

## 深度学习在单分子定位显微镜中的应用

罗婷丹, 李依明\*

南方科技大学生物医学工程系, 广东 深圳 518055

**摘要** 单分子定位显微成像广泛应用于亚细胞结构的三维空间解析。传统的单分子定位分析方法依赖于点扩展函数(PSF)模型拟合, 制约了一些较为复杂的 PSF 模型的应用。深度学习的出现正在逐渐改变单分子定位的分析方法, 有望用来分析复杂的 PSF 模型, 并能提供新的方向和思路。介绍了深度学习在单分子定位、稀疏图像重建、PSF 设计和高内涵信息提取等方面的应用, 并介绍了目前主流的数据模拟方法。最后对深度学习在单分子定位显微成像领域内的发展方向进行了展望。

**关键词** 显微; 单分子定位显微镜; 深度学习; 超分辨; 点扩展函数

**中图分类号** O436 **文献标志码** A

**DOI:** 10.3788/CJL202249.2407206

## 1 引言

空间分辨率是光学显微镜最重要的特征, 它被定义为能分辨的两点之间的最短距离<sup>[1]</sup>。长期以来, 人们认为它受到光的衍射性质的物理限制, 这一限制也被称为阿贝衍射极限<sup>[2]</sup>。受到衍射的限制, 即使是无限小的点光源经过透镜后也只能形成有限大小的光斑, 当两个点光源距离较近时便无法将其分辨开。

为了克服这一限制, 近年来各种超分辨荧光显微技术被开发用于纳米级成像, 使得观察细胞内部单个分子之间的相互作用成为可能。出现了数种能够突破光学衍射极限达到高分辨率的成像方法, 包括结构照明显微镜(SIM)<sup>[3-4]</sup>、受激发射损耗(STED)显微镜<sup>[5]</sup>以及单分子定位显微镜(SMLM)<sup>[6-8]</sup>。其中 SMLM 由于不需要复杂的光路设计且在三者中具有最高的分辨能力而被广泛应用于细胞器以及蛋白质结构的研究中。SMLM 利用荧光物质的光开关特性, 使用点扩展函数(PSF)模型拟合数据, 以精确定位每个分子的位置, 重建高分辨率图像。

近年来随着深度学习的发展, 越来越多的研究人员将深度学习应用在 SMLM 的数据分析当中<sup>[9-11]</sup>。在数据分析中, 深度学习展现了其从复杂环境高效获取信息的优势。通过单分子显微镜进行数据采集, 使用深度学习进行数据分析, 能够提高从光子受限的图像或测量中提取信息的能力。本文将阐述深度学习在单分子定位显微技术当中的应用。在简要描述 SMLM 的技术原理和发展、介绍深度学习方法的基本组成部分和背景后, 重点讨论了基于深度学习的分析

方法在 SMLM 中的近期应用以及数据集的模拟。针对深度学习在单分子定位显微技术中的优势和限制, 对其在未来的发展进行了展望。

## 2 单分子定位显微镜

SMLM 成像的主要思想是避免衍射受限区域内多个点光源的重叠, 用激活光和激发光控制成像区每次仅有少量、随机、离散的单个荧光分子发光。将采集到的图像与 PSF 模型相拟合, 得到单个分子的质心坐标, 重建最终的高分辨率图像。与衍射受限的传统显微技术相比, SMLM 技术极大地提高了空间分辨率, 可以在分子尺度上对生物结构成像。

相比传统光学显微镜的空间分辨率, SMLM 技术使其提高了一个数量级。随着近年来技术的发展, SMLM 还可以从生物样品中获得更多维度的信息, 譬如三维空间信息<sup>[12]</sup>、多色成像<sup>[13-16]</sup>、光谱成像<sup>[17]</sup>、偏振<sup>[18-20]</sup>、单分子动态信息<sup>[21-22]</sup>和荧光寿命<sup>[23-24]</sup>等。其中三维 SMLM 成像是目前发展最多也最成熟的技术。生物结构的三维特性要求超分辨显微镜不仅在横向, 还要在轴向获得较高的分辨率。三维 SMLM 通常是利用 PSF 工程使得 PSF 在不同的轴向位置形状不同, 从而通过形状来精确定位。其中基于柱面镜的散光 PSF<sup>[12]</sup>方法由于硬件结构和数据分析算法简单, 成为了目前应用最广泛的三维超分辨技术。其他三维 SMLM 方法包括使用不同类型的工程 PSF(比如双螺旋 PSF<sup>[25]</sup>、Tetrapod PSF<sup>[26]</sup>)、双平面结构焦面成像<sup>[27]</sup>、基于单分子干涉的干涉式染料光激活定位显微镜(iPALM)成像<sup>[27-28]</sup>和基于超临界角荧光的三维成

像<sup>[29-30]</sup>等。这些技术在研究生物样品的三维空间结构方面具有强大的能力。

多色成像作为超分辨成像技术的重要延伸,极大地增强了人们研究不同生物大分子相互作用的能力。依靠不同的激发光让不同荧光分子发射出不同颜色的荧光是进行多色成像最常规的思路,利用荧光蛋白<sup>[13-14]</sup>和染料<sup>[15-16]</sup>都获得了较好的应用。2015 年 Zhang 等<sup>[17]</sup>利用三棱镜分光原理,实现了光谱分辨的随机光学重建显微成像术(SR-STORM)。2016 年 Shechtman 等<sup>[31]</sup>独辟蹊径,利用空间光调制器对不同波长荧光响应不同的特性,提出了一种基于 PSF 工程技术的双色 SMLM 成像方法。此外,随着 SMLM 的发展,人们将 SMLM 技术应用扩展至新维度,比如探究荧光偏振、分子运动等。利用 SMLM 探究单分子荧光各向异性、偏振方向等信息为分子方向和旋转运动的研究提供了有价值的见解<sup>[18-20]</sup>。SMLM 的发展还促进了单粒子跟踪技术的发展,例如通过单分子的开关特性,可以在同一视图中跟踪数千个分子<sup>[32-33]</sup>。

SMLM 技术的发展使我们能够在纳米尺度上在细胞原位生物大分子发生功能之处实现其在亚细胞器中的三维空间分布的可视化<sup>[15,34-36]</sup>。该技术还帮助我们取得了一系列重大的科学发现例如 Xu 等<sup>[15]</sup>基于三维双色 STORM 首次揭示出常规显微镜无法分辨的神经元轴突上肌动蛋白与血影蛋白交替排布形成间隔约 180 ~ 190 nm 的膜相关周期性骨架结构。在细胞核中,超分辨成像揭示了 DNA 和蛋白质的相互作用,例如发现了对 DNA 末端保护至关重要的 TRF2 依赖性端粒环的形成依赖于 TRF2 蛋白<sup>[37]</sup>。随着 SMLM 技术的发展,我们将会更好地认识纳米级别尺度的微观世界。

### 3 深度学习

人工神经网络(ANN)又名多层感知器(MLP),是一种模拟人的大脑处理信息的流程而进行计算的模型<sup>[38]</sup>,如图 1(a)所示。它由大量的人工神经元联结进行计算,能在外界信息的基础上改变内部结构,是一种自适应系统<sup>[39]</sup>。神经网络可以认为是一种非线性隐函数,常用来对输入和输出间复杂的关系进行建模。

神经网络是一种运算模型,由大量的节点(“神经元”)和节点之间的相互联结构成。每一个节点代表一种特定的输出函数,称为激活函数。每两个节点之间的联结都代表一个加权值,可以认为是神经网络的记忆。根据网络不同的输入和输出选择不同的网络结构和激活函数。在学习或训练过程中,通过计算每次输出与目标值的差异对网络中各个节点的权值和偏置进行调整,在一次次循环中降低网络输出与目标值之间的差异,不断优化网络<sup>[40]</sup>,最终得到一个具有较高准

确率的目标网络模型。

同一网络因学习方式及内容不同可具有不同的功能。神经网络的学习训练方式通常可分为两种:一种是有监督学习<sup>[41]</sup>,利用给定的样本标准进行分类或模仿,分析目标值的特征;另一种是无监督学习<sup>[42]</sup>,只规定学习方式或某些规则,具体的学习内容因系统所处环境(即输入信号情况)而异。神经网络可以不断地训练发现环境特征和规律性,在规则中找到最优策略。

传统的神经网络采用反向传播的机制进行训练,当网络层数增多时会出现梯度消失、网络无法收敛的问题<sup>[43-44]</sup>。随后出现了深度神经网络(DNN)。相比传统的 ANN,DNN 由于具有更多的网络层数,因而对事物的建模或抽象表现能力更强,也能模拟更复杂的模型。

深度学习在图像处理领域已经应用十分广泛,在如图像分割、图像识别、目标检测等方面均已应用<sup>[45]</sup>。深度学习在图像处理领域取得如此广泛的应用,主要归功于以下因素:高性能的中央处理器(CPU)和图形处理器(GPU)的发展<sup>[46-47]</sup>;可供训练的大规模数据集;深度学习算法的发展。从技术上讲,深度学习可以看作是对传统 ANN 的改进。通过多层网络进行特征提取使得可以从较低级别的特征导出较高级别的特征<sup>[48]</sup>。

目前广泛应用于图像分析的网络架构是卷积神经网络(CNN)<sup>[49]</sup>。CNN 从图像中获取输入、提取特征,并提供可学习的参数以有效地进行分类、检测和其他任务<sup>[50-51]</sup>。卷积运算的输入和输出是具有宽度、高度和深度信息(信道)的三维矢量。每个卷积都有一个由权重参数组成的小型滤波器,以滑动窗口的方式作用于输入数据,生成一个输出神经元。这保留了输入数据的宽度和高度结构。因此,通过保留数据转发路径中的图像结构,CNN 可以更容易地捕获重要的特征。

20 世纪 90 年代 LeCun 等<sup>[52]</sup>提出了一种五层的 CNN 模型[图 1(b)]。该网络将较简单的特征聚合成较复杂的特征,可成功地对手写数字进行识别<sup>[49]</sup>。基于 CNN 的 AlexNet 在计算机视觉领域最著名的分类基准 ImageNet 上取得了较好的性能<sup>[53-54]</sup>。之后又出现了一些更深层次的架构,如 ZF-Net 和 VGG (Visual Geometry Group)<sup>[52,55]</sup>。GoogLeNet 是 Google 推出的基于 Inception 模块的深度学习神经网络模型,在 2014 年的 ImageNet 竞赛中夺得了冠军,在随后的两年中持续改进,形成了 Inception v2、Inception v3、Inception v4 等版本<sup>[56-57]</sup>。Inception 类型的网络是以 Inception 模块为基础搭建的网络<sup>[58]</sup>,主要针对非均匀的稀疏数据计算问题[图 1(c)]。Inception 模块利用多尺度融合将稀疏矩阵聚类成相对密集的子矩阵,能提高计算性能。

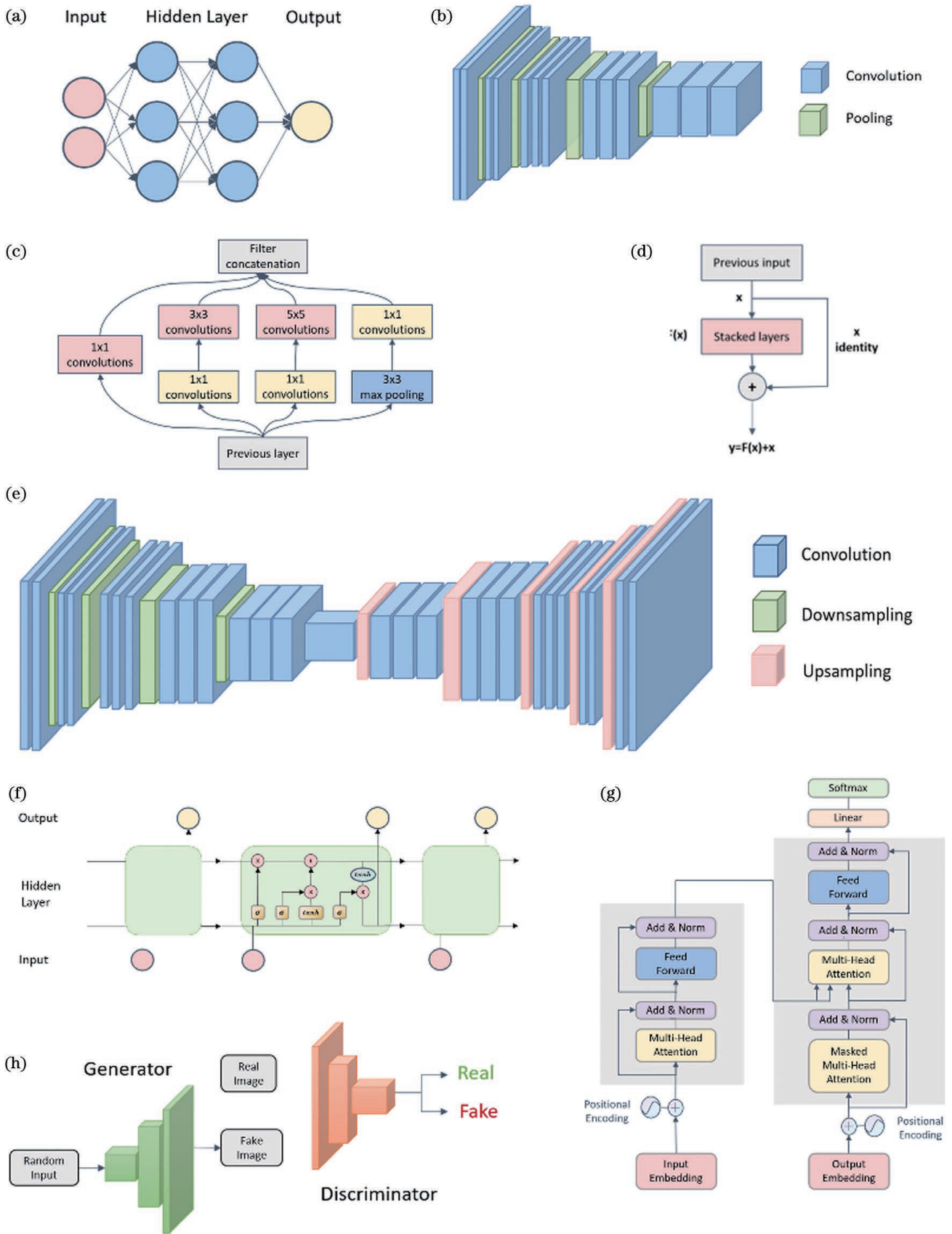


图 1 神经网络模型框架。(a)多层感知器;(b)基于卷积神经网络(CNN)的特征网络;(c) Inception 模块;(d) 残差网络 (ResNet);(e)编码器-解码器架构;(f) 长短期记忆 (LSTM)网络;(g) Transformer;(h)生成式对抗网络

Fig. 1 Neural network model framework. (a) Multilayer perceptron; (b) convolutional neural network (CNN) based feature network; (c) Inception module; (d) residual network (ResNet); (e) encoder-decoder architecture; (f) long short-term memory (LSTM) network; (g) Transformer; (h) generative adversarial network



传统的神经网络都是执行前向传播、计算差值、反向传播等操作的反复循环,直至结果收敛至目标值。但是随着网络层数加深会出现梯度消失的问题,这一点会极大地制约特征细节等信息的提取。为了解决这个问题,2016 年何恺明(He)等<sup>[56]</sup>提出了残差网络(ResNet)。其核心点在于跳跃式连接,允许残留映射[图 1(d)]。跳跃式连接将两层的输出相加,得到  $H(x) = F(x) + x$ ,  $F(x)$  调整  $x$  以减少真实值与预测值之间的残差,解决了梯度消失问题,使网络的深度不再成为制约训练的因素。

不同的应用场景会有不同的输入和不同的输出,因此针对不同的需求产生了不同的网络框架。针对图像分割问题,Badrinarayanan 等<sup>[59-60]</sup>提出了一个编码器-解码器架构。完全卷积网络(FCN)是最早采用编码器-解码器结构的网络模型<sup>[61]</sup>,它使用几个转置卷积对中间特征映射进行上采样。SegNet 和 DeconvNet 提出了一种去池化方法,编码器部分进行最大池化时记住每个最大激活值的位置<sup>[50,59]</sup>。Hourglass 网络增加了编码器和解码器组件之间的残差连接<sup>[62]</sup>。编码器-解码器架构中最为常用的模型应是 U-net<sup>[63]</sup>[图 1(e)],该模型结合了编码器-解码器和 ResNet 的优点,一方面确保了网络的输出是与输入图像具有相同空间尺度和信道深度的图像,另一方面采用 ResNet 的跳跃式连接对相同尺度图像进行连接,补偿了因采样而丢失的信息,提高了精度。

针对时间序列问题或者上下文相关问题,循环神经网络(RNN)能够捕捉到输入数据之间的依赖性。

RNN 允许将以前的输出作为具有隐藏状态的输入,它的每次输入均为网络上一次输出的结果。该方法有助于序列数据建模,主要应用于自然语言处理和语音识别领域。在标准 RNN 中经常遇到梯度消失和爆炸现象。长期短期记忆(LSTM)网络的设计是为了避免长期依赖问题<sup>[64]</sup>,如消失梯度。门控循环单元(GRU)融合了更新门和复位门两种门操作机构来解决标准 RNN 的问题<sup>[65]</sup>[图 1(f)]。2017 年 Transformer 出现,随后它迅速取代了 RNN<sup>[66]</sup>。Transformer 采用自关注机制<sup>[67]</sup>,利用差分加权输入数据各个特征的权重[图 1(g)]。Transformer 在自然语言处理中表现出出色的性能,并且最近已应用于图像处理,其效果可与 CNN 相媲美。在计算机视觉的各个领域(包括图像分类、目标检测和分割),Transformer 将会有更广泛的应用。

生成式对抗网络(GAN)<sup>[68]</sup>目前在计算机视觉方面应用广泛。GAN 由两个基本的神经网络组成,分别是生成器和判别器[图 1(h)]。GAN 受到博弈论中的零和博弈的启发,将问题视为两个网络之间的对抗和博弈。生成器从给定的噪声(通常为均匀或正态分布)中生成合成数据,判别器将生成器的输出与真实数据区分开来。目前常用该类型的网络训练得到高分辨率的图像或者进行模态转换。自 2017 年以来,GAN 的研究已经相当流行,DCGAN、CoGAN、Progressive GAN、CycleGAN 和 StyleGAN 是几种流行的架构<sup>[69-73]</sup>。

以上提及的网络和特点汇总总结如表 1 所示。

表 1 常用神经网络框架特点

Table 1 Characteristics of common neural network frameworks

| Model       | Characteristics  |
|-------------|--|
| MLP         | Fully connected class of feedforward ANN   |
| CNN         | Deep learning model based on shared-weight architecture of convolution kernels or filters.   |
| Inception   | Deep learning model solving non-uniform sparse data calculation problems by adding $1 \times 1$ convolutional layer to reduce dimensionality |
| ResNet      | Deep learning model solving vanishing/exploding gradient problems and improving classification/recognition accuracy                          |
| LSTM        | Special kind of recurrent neural network that is capable of learning long term dependencies in data  |
| Transformer | Deep learning model that adopts self-attention mechanism, differentially weighting significance of each part of input data                   |
| GAN         | Deep learning framework to train generative models via adversarial process   |

## 4 深度学习在单分子定位显微镜中的应用

近年来深度学习极大地改变了 SMLM 的数据分析方法,例如从衍射受限的图像中提取定位信息<sup>[9,74-78]</sup>,从稀疏的定位当中重建高质量的超分辨图像<sup>[11,79]</sup>,针对特定应用场景对 PSF 模型进行优化等<sup>[80-81]</sup>。接下来我们将从四个方面介绍深度学习在 SMLM 中的应用。

### 4.1 单分子定位

SMLM 的成像关键在于确保每次仅有少量、随机、离散的单个荧光分子发光。但这一原则导致较长的采集时间以及动态成像过程中较差的时间分辨率。此外,单分子成像所需的长时间采集和高强度激发光会导致强光毒性。这些限制可以通过高密度激发荧光分子进行缓解。但在这种情况下对于荧光分子的检测和精准定位都具有挑战性。Nehme 等<sup>[9]</sup>提出的 DeepSTORM 模型是第一次将深度学习应用于单分

子定位中的方法。DeepSTORM 利用编码器-解码器网络结构从由高密度的单分子图像组成的图像中获得微管图像的超分辨重建(图 2)。传统的单分子定位方法往往需要较长的计算时间并且需要对相关参数进行调整<sup>[82]</sup>,而 DeepSTORM 在具有挑战性的信噪比条件和高密度下实现了较高的分辨率,并提高了定位速度。

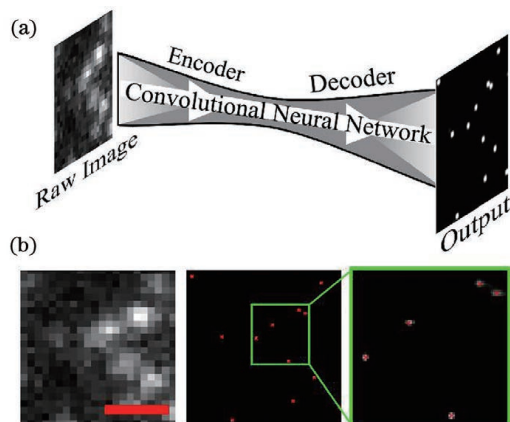


图 2 基于深度学习的单分子定位方法 DeepSTORM 以及模拟高密度数据<sup>[9]</sup>。(a) DeepSTORM 网络结构;(b)左图为低分辨率的图像;中间为 DeepSTORM 重建与真实值(红色叉号)的重合情况,右图为选定区域的放大视图

Fig. 2 Single-molecule localization method DeepSTORM based on deep learning and simulated dense data<sup>[9]</sup>. (a) Network architecture of DeepSTORM; (b) diffraction-limited low resolution image (left), DeepSTORM reconstruction with ground truth emitter positions (red crosses) (middle), and magnified view of the selected region (right)

与 DeepSTORM 不同,DeepLoco 通过模拟 PSF 来生成训练数据对网络进行训练,将单帧图像映射到定位列表,实现对二维或三维数据的快速定位<sup>[75]</sup>。DeepLoco 使用一种基于高斯核函数的损失函数,可以使网络在训练中直接定位单分子点。相比 DeepSTORM 利用网络直接获取突破衍射极限的高分辨率的图像,DeepLoco 从逆问题出发,获取单分子的点坐标,根据点坐标重建超分辨图像。这种方法可以提供更复杂的结构化输出。在速度方面,DeepLoco 相比 ThunderSTORM 的极大似然估计(MLE)拟合方法<sup>[83]</sup>提高了 3 个数量级,它能够在大约 1 s 内分析 20000 frame 3D SMLM 数据集,同时保证了一定的精度。

Zelger 等<sup>[76]</sup>利用矢量衍射理论模拟生成的训练数据,利用模拟数据训练网络,最终达到三维精准定位。该网络在保证精度的同时具有较高的速度。实验证明网络能够达到 ThunderSTORM 中 MLE<sup>[83]</sup>相同的拟合精度,同时保持较高的速率,可以达到 1 s 分析 22000 个定位点。不过该方法主要是针对分割好的单分子数据进行训练,不能对高密度的单分子进行分析。

2020 年 Nehme 等<sup>[77]</sup>在 DeepSTORM 的基础上进一步开发出能够进行三维定位的神经网络 DeepSTORM3D,并优化得到适合高密度情况的 PSF 模型。DeepSTORM3D 将物理模型融合到网络训练当中,实现端到端的优化。通过超分辨的线粒体重建和细胞中荧光标记端粒的体积成像实验证明了 DeepSTORM3D 的可行性。在网络结构上,在 DeepSTORM 编码器-解码器模型的基础上,DeepSTORM3D 增加了多尺度信息融合模块进行特征提取,并测试选取了最优的感知域进行训练。在网络中增加了一个带有可调相位的物理模拟层进行数据模拟共同学习最优的 PSF(编码)和相关的定位算法(解码)。这种方法具有极强的通用性,可以适用于不同的 3D SMLM 数据集参数(密度、信噪比和轴向位置范围)。而在此之前,通常是对 PSF 进行单独优化,再利用优化的 PSF 进行网络训练定位单分子,两个步骤是分离进行的。由 Nehme 等<sup>[77]</sup>提出的共同优化方法给后续研究提供了新的思路。

2021 年 Speiser 等<sup>[78]</sup>考虑到相邻帧中所蕴含的信息的相关性,开发出上下文依赖的深度学习定位方法 DECODE。与 DeepSTORM3D 相比,其在高密度 PSF 的定位中将定位速度提高了 10 倍、将准确率提高了 2 倍,在达到理论定位精度的同时还对定位结果做出了不确定程度预测。根据不确定程度过滤不精确的定位,进一步提高定位的准确率。该网络结构由两个堆叠的 U-net<sup>[63]</sup>组成。最终输出由多个与输入图像相同大小的通道组成,分别为定位点在该位置出现的概率、亮度、3D 坐标的偏移值、每个像素的背景强度以及定位和亮度的不确定性。这些额外信息有助于更准确地评估网络输出结果的可信度。

#### 4.2 稀疏图像重建

传统的 SMLM 方法存在低通量的缺点,需要大量精确定位的分子才能重建高质量的超分辨图像。一旦没有获得足够数量的定位,图像质量和分辨率将会下降。Ouyang 等<sup>[11]</sup>提出的 ANNA-PALM,是一种利用 GAN 网络与 U-net 网络结构相结合的网络模型<sup>[63,84]</sup>,从稀疏的定位图像快速重建超分辨视图。对微管、核孔和线粒体的模拟和实验成像表明,该方法可以在不影响空间分辨率的情况下,用少量帧重建高质量的超分辨图像。在训练过程中,ANNA-PALM 使用了数据集中部分帧的稀疏的高分辨率图像和数据集中密集的高分辨率图像。该生成器以稀疏图像作为输入,输出预测图像。判别器的输入则为生成的预测图像和密集的高分辨率图像。实验结果表明利用 ANNA-PALM 可以很好地缩短采集时间和减少样品辐照,提高空间分辨率,并大大降低漂移的不利影响。ANNA-PALM 从稀疏的图像中重建微管、线粒体和核孔的高分辨率图像与传统方法重建的图像质量相当。在实验中测试发现 ANNA-PALM 使用 7477 个



定位点进行重建,其效果与标准 PALM 使用 644844 个定位点重建的图像效果相同。在未来的实验中, ANNA-PALM 可能被用于活细胞高分辨率成像和大视场高通量成像。

Gaire 等<sup>[79]</sup>提出了另一种基于 CNN 的方法来生成高密度的多光谱超分辨图像。与 ANNA-PALM 相似,该网络同样利用低密度图像来恢复高密度图像。该方法在不改变现有光谱单分子定位显微镜 (sSMLM) 硬件系统的情况下,可以在较短的时间内生成多色超分辨图像。高质量的超分辨图像重建使用的帧数仅为通常需要的帧数的 1/8,并实现了双色和三色 sSMLM 超分辨成像。在网络模型方面, ANNA-PALM 使用 GAN 网络进行数据生成帮助重建,而

Gaire 等<sup>[79]</sup>使用 CNN 进行数据重建。对于使用者来说 CNN 的稳定性和易训练性是高于 GAN 网络的。

Zhou 等<sup>[85]</sup>在 ANNA-PALM 网络的基础上进行改造,将由 U-net 构成的生成器替换成由 ResNet 构成的生成器进行图像重建,创建了 DI-STORM(图 3)。由于 U-net 网络结构多次下采样增加了感受野导致局部信息丢失,基于 U-net 网络结构的 ANNA-PALM 生成的修复图像中存在许多图像伪影,如细丝错误连接或分裂等。相比之下,ResNet 具有较强的提取局部语义特征的能力。DI-STORM 相比传统 SMLM 在模拟数据上速度提高了 55 倍,在实验数据上速度提高了 9 倍,远高于 ANNA-PALM 相应的 17 倍和 5 倍的速度提升。

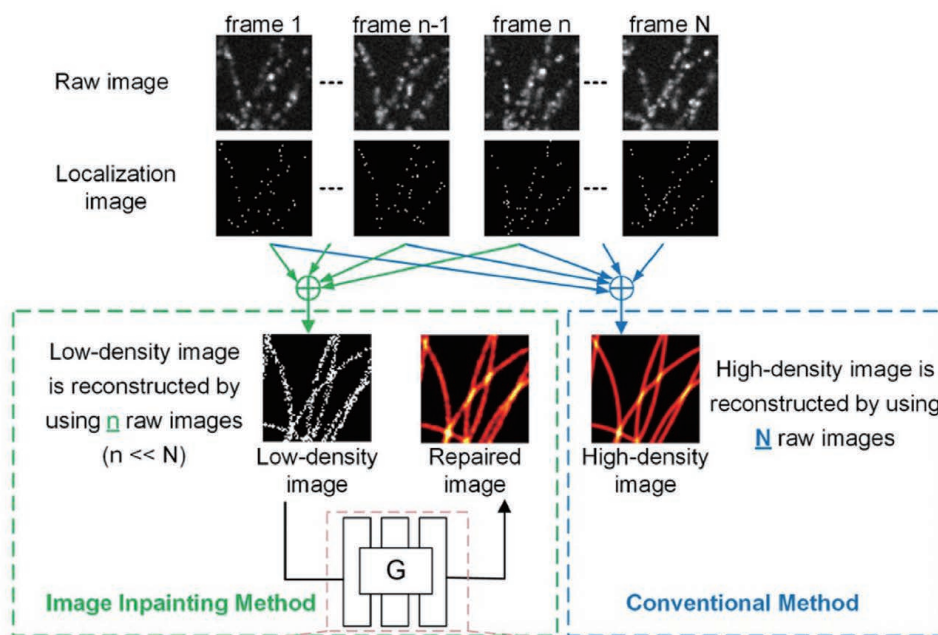


图 3 DI-STORM 的图像重建过程与传统的 SMLM 方法<sup>[85]</sup>

Fig. 3 Image reconstruction processes of DI-STORM and conventional SMLM method<sup>[85]</sup>

DNA 点累积成像技术(DNA-PAINT)是一种相对容易实现的多色超分辨成像技术。Narayanasamy 等<sup>[86]</sup>将 DeepSTORM 与 DNA-PAINT 相结合,从稀疏图像中重建高分辨率图像,实现重建加速。该方法与深度学习相结合取得较好的重建结果,其主要原因在于实验数据的密度可调整。使用实验成像数据作为模拟单分子数据的替代方法进行网络训练避免了实验数据与模拟数据的差异性。该模型具有较好的泛化性,在不同密度和结构形态下均能表现较好的效果。浙江大学 Zhu 等<sup>[87]</sup>的工作同样验证了 DNA-PAINT 与神经网络结合的优越性。他们开发的 U-PAINT 与传统方法相比,只需要原始数据的 1/10 即可实现微管的快速超分辨成像和重建。

#### 4.3 点扩展函数设计

PSF 的设计通常是通过设计不同的光瞳函数来实现某些参数指标的最优化。2014 年,Shechtman 等<sup>[26]</sup>把三维 Cramer-Rao 下界(CRLB)作为目标函数,

针对不同景深的成像,优化了一系列以泽尼克多项式为基函数的光瞳函数,比如 Saddle Point PSF<sup>[26]</sup>、Tetrapod PSF<sup>[26,88]</sup>等。然而,这种方法常常只是针对分割后的单分子数据定位精度的优化。在复杂的实际成像过程中,这些 PSF 的表现(特别是在互相重叠的单分子数据中)往往并不一样。传统的数学建模方法通常难以对这一复杂的模型直接建模并优化。深度学习的出现可以把复杂的数学模型通过数据训练的方法用神经网络进行建模,并通过损失函数来优化网络。这使得很多复杂的 PSF 设计成为可能。

Hershko 等<sup>[80]</sup>设计了一种由两部分组成的神经网络,包括空间光调制器(SLM)优化器和重构网络,用来优化一个对波长变化最敏感的 PSF(图 4)。利用相位掩模模拟显微镜的光瞳平面图像,得到了与颜色相关的 PSF。然后将其输入到重建网络中,预测原始点源在全视场图像中的位置和颜色。随着重构网络的不断优化,模拟器也将在这个相互依赖的过程中不断

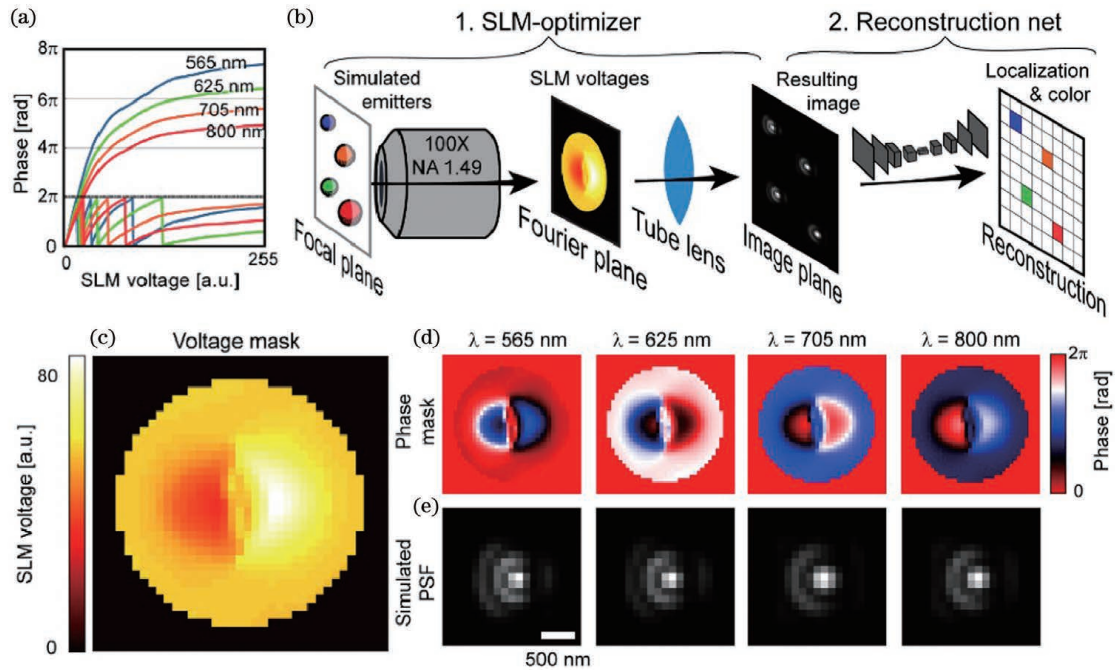


图 4 基于神经网络的多色 PSF 的优化设计<sup>[80]</sup>。(a)各波长在经过相同电压 SLM 后的相位改变量；(b)多波长 PSF 基于神经网络的优化流程；(c)优化的 SLM 电压模式用于神经网络分辨颜色；(d)在同一个 SLM 电压调制下，对 565 nm、625 nm、705 nm 和 800 nm 四种波长的荧光带来相位改变；(e)模拟的不同颜色的 PSF

Fig. 4 Design of an optimal PSF using neural networks for multicolor imaging<sup>[80]</sup>. (a) Amount of phase change of each wavelength after passing through SLMs with same voltage; (b) optimization flow of multi-wavelength PSF based on neural network; (c) optimized SLM voltage pattern for color determination by neural network; (d) phase delay for 565 nm, 625 nm, 705 nm and 800 nm light with same voltage; (e) simulated PSFs of different wavelengths

优化。Hershko 等<sup>[80]</sup>通过模拟数据证明了新设计的 PSF 可以在最小程度降低定位精度的情况下，以较高的精度恢复颜色。

定位重叠 PSF 是一个复杂的计算问题。Tetrapod PSF 最初的设计是为了在单分子拟合条件下获得最佳的三维定位精度。在实际成像过程中，Tetrapod PSF 由于其较大的尺寸，在大轴向范围内经常会重叠。Nehme 等<sup>[77]</sup>利用深度学习设计优化一个 PSF 用来定位重叠的单分子，从而达到更高的定位精度。这种优化的 PSF 与 Tetrapod PSF 相比，空间占用要小得多。Nehme 等<sup>[77]</sup>在不同密度的模拟图像上比较了优化的 PSF 和 Tetrapod PSF 的性能，结果发现与 Tetrapod PSF 相比，优化的 PSF 定位更加准确。使用荧光标记 DNA 端粒的实验数据进行比较，结果表明优化的 PSF 比 Tetrapod PSF 表现得更准确，假阳性更少，从而验证了该方法的可行性。随后，Nehme 等<sup>[81]</sup>在 DeepSTORM3D 的基础上实现双通道 PSF 端到端的优化。在数据模拟上，使用两个物理层分别对应两个通道进行模拟。实验结果表明在信号强度为 15000 光子、背景为 500 光子的条件下，优化的双通道 PSF 性能明显优于 Tetrapod PSF。

#### 4.4 点扩展函数多维度信息提取与其他应用

PSF 通常隐藏丰富的多维信息，如横向和轴向位置以及发射波长、色彩等。深度学习依赖于神经网络所提取的特征，利用深度学习可以充分提取 PSF 所包

含的信息，从而提高定位精度。

Zhang 等<sup>[74]</sup>提出一种基于 ResNet 的网络，smNet，用来提取 PSF 的三维位置、偶极子振动方向以及波前畸变。smNet 是一个 27~36 层的深度网络，由卷积层、残差块、全连通层以及批量归一化和 ReLU 单元组成。为了充分利用空间域所包含的信息，smNet 堆叠了大量的残差块，并选用较大的卷积核进行特征提取，尽可能捕获更多的信息。训练完成后，Zhang 等<sup>[74]</sup>使用模拟数据和实验数据对 smNet 的性能进行基准测试，证明了与传统的椭圆高斯拟合相比，smNet 更精确地定位了单分子的多维度信息。在损失函数设计中，他们引入了 PSF 的 CRLB<sup>[89]</sup>来加权每个参数的误差。通过该设计，smNet 在训练过程中对参数进行调优，从而能够对不同参数范围的 PSF 提取接近理论极限的信息。

PSF 不仅包含位置信息，也包含了颜色、偶极子方向和像差等其他参数。在传统的双色实验中，通过两台相机或同一相机的两个区域将收集到的荧光分成两个通道<sup>[90]</sup>。叠加两个通道的超分辨重建需要使用一种数学算法对通道进行配准，但这种配准算法的鲁棒性有限，计算量大，尤其是在三维图像中容易产生误差，从而降低图像质量。尽管可以使用相同的通道依次对两种颜色进行成像，但采集时间增加了一倍，大大降低了吞吐量。最近发表的两项研究利用训练的神经网络从它们的 PSF 形状中提取单分子的颜色<sup>[10,80]</sup>。



通过训练的网络,在同一通道中同时对两种颜色成像成为可能。

Kim 等<sup>[10]</sup>利用深度学习从 PSF 中同时预测单分子的轴向位置和发射颜色(图 5)。该网络的最终输出是由定位的结果和分色的结果共同决定的。网络输出由全连接层组成,为荧光团颜色的概率分布的估计,这使每个 PSF 图像能够以已知的置信度进行分类。它

简化了实验,不需要修改 PSF 形状,也不需要单分子荧光发射分成不同的光路。在训练数据上,由于它包含了足够的不同轴向位置的荧光团样本,因此训练神经网络可以识别所有轴向位置的荧光团。他们将轴向定位神经网络与基于三次样条插值的 PSF 模型<sup>[91]</sup>的 MLE 结果进行了比较,其结果总体上遵循 CRLB 的趋势。

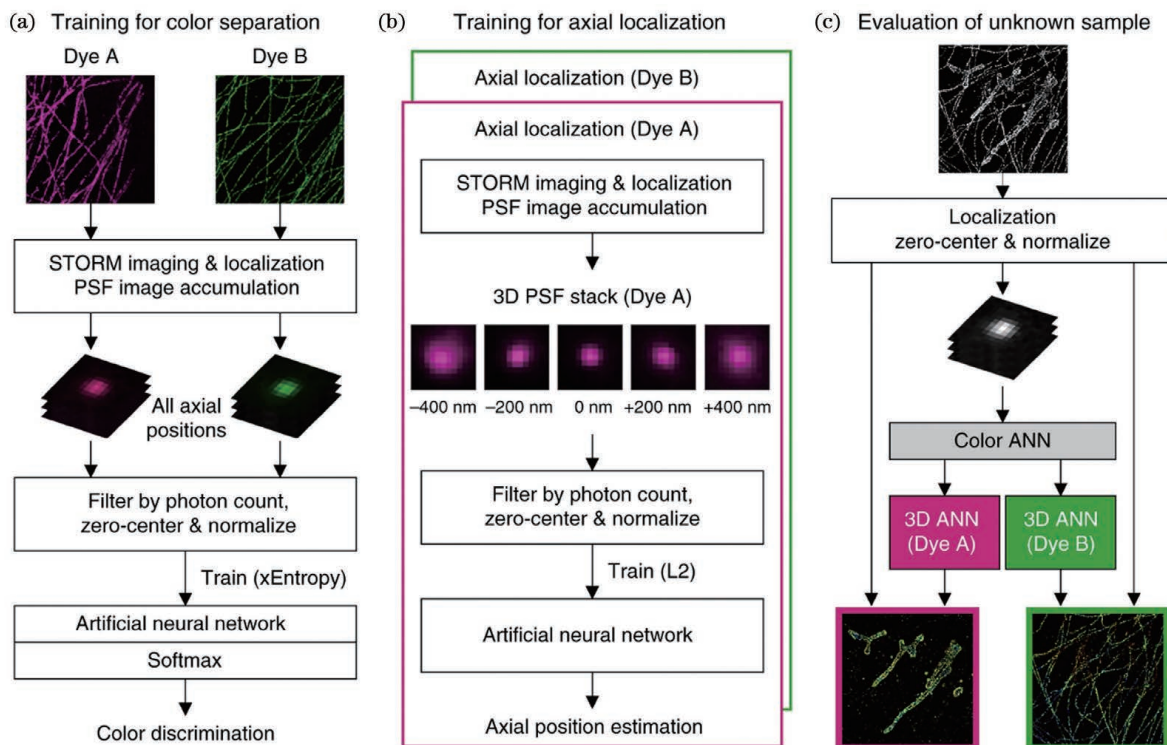


图 5 基于机器学习的多维 SMLM 的工作流程<sup>[10]</sup>。(a) 分色神经网络;(b) 解析轴向位置的神经网络;(c) 对于未知样本的分析

Fig. 5 Workflow of machine learning-based multidimensional SMLM<sup>[10]</sup>. (a) Color-separating ANN; (b) ANNs for resolving

axial position; (c) analysis of unknown samples

背景荧光是影响光学显微镜成像质量的主要因素。在进行三维定位显微镜或单分子跟踪的单分子图像分析时,背景光的危害尤其严重。Möckel 等<sup>[92]</sup>设计了一种基于 U-net 的神经网络 BGnet,对 PSF 图像进行背景校正。BGnet 是一种快速估计点源图像背景的通用方法,即使在使用 PSF 工程创建复杂的 PSF 形状时,也能以优异的精度快速估计出点源图像的背景。实验证明利用该网络进行背景校正后的 PSF 图像,无论是对仿真数据还是实验数据,都提高了定位精度。

最近开发的光谱分辨 SMLM 方法(SR-SMLM)<sup>[93]</sup>也采用了深度学习进行信息特征提取,提高分类的准确率。sSMLM 同时捕获单分子发射的空间位置和发射光谱,同时实现多色超分辨成像。现有的 SR-SMLM 依赖于提取光谱特征,例如加权光谱质心,来区分不同的分子标记。然而,完整光谱剖面所承载的丰富信息并没有得到充分利用。光谱分辨率受到探测光子数的限制<sup>[94]</sup>,只有少数定位被用于高质量的光谱分析。Zhang 等<sup>[93]</sup>通过使用全连接层的架构来对光谱进行分类,充分利用光谱中所涵盖的信息。在

他们的方法中,SR-SMLM 得到的全光谱作为网络的输入,网络的输出是一个颜色通道。Zhang 等<sup>[93]</sup>通过展示 AF647 标记的线粒体和 CF660 标记的微管的双色 STORM 图像进行验证。与之前的传统方法相比,该方法在光谱数据利用率上提高了 2 倍,错误率降至 1/10。该方法有望与荧光团设计相结合,进一步实现 SR-SMLM 多路同时成像。

## 5 单分子定位的数据模拟

机器学习技术在 SMLM 逆问题上的成功应用为 SMLM 算法提升开辟了一条新道路:开发更精确的模拟器。深度学习主要是由数据驱动进行特征提取,通过大量样本的学习能够得到深层的、数据集特定的特征表示。因此数据的多样性是极其重要的。与传统的自然图像数据集不同的是,通过模拟器可以模拟出所需要的单分子图像。

模拟器能够通过改变参数(例如噪声、相机种类、帧数、成像时间、密度、背景)来模拟 SMLM 成像<sup>[95]</sup>。通过改变这些参数可以获得更多样的数据,为网络训



练提供更多的可能场景。此外,通过模拟仿真不同情况下的图片可以用来评价不同算法对于图像的处理效果。例如 3D SMLM 挑战赛<sup>[96]</sup> 提供一个公共平台测评不同类型算法的性能。参赛者在指定数据集上进行公开测试,使用客观评价指标评估不同的 SMLM 算法

在模拟和真实参考数据集上的表现。

目前已有许多 SMLM 数据模拟器作为公开的软件工具发布。常用的 SMLM 模拟器软件包括 SuReSim<sup>[97]</sup>, TestSTORM<sup>[98]</sup> 和 SMeagol<sup>[99]</sup> (图 6)。这些模拟器的特点如表 2 所示。这些模拟器的合成数

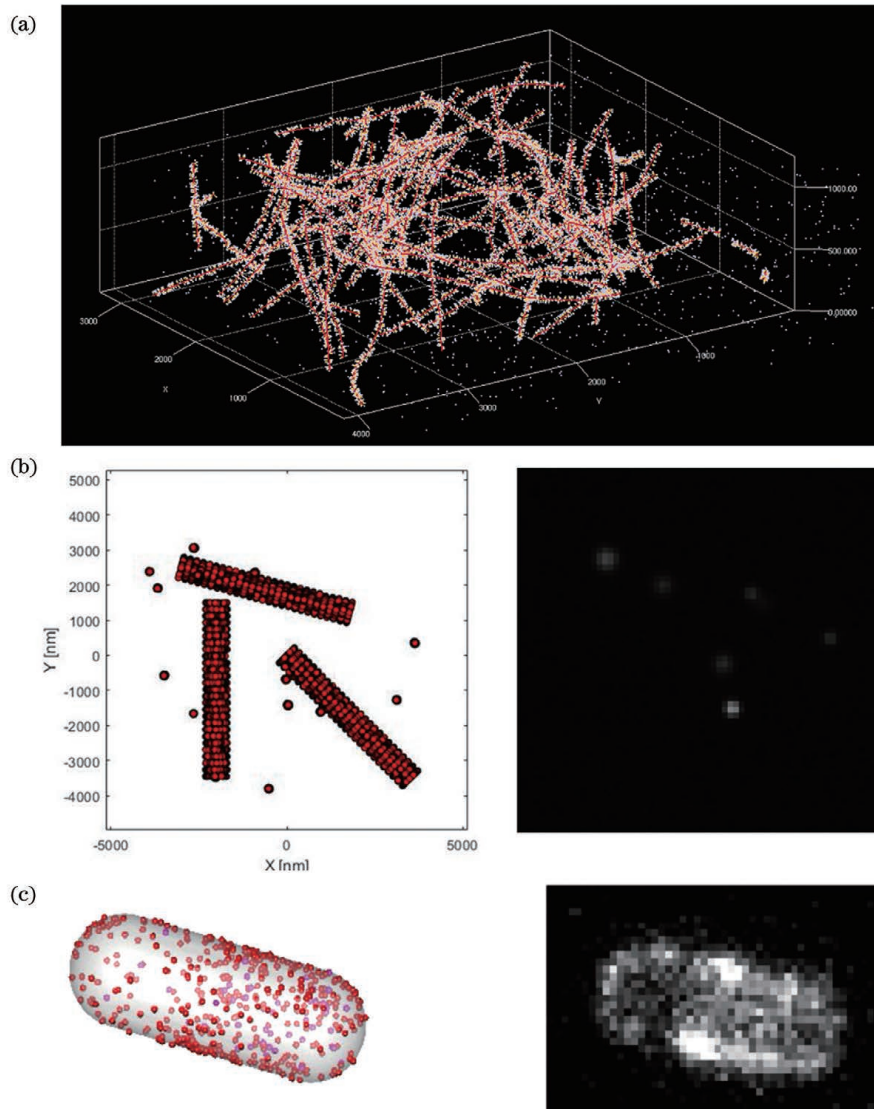


图 6 基于 SMLM 模拟器软件进行数据模拟。(a)SuReSim 对微管的三维模拟<sup>[97]</sup>; (b)左图为 TestSTORM 轴突模式模拟, 右图为生成的 tiff 图像<sup>[98]</sup>; (c)SMeagol 模拟荧光标记的 MinE 蛋白<sup>[99]</sup>

Fig. 6 Data simulation based on SMLM simulator software. (a) Three-dimensional simulation of microtubules by SuReSim<sup>[97]</sup>; (b) TestSTORM simulated axonal patterns (left) and generated TIFF image (right)<sup>[98]</sup>; (c) SMeagol simulated fluorescently labeled MinE proteins<sup>[99]</sup>

表 2 常用数据模拟方法及特点

Table 2 Common data simulation methods and characteristics

| Method    | Characteristics   |
|-----------|---|
| SuReSim   | Generate simulated localizations from ground truth models in two alternative modes; instant 3D visualization; raw SMLM data in form of image stacks   |
| TestSTORM | Generate image stacks for conventional localization microscopes with several new features such as scalar and vector diffraction based PSF models, drift, structured background, multicolor imaging, polarization sensitive excitation and detection |
| SMeagol   | Generate highly realistic single-molecule microscopy time-lapse image series, aiming primarily at single-particle tracking applications   |

据生成提供了可控制的背景创建能力,可通过密度分布和成像形状(如圆形、管状)进行控制。模拟的数据确保了数据的多样性,例如,Levet 等<sup>[100]</sup>通过改变集群的数量、相对位置、直径、密度比和背景/噪声水平来模拟多种场景,验证自身算法的鲁棒性,并通过横向对比验证自身算法的优越性。

另一方面,数据模拟器被设计用来模拟真实的标记和成像条件<sup>[101-102]</sup>。SMLM 数据模拟考虑了数据生成过程中 SMLM 成像技术的内部工作原理和标记参数(如标记效率、定位精度、帧数、每帧闪烁个数),但不直接控制结果数据。例如,Spahn 等<sup>[103]</sup>使用具有特定 SMLM 成像参数的 SuReSim 模拟器<sup>[97]</sup>生成一些视场为  $15\ \mu\text{m} \times 15\ \mu\text{m}$  的生物结构的图像。Sieben 等<sup>[104]</sup>使用了模拟器来验证他们的工作,通过控制标记效率、定位位置、噪声分子和荧光团参数(例如光子计数分布、定位精度)来生成真实模型。Cardoen 等<sup>[105]</sup>模拟常用 Alexa 647 荧光团标记的微管的图像,与真实采集的数据进行对比验证,结果表明模拟数据和真实数据在光强分布上一致,验证了模拟数据的可信性。

## 6 结束语

深度学习对计算机视觉和图像处理产生了巨大的影响,已经成为一种新兴的强大的图像分析工具。最近发展起来的超分辨显微技术与深度学习相结合,取得了较好的成果。根据输入数据和目标输出数据的类型,SMLM 图像数据分析采用了各种有效且性能良好的神经网络架构。目前,深度学习已经被广泛应用于从衍射受限的图像中提取单分子定位。与传统模型拟合方法相比,深度学习可以在保证一定准确率的情况下,大大减少处理时间。通过学习相似图像的结构冗余后,深度学习能够从稀疏的图像中重建高质量的超分辨图像。这种方法可以最大限度地减少光毒性,促进活细胞的超分辨成像。同时深度学习解决了传统方法在高密度单分子定位方面的制约,能够提高定位精度。除此之外,采用基于深度学习的方法对提取各种隐藏的分子信息也有很大的帮助,如单分子的轴向位置和发射颜色信息,这些信息可以提高最终的定位和光谱精度。在 PSF 的优化方面,深度学习也有所助益,两方面相结合可以使整个网络的解释性更强。

从 2018 年深度学习第一次应用在单分子定位当中,到如今深度学习已经渗透到 SMLM 的各个方向。对于网络模型来说,数据集、网络框架和损失函数是不可缺少的三部分。针对单分子定位与深度学习相结合,从这三个方面进行总结。

深度学习是一种依赖训练数据的方法,从大量数据中提取特征得到目标结果。例如基于稀疏图像重建的深度学习算法往往依赖于训练数据集的生物样本结构。另一方面,基于单分子的超分辨显微成像可以利用 PSF 来模拟单分子信号,可以生成几乎无限数量的

不依赖生物样本结构的模拟训练数据。这对任何类型的机器学习方法都是非常重要的资源。然而风险同样存在:网络输出可能会局限于其输入数据;生物样品背景的不均匀性、误差等其他因素导致模拟数据和真实数据存在一定差异;利用模拟数据训练得到的网络对真实数据进行预测会导致一定的误差和性能的降低。因此,在使用深度学习训练时需要考虑到数据不匹配这一问题,一方面开发出能够意识到或稳健地应对测试和训练数据之间差异性的方法,另一方面提高模拟数据与真实数据的相似性,降低误差。例如对于利用 PSF 生成的训练数据集,需要更为准确的 PSF 模型进行单分子点模拟,如 INSPR<sup>[106]</sup>。将光学知识融入到网络中进行训练也是新的途径。DeepSTORM3D 中使用了自定义的物理层,将数据模拟放入网络训练当中,将 PSF 模拟和反向传播的过程加入到整个模型的训练当中,实现了共同优化的效果。将物理模型融入到网络的训练当中可以增强网络模型的可解释性。

对于网络框架来说,目前应用在单分子定位领域的网络框架多为比较经典的稳定的网络模型,比如 U-net。深度学习在计算机视觉领域中有许多优秀的框架,将不同的网络框架应用在单分子定位领域是必然的趋势。考虑到单分子数据集的特点和网络的训练难度、参数数量以及训练时间,采用迁移学习是一种可行的方法。单分子数据集具有一定的结构相似性,在这样一个预训练的网络上再次进行训练会极大地降低训练时间和训练难度。

此外,根据数据采集特点可以选择具有时间特性的网络框架,例如 RNN 类型的网络。传统的 SMLM 主要利用荧光分子的光开关特性和稀疏成像进行重建。网络的输入数据往往是具有时间顺序的连续帧,根据这一特性,采用 RNN 类型的网络能够更好地提取特征。例如相比单帧输入,DECODE 使用连续三帧输入,能够提高定位精度。又如 Saguy 等<sup>[107]</sup>开发了一种基于 LSTM 的网络,名为 DBLink,利用上下文信息提高特征提取的能力和预测的准确性,实现动态 SMLM 数据的重建。

损失函数作为网络模型中必不可少的一部分,可以帮助网络快速收敛,获得更精准的结果。损失函数一般分为三类:目标值与预测值的差异,例如平均绝对误差、均方误差;网络结构相关的损失函数,例如 GAN 网络的对抗性损失函数;带有先验知识的特定损失函数,例如针对 PSF 优化的 CRLB 损失函数等。目前,网络模型的发展更倾向于结构化输出,即在一次前向传播中可以最大化地获取特征信息。而结构化输出意味着网络需要对多个指标进行优化,即损失函数由多个部分组成。在构建损失函数时需要考虑损失函数的组成部分是否存在相关性,能否进行简单的求和,不同部分的权重该如何选择。这些都会影响到最终网络的优化性能和训练速度。



总的来说,深度学习算法已经对许多科学研究领域产生了巨大的影响。我们期待深度学习能够与单分子定位实现更深的融合、取得更显著的成绩。

### 参 考 文 献

- [1] Zhuang X W. Nano-imaging with STORM [J]. *Nature Photonics*, 2009, 3(7): 365-367.
- [2] Abbe E. Note on the proper definition of the amplifying power of a lens or lens-system[J]. *Journal of the Royal Microscopical Society*, 1884, 4(3): 348-351.
- [3] Gustafsson M G L. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy [J]. *Journal of Microscopy*, 2000, 198(2): 82-87.
- [4] Heintzmann R, Jovin T M, Cremer C. Saturated patterned excitation microscopy: a concept for optical resolution improvement[J]. *Journal of the Optical Society of America A*, 2002, 19(8): 1599-1609.
- [5] Hell S W, Wichmann J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy [J]. *Optics Letters*, 1994, 19(11): 780-782.
- [6] Betzig E, Patterson G H, Sougrat R, et al. Imaging intracellular fluorescent proteins at nanometer resolution [J]. *Science*, 2006, 313(5793): 1642-1645.
- [7] Hess S T, Girirajan T P K, Mason M D. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy [J]. *Biophysical Journal*, 2006, 91(11): 4258-4272.
- [8] Rust M J, Bates M, Zhuang X W. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM) [J]. *Nature Methods*, 2006, 3(10): 793-796.
- [9] Nehme E, Weiss L E, Michaeli T, et al. Deep-STORM: super-resolution single-molecule microscopy by deep learning [J]. *Optica*, 2018, 5(4): 458-464.
- [10] Kim T, Moon S, Xu K. Information-rich localization microscopy through machine learning [J]. *Nature Communications*, 2019, 10(1): 1-8.
- [11] Ouyang W, Aristov A, Lelek M, et al. Deep learning massively accelerates super-resolution localization microscopy [J]. *Nature Biotechnology*, 2018, 36(5): 460-468.
- [12] Huang B, Wang W Q, Bates M, et al. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy [J]. *Science*, 2008, 319(5864): 810-813.
- [13] Shroff H, Galbraith C G, Galbraith J A, et al. Dual-color superresolution imaging of genetically expressed probes within individual adhesion complexes [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2007, 104(51): 20308-20313.
- [14] Subach F V, Patterson G H, Manley S, et al. Photoactivatable mCherry for high-resolution two-color fluorescence microscopy [J]. *Nature Methods*, 2009, 6(2): 153-159.
- [15] Xu K, Zhong G S, Zhuang X W. Actin, spectrin, and associated proteins form a periodic cytoskeletal structure in axons [J]. *Science*, 2013, 339(6118): 452-456.
- [16] Dempsey G T, Vaughan J C, Chen K H, et al. Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging [J]. *Nature Methods*, 2011, 8(12): 1027-1036.
- [17] Zhang Z Y, Kenny S J, Hauser M, et al. Ultrahigh-throughput single-molecule spectroscopy and spectrally resolved super-resolution microscopy [J]. *Nature Methods*, 2015, 12(10): 935-938.
- [18] Jameson D M, Ross J A. Fluorescence polarization/anisotropy in diagnostics and imaging [J]. *Chemical Reviews*, 2010, 110(5): 2685-2708.
- [19] Backlund M P, Lew M D, Backer A S, et al. The role of molecular dipole orientation in single-molecule fluorescence microscopy and implications for super-resolution imaging [J]. *Chemphyschem*, 2014, 15(4): 587-599.
- [20] Shroder D Y, Lippert L G, Goldman Y E. Single molecule optical measurements of orientation and rotations of biological macromolecules [J]. *Methods and Applications in Fluorescence*, 2016, 4(4): 042004.
- [21] Manzo C, Garcia-Parajo M F. A review of progress in single particle tracking: from methods to biophysical insights [J]. *Reports on Progress in Physics*, 2015, 78(12): 124601.
- [22] Kusumi A, Tsunoyama T A, Hirokawa K M, et al. Tracking single molecules at work in living cells [J]. *Nature Chemical Biology*, 2014, 10(7): 524-532.
- [23] Berezin M Y, Achilefu S. Fluorescence lifetime measurements and biological imaging [J]. *Chemical Reviews*, 2010, 110(5): 2641-2684.
- [24] Datta R, Heaster T M, Sharick J T, et al. Fluorescence lifetime imaging microscopy: fundamentals and advances in instrumentation, analysis, and applications [J]. *Journal of Biomedical Optics*, 2020, 25(7): 071203.
- [25] Pavani S R P, Thompson M A, Biteen J S, et al. Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2009, 106(9): 2995-2999.
- [26] Shechtman Y, Sahl S J, Backer A S, et al. Optimal point spread function design for 3D imaging [J]. *Physical Review Letters*, 2014, 113(13): 133902.
- [27] Juette M F, Gould T J, Lessard M D, et al. Three-dimensional sub-100 nm resolution fluorescence microscopy of thick samples [J]. *Nature Methods*, 2008, 5(6): 527-529.
- [28] Babcock H P. Multiplane and spectrally-resolved single molecule localization microscopy with industrial grade CMOS cameras [J]. *Scientific Reports*, 2018, 8: 1726.
- [29] Cabriel C, Bourg N, Jouchet P, et al. Combining 3D single molecule localization strategies for reproducible bioimaging [J]. *Nature Communications*, 2019, 10: 1980.
- [30] Dasgupta A, Deschamps J, Matti U, et al. Direct supercritical angle localization microscopy for nanometer 3D superresolution [J]. *Nature Communications*, 2021, 12: 1180.
- [31] Shechtman Y, Weiss L E, Backer A S, et al. Multicolour localization microscopy by point-spread-function engineering [J]. *Nature Photonics*, 2016, 10(9): 590-594.
- [32] Manley S, Gillette J M, Patterson G H, et al. High-density mapping of single-molecule trajectories with photoactivated localization microscopy [J]. *Nature Methods*, 2008, 5(2): 155-157.
- [33] Shim S H, Xia C L, Zhong G S, et al. Super-resolution fluorescence imaging of organelles in live cells with photoswitchable membrane probes [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2012, 109(35): 13978-13983.
- [34] Mohan N, Sorokina E M, Verdeny I V, et al. Detyrosinated microtubules spatially constrain lysosomes facilitating lysosome-autophagosome fusion [J]. *The Journal of Cell Biology*, 2019, 218(2): 632-643.
- [35] Otterstrom J, Castells-Garcia A, Vicario C, et al. Super-resolution microscopy reveals how histone tail acetylation affects DNA compaction within nucleosomes *in vivo* [J]. *Nucleic Acids Research*, 2019, 47(16): 8470-8484.
- [36] Bálint Š, Verdeny Vilanova I, Sandoval Álvarez Á, et al. Correlative live-cell and superresolution microscopy reveals cargo transport dynamics at microtubule intersections [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2013, 110(9): 3375-3380.
- [37] Doksan Y, Wu J Y, de Lange T, et al. Super-resolution fluorescence imaging of telomeres reveals TRF2-dependent T-loop formation [J]. *Cell*, 2013, 155(2): 345-356.
- [38] LeCun Y, Bengio Y, Hinton G. Deep learning [J]. *Nature*,

- 2015, 521(7553): 436-444.
- [39] Leshno M, Lin V Y, Pinkus A, et al. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function[J]. *Neural Networks*, 1993, 6(6): 861-867.
- [40] Lu Z, Pu H M, Wang F C, et al. The expressive power of neural networks: a view from the width[C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*, December 4-9, 2017, Long Beach, CA, USA. Copenhagen: MLR Press, 2017: 6232-6240.
- [41] Cunningham P, Cord M, Delany S J. Supervised learning[M]//Cord M, Cunningham P. *Machine learning techniques for multimedia*. Cognitive technologies. Heidelberg: Springer, 2008: 21-49.
- [42] Barlow H B. Unsupervised learning[J]. *Neural Computation*, 1989, 1(3): 295-311.
- [43] Griewank A. Who invented the reverse mode of differentiation [EB/OL]. [2022-03-06]. [https://www.math.uni-bielefeld.de/documenta/vol-ismp/52\\_griewank-andreas-b.pdf](https://www.math.uni-bielefeld.de/documenta/vol-ismp/52_griewank-andreas-b.pdf).
- [44] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors[J]. *Nature*, 1986, 323(6088): 533-536.
- [45] Voulodimos A, Doulamis N, Doulamis A, et al. Deep learning for computer vision: a brief review [J]. *Computational Intelligence and Neuroscience*, 2018, 2018: 7068349.
- [46] Steinkraus D, Buck I, Simard P Y. Using GPUs for machine learning algorithms [C]//*Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*, August 31-September 1, 2005, Seoul, Korea. New York: IEEE Press, 2005: 1115-1120.
- [47] Raina R, Madhavan A, Ng A Y. Large-scale deep unsupervised learning using graphics processors[C]//*ICML'09: Proceedings of the 26th Annual International Conference on Machine Learning*, June 14-18, 2009, Montreal, Quebec, Canada. New York: ACM Press, 2009: 873-880.
- [48] Widrow B, Lehr M A. 30 years of adaptive neural networks: perceptron, madaline, and backpropagation[J]. *Proceedings of the IEEE*, 1990, 78(9): 1415-1442.
- [49] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks[M]//Fleet D, Pajdla T, Schiele B, et al. *Computervision-ECCV2014. Lecture Notes in Computer Science*. Cham: Springer, 2014, 8689: 818-833.
- [50] Noh H, Hong S, Han B. Learning deconvolution network for semantic segmentation[C]//*2015 IEEE International Conference on Computer Vision*, December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1520-1528.
- [51] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [52] 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.
- [53] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [EB/OL]. (2014-09-04) [2022-02-05]. <https://arxiv.org/abs/1409.1556>.
- [54] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions[C]//*2015 IEEE Conference on Computer Vision and Pattern Recognition*, June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015.
- [55] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//*2016 IEEE Conference on Computer Vision and Pattern Recognition*, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 2818-2826.
- [56] 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*, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [57] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//*2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7132-7141.
- [58] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-ResNet and the impact of residual connections on learning[C]//*AAAI'17: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, February 4-9, 2017, San Francisco, CA, USA. Menlo Park: AAAI Press, 2017: 4278-4284.
- [59] Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(12): 2481-2495.
- [60] Kingma D P, Welling M. Auto-encoding variational Bayes [EB/OL]. (2013-12-20) [2022-03-05]. <https://arxiv.org/abs/1312.6114>.
- [61] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//*2015 IEEE Conference on Computer Vision and Pattern Recognition*, June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 3431-3440.
- [62] Newell A, Yang K Y, Deng J. Stacked hourglass networks for human pose estimation[M]//Leibe B, Matas J, Sebe N, et al. *Computervision-ECCV2016. Lecture Notes in Computer Science*. Cham: Springer, 2016, 9912: 483-499.
- [63] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[M]//Navab N, Hornegger J, Wells W M, et al. *Medical image computing and computer-assisted intervention-MICCAI 2015. Lecture Notes in Computer Science*. Cham: Springer, 2015, 9351: 234-241.
- [64] Hochreiter S, Schmidhuber J. Long short-term memory[J]. *Neural Computation*, 1997, 9(8): 1735-1780.
- [65] Cho K, van Merriënboer B, Bahdanau D, et al. On the properties of neural machine translation: encoder-decoder approaches[EB/OL]. (2014-09-03) [2022-03-06]. <https://arxiv.org/abs/1409.1259>.
- [66] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth  $16 \times 16$  words: transformers for image recognition at scale [EB/OL]. (2020-10-22) [2022-06-03]. <https://arxiv.org/abs/2010.11929>.
- [67] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [EB/OL]. (2017-06-12) [2022-03-06]. <https://arxiv.org/abs/1706.03762>.
- [68] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks [EB/OL]. (2014-06-10) [2022-06-05]. <https://arxiv.org/abs/1406.2661>.
- [69] Mirza M, Osindero S. Conditional generative adversarial nets [EB/OL]. (2014-11-06) [2022-03-06]. <https://arxiv.org/abs/1411.1784>.
- [70] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks [EB/OL]. (2015-11-19) [2022-03-06]. <https://arxiv.org/abs/1511.06434>.
- [71] Karras T, Aila T, Laine S, et al. Progressive growing of GANs for improved quality, stability, and variation [EB/OL]. (2017-10-27) [2022-03-06]. <https://arxiv.org/abs/1710.10196>.
- [72] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks [C]//*2017 IEEE International Conference on Computer Vision*, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2242-2251.
- [73] Karras T, Laine S, Aila T M. A style-based generator architecture for generative adversarial networks[C]//*2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 15-20, 2019, Long Beach, CA,



- USA. New York: IEEE Press, 2019: 4396-4405.
- [74] Zhang P Y, Liu S, Chaurasia A, et al. Analyzing complex single-molecule emission patterns with deep learning [J]. *Nature Methods*, 2018, 15(11): 913-916.
- [75] Boyd N, Jonas E, Babcock H, et al. DeepLoco: fast 3D localization microscopy using neural networks [EB/OL]. (2018-02-16) [2022-03-06]. <https://www.biorxiv.org/content/10.1101/267096v1>.
- [76] Zelger P, Kaser K, Rossboth B, et al. Three-dimensional localization microscopy using deep learning [J]. *Optics Express*, 2018, 26(25): 33166-33179.
- [77] Nehme E, Freedman D, Gordon R, et al. DeepSTORM3D: dense 3D localization microscopy and PSF design by deep learning [J]. *Nature Methods*, 2020, 17(7): 734-740.
- [78] Speiser A, Müller L R, Hoess P, et al. Deep learning enables fast and dense single-molecule localization with high accuracy [J]. *Nature Methods*, 2021, 18(9): 1082-1090.
- [79] Gaire S K, Zhang Y, Li H Y, et al. Accelerating multicolor spectroscopic single-molecule localization microscopy using deep learning [J]. *Biomedical Optics Express*, 2020, 11(5): 2705-2721.
- [80] Hershko E, Weiss L E, Michaeli T, et al. Multicolor localization microscopy and point-spread-function engineering by deep learning [J]. *Optics Express*, 2019, 27(5): 6158-6183.
- [81] Nehme E, Ferdman B, Weiss L E, et al. Learning optimal wavefront shaping for multi-channel imaging [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(7): 2179-2192.
- [82] Holden S J, Uphoff S, Kapanidis A N. DAOSTORM: an algorithm for high-density super-resolution microscopy [J]. *Nature Methods*, 2011, 8(4): 279-280.
- [83] Ovesný M, Křížek P, Borkovec J, et al. ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging [J]. *Bioinformatics*, 2014, 30(16): 2389-2390.
- [84] Isola P, Zhu J Y, Zhou T H, et al. Image-to-image translation with conditional adversarial networks [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 5967-5976.
- [85] Zhou Z W, Kuang W B, Wang Z X, et al. ResNet-based image inpainting method for enhancing the imaging speed of single molecule localization microscopy [J]. *Optics Express*, 2022, 30(18): 31766-31784.
- [86] Narayanasamy K K, Rahm J V, Tourani S, et al. Fast DNA-PAINT imaging using a deep neural network [J]. *Nature Communications*, 2022, 13: 5047.
- [87] Zhu M, Zhang L H, Jin L H, et al. DNA-PAINT imaging accelerated by machine learning [J]. *Frontiers in Chemistry*, 2022, 10: 864701.
- [88] Shechtman Y, Weiss L E, Backer A S, et al. Precise three-dimensional scan-free multiple-particle tracking over large axial ranges with tetrapod point spread functions [J]. *Nano Letters*, 2015, 15(6): 4194-4199.
- [89] Kay S M. *Fundamentals of statistical signal processing [M]*. Englewood Cliffs: Prentice-Hall PTR, 1993.
- [90] Li Y M, Shi W, Liu S, et al. Global fitting for high-accuracy multi-channel single-molecule localization [J]. *Nature Communications*, 2022, 13: 3133.
- [91] Li Y M, Mund M, Hoess P, et al. Real-time 3D single-molecule localization using experimental point spread functions [J]. *Nature Methods*, 2018, 15(5): 367-369.
- [92] Möckl L, Roy A R, Petrov P N, et al. Accurate and rapid background estimation in single-molecule localization microscopy using the deep neural network BGnet [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2020, 117(1): 60-67.
- [93] Zhang Z Y, Zhang Y, Ying L, et al. Machine-learning based spectral classification for spectroscopic single-molecule localization microscopy [J]. *Optics Letters*, 2019, 44(23): 5864-5867.
- [94] Jeong D, Kim D. Super-resolution fluorescence microscopy-based single-molecule spectroscopy [J]. *Bulletin of the Korean Chemical Society*, 2022, 43(3): 316-327.
- [95] Khater I M, Nabi I R, Hamarneh G. A review of super-resolution single-molecule localization microscopy cluster analysis and quantification methods [J]. *Patterns*, 2020, 1(3): 100038.
- [96] Sage D, Pham T A, Babcock H, et al. Super-resolution fight club: assessment of 2D and 3D single-molecule localization microscopy software [J]. *Nature Methods*, 2019, 16(5): 387-395.
- [97] Venkataramani V, Herrmannsdörfer F, Heilemann M, et al. SuReSim: simulating localization microscopy experiments from ground truth models [J]. *Nature Methods*, 2016, 13(4): 319-321.
- [98] Novák T, Gajdos T, Sinkó J, et al. TestSTORM: versatile simulator software for multimodal super-resolution localization fluorescence microscopy [J]. *Scientific Reports*, 2017, 7: 951.
- [99] Lindén M, Čurić V, Boucharin A, et al. Simulated single molecule microscopy with SMeagol [J]. *Bioinformatics*, 2016, 32(15): 2394-2395.
- [100] Levet F, Julien G, Galland R, et al. A tessellation-based colocalization analysis approach for single-molecule localization microscopy [J]. *Nature Communications*, 2019, 10(1): 2379.
- [101] Culley S, Albrecht D, Jacobs C, et al. Quantitative mapping and minimization of super-resolution optical imaging artifacts [J]. *Nature Methods*, 2018, 15(4): 263-266.
- [102] Marsh R J, Pfisterer K, Bennett P, et al. Artifact-free high-density localization microscopy analysis [J]. *Nature Methods*, 2018, 15(9): 689-692.
- [103] Spahn C, Herrmannsdörfer F, Kuner T, et al. Temporal accumulation analysis provides simplified artifact-free analysis of membrane-protein nanoclusters [J]. *Nature Methods*, 2016, 13(12): 963-964.
- [104] Sieben C, Banterle N, Douglass K M, et al. Multicolor single-particle reconstruction of protein complexes [J]. *Nature Methods*, 2018, 15(10): 777-780.
- [105] Cardoen B, Yedder H B, Sharma A, et al. ERGO: efficient recurrent graph optimized emitter density estimation in single molecule localization microscopy [J]. *IEEE Transactions on Medical Imaging*, 2020, 39(6): 1942-1956.
- [106] Xu F, Ma D H, MacPherson K P, et al. Three-dimensional nanoscopy of whole cells and tissues with *in situ* point spread function retrieval [J]. *Nature Methods*, 2020, 17(5): 531-540.
- [107] Saguy A, Alalouf O, Opatovski N, et al. DBlink: Dynamic localization microscopy in super spatiotemporal resolution via deep learning [EB/OL]. (2022-07-03) [2022-07-08] DOI: 10.1101/2022.07.01.498428.

# Deep Learning in Single-Molecule Localization Microscopy

Luo Tingdan, Li Yiming\*

*Department of Biomedical Engineering, Southern University of Science and Technology, Shenzhen 518055, Guangdong, China*

## Abstract

**Significance** Single-molecule localization microscopy (SMLM) is widely used for 3D spatial resolution of subcellular structures. Traditional single-molecule localization analysis methods rely on point spread function (PSF) model fitting, which restricts the application of some more complex PSF models. The emergence of deep learning is gradually changing the analysis methods of single-molecule localization, which is expected to be used to analyze complex PSF models and can provide new directions and ideas.

**Progress** To break through the diffraction limit, various super-resolution fluorescence microscopy techniques have been developed for nanoscale imaging in recent years, making it possible to observe the interactions between individual molecules inside cells. Several imaging methods have emerged to achieve high resolution beyond the optical diffraction limit, including structured illumination microscopy (SIM), stimulated emission depletion (STED) microscopy, and SMLM. Among them, SMLM is widely used in the study of organelles and protein structures. The main idea of SMLM imaging is to avoid the overlapping of multiple point light sources in the diffraction-limited area, and control the imaging area with activation light and excitation light to emit only a small number of random and discrete single fluorescent molecules each time.

With the development of deep learning, a large number of excellent network frameworks have emerged, which have achieved good results in the field of computer vision, for example, convolutional neural network (CNN) widely used in feature extraction, recurrent neural network (RNN) for time series data extraction, generative adversarial network (GAN) for generating data, and so on. Developed super-resolution microscopy techniques combined with deep learning have achieved good results. Depending on the type of input data and target output data, SMLM image data analysis employs various efficient and well-performing neural network architectures. Currently, deep learning has been widely used to extract single-molecule localization from diffraction-limited images. Compared with traditional model fitting methods, deep learning can greatly reduce the processing time while ensuring a certain accuracy. By learning the structural redundancy of similar images, deep learning can reconstruct high-quality super-resolution images from sparse images. This approach can minimize phototoxicity and facilitate super-resolution imaging of live cells. Deep learning solves the limitation of traditional methods in high-density single-molecule localization and improves the localization accuracy. Additionally, employing deep learning-based methods is of great help in extracting various hidden molecular information, such as single-molecule axial position and emission color information, which can improve the localization and spectral accuracy.

For specific application scenarios, deep learning is used to model complex mathematical models by data training method, and the network is optimized by loss function, which makes many complex PSF designs possible. PSFs usually contain rich multidimensional information, such as transverse and axial position, emission wavelength, color and so on. Using deep learning can fully extract the information contained in PSFs, thereby improving localization accuracy.

Deep learning is a data-driven feature extraction method. By learning a large number of samples, a deep and specific feature representation of the dataset can be obtained. So diversity of data is extremely important. Different from the traditional natural image dataset, the desired single-molecule image can be simulated by the simulator. The emergence of a large number of excellent single-molecule data simulators also provides a favorable tool for the combination of deep learning and single-molecule localization.

**Conclusions and Prospects** Deep learning has had a huge impact on computer vision and image processing, and has become an emerging and powerful tool for image analysis. Recently developed super-resolution microscopy techniques combined with deep learning have achieved good results. Depending on the type of input data and target output data, SMLM image data analysis employs various efficient and well-performing neural network architectures. For the combination of single-molecule localization and deep learning, we analyze it from three aspects: dataset, network framework, and loss function.

Deep learning is a method that relies on training data to extract features from a large amount of data to obtain target results. For example, deep learning algorithms based on sparse image reconstruction often rely on the biological sample structure of the training dataset. Data mismatches need to be taken into account when training with deep learning, while



developing methods that are aware or robust to differences between test and training data is needed. On the other hand, the similarity between the data simulation and the real data is improved, and the error is reduced. Considering the characteristics of single-molecule datasets and the training difficulty, number of parameters and training time of the network, transfer learning is a feasible method. Considering the structural similarity of the single-molecule datasets, retraining on the pretrained network will greatly reduce the training time and training difficulty. In addition to the feature of structural similarity, the input of the network is a continuous frame with time sequence. According to this feature, the RNN-type network can better extract features. For the loss function, we need to consider these issues. When constructing the loss function, we need to consider whether the components of the loss function are correlated, whether a simple summation can be performed, and how to choose the weights of different parts. Deep learning algorithms have had a huge impact on many fields of scientific research. We expect deep learning to be more deeply integrated with single-molecule localization and achieve more remarkable achievements.

**Key words** microscopy; single-molecule localization microscopy; deep learning; super resolution; point spread function