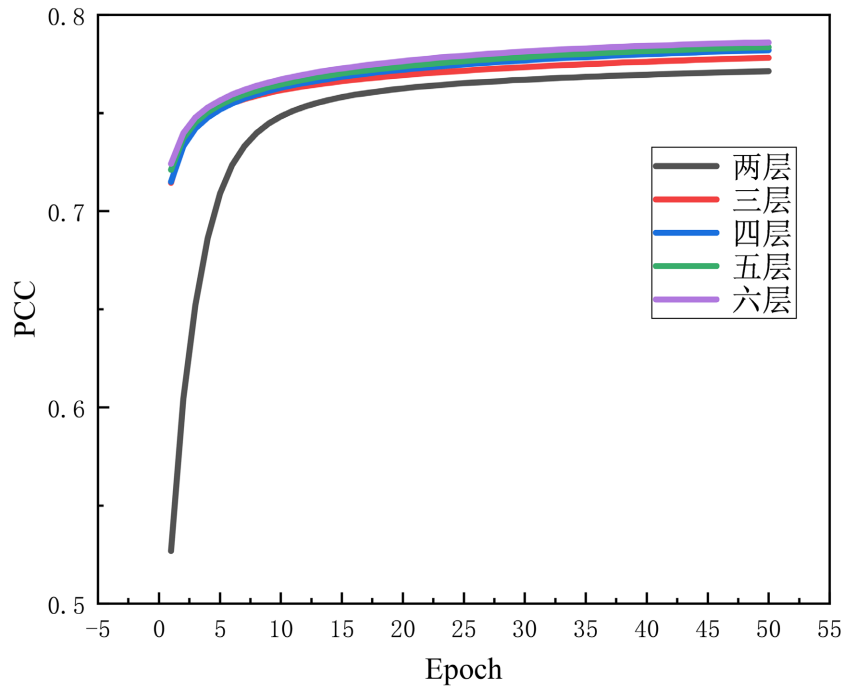


Figure 5. Comparison of network performance at different layers

图 5. 不同层数网络性能对比图

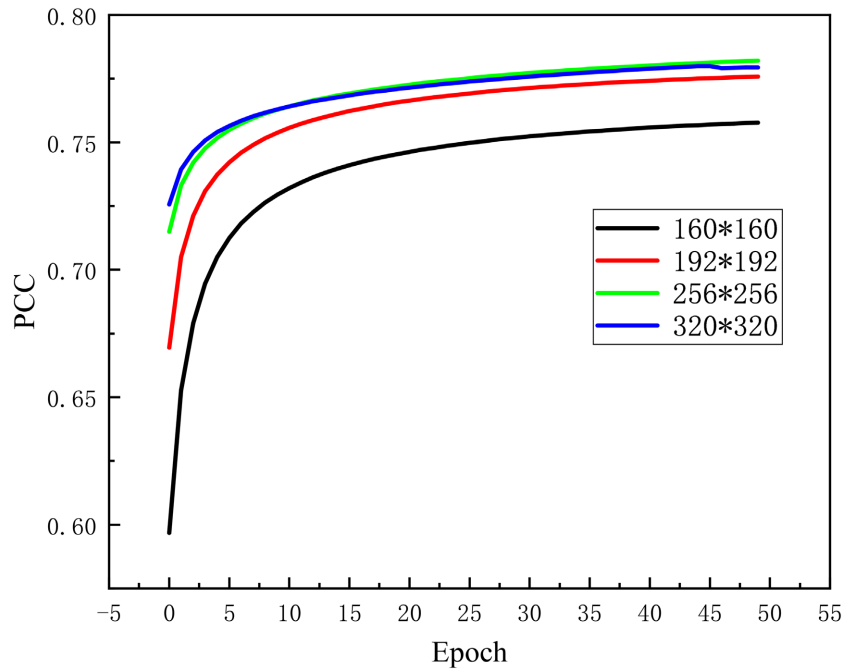


Figure 6. A comparison of the effects of different filling sizes on network performance

图 6. 不同填充尺寸对网络性能的影响对比图

### 4.3. 小样本容量对网络性能的影响分析

合适的小批量随机梯度下降算法能够提高计算效率、加快计算速度, 收敛更加彻底同时拥有很好的泛化性。因此, 我们针对不同批量的输入, 探究网络的性能。我们选取了样本容量分别为 1、5、10、20、

25, 其余条件不变。测试结果如图 7 所示, 从对比图我们可以发现, 当 batchsize 为 1 时, 出现了过拟合的现象, 而当 batchsize 取 10, 20, 25 时, 其拟合速度都随着 batchsize 数量的增加而减慢, 因此综合考虑 batchsize 选用 5 为最佳。

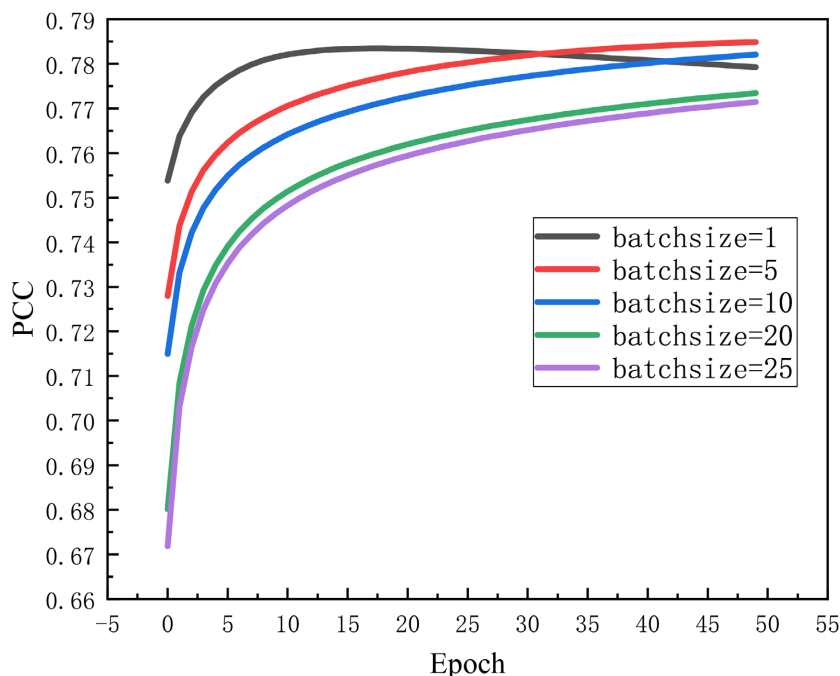


Figure 7. Comparison of network performance with different sample sizes  
图 7. 不同样本容量网络性能对比图

## 5. 总结和展望

本文基于光学神经网络成功地实现了像素级别的语义分割任务, 可以从包含人脸的图片中提取出人脸的轮廓, 并为图片中的每一个像素分配其属性, 属于人脸的像素相对于背景像素有着更高的输出光强度。通过这种光学神经网络架构, 我们解决了目前传统电学神经网络所面临的算力, 架构, 能耗, 散热等困境。经过对比测试, 我们对所设计的光学神经网络的层数, 填充像素数, 以及 batchsize 等训练参数进行了分析, 得出了当使用 4 层网络架构, 图片填充至  $256 * 256$ , 以及使用 batchsize 为 5 的组合时, 该光学神经网络可以获得最佳的语义分割效果。该光学网络在语义分割任务的实现可以推动其进一步实现更复杂的视觉学习任务。同时, 它以光速进行推理计算, 完成一次  $128 * 128$  像素图片的语义分割仅需  $1.1 \text{ ns}$ , 并且仅需极低的能耗来进行计算, 只需要普通的毫瓦激光器就可以完成光学计算, 具备极高的能量使用效率, 可以为人工智能提供新的平台。

目前, 该网络只能实现二分类的语义分割任务, 在实际生活中, 比如说自动驾驶中, 机器视觉系统需要分辨不同种类的物体, 因此实现多任务多目标的语义分割是我们下一步的工作内容。我们在之后会继续对网络进行设计, 构建多任务多目标的语义分割光学神经网络框架, 进一步填充光学神经网络在复杂视觉任务领域的空白。

## 参考文献

- [1] 高荣芳, 吉春旭, 程国建. 从浅层网络到深度网络的图像分类研究综述[J]. 计算机与数字工程, 2016, 44(10): 2020-2024.



- 
- [2] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Munich, 5-9 October 2015, 234-241. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- [3] Li, Y., Liu, Y. and Sun, Q. (2021) Real-Time Semantic Segmentation via Region and Pixel Context Network. *Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR)*, Milan, 10-15 January 2021, 7043-7049. <https://doi.org/10.1109/ICPR48806.2021.9413018>
- [4] Sourav, G., Niko, S., Feras, D., et al. (2020) Semantics for Robotic Mapping, Perception and Interaction: A Survey. <https://doi.org/10.48550/arXiv.2101.00443>
- [5] Adam, J.M., Liu, W., Zang, Y., et al. (2023) Deep Learning-Based Semantic Segmentation of Urban-Scale 3D Meshes in Remote Sensing: A Survey. *International Journal of Applied Earth Observation and Geoinformation*, **121**, Article ID: 103365. <https://doi.org/10.1016/j.jag.2023.103365>
- [6] Shen, Y., Harris, N.C., Skirlo, S., et al. (2017) Deep Learning with Coherent Nanophotonic Circuits. *Nature Photonics*, **11**, 441-446. <https://doi.org/10.1038/nphoton.2017.93>
- [7] Xu, X., Tan, M., Corcoran, B., et al. (2021) 11 TOPS Photonic Convolutional Accelerator for Optical Neural Networks. *Nature*, **589**, 44-51. <https://doi.org/10.1038/s41586-020-03063-0>
- [8] Lin, X., Rivenson, Y., Yardimci, N.T., et al. (2018) All-Optical Machine Learning Using Diffractive Deep Neural Networks. *Science*, **361**, 1004-1008. <https://doi.org/10.1126/science.aat8084>
- [9] 刘硕. 阈值分割技术发展现状综述[J]. 科技创新与应用, 2020(24): 129-130.
- [10] 李改梅, 杨润玲, 周军妮. 基于二维直方图与FCM相结合的图像快速分割方法[J]. 现代电子技术, 2007, 30(15): 176-178.
- [11] 陈方昕. 基于区域生长法的图像分割技术[J]. 科技信息, 2008(15): 58-59.
- [12] 李飞, 薛彬, 黄亚楼. 初始中心优化的K-Means聚类算法[J]. 计算机科学, 2002, 29(7): 94-96.
- [13] 赵梦琦. 基于分水岭算法的图像分割算法研究[D]: [硕士学位论文]. 西安: 西安邮电大学, 2020.
- [14] Shelhamer, E., Long, J. and Darrell, T. (2017) Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 640-651. <https://doi.org/10.1109/TPAMI.2016.2572683>
- [15] Zhou, T., Fang, L., Yan, T., et al. (2020) In situ Optical Backpropagation Training of Diffractive Optical Neural Networks. *Photonics Research*, **8**, Article 940. <https://doi.org/10.1364/PRJ.389553>
- [16] Krizhevsky, A., Sutskever, I. and Hinton, G. (2012) ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, **60**, 84-90. <https://doi.org/10.1145/3065386>
- [17] Shen, X., Tao, X., Gao, H., et al. (2016) Deep Automatic Portrait Matting. *Computer Vision—ECCV 2016*, Amsterdam, 11-14 October 2016, 92-107. [https://doi.org/10.1007/978-3-319-46448-0\\_6](https://doi.org/10.1007/978-3-319-46448-0_6)
- [18] Chen, H., Feng, J., Jiang, M., et al. (2021) Diffractive Deep Neural Networks at Visible Wavelengths. *Engineering*, **7**, 1483-1491. <https://doi.org/10.1016/j.eng.2020.07.032>