arXiv:2301.02757v1 [physics.ins-det] 7 Jan 2023

利用神经风格迁移模拟非理想仪器行为以处理全息图

John S. Schreck[1], Matthew Hayman[2], Gabrielle Gantos[1],Aaron Bansemer[3],和 David John Gagne[1]1美国国家大气研究中心（NCAR），计算与信息系统实验室，博尔德，科罗拉多州，美国2美国国家大气研究中心（NCAR），地球观测实验室，博尔德，科罗拉多州，美国3美国国家大气研究中心（NCAR），中尺度与微尺度气象实验室，博尔德，科罗拉多州，美国

通信作者： schreck@ucar.edu; mhayman@ucar.edu

摘要。

全息云探测仪能够提供前所未有的云粒子密度、尺寸和位置信息。每次激光脉冲都会捕获大体积范围内的粒子，通过计算方法对图像重新聚焦，即可确定粒子的尺寸和形状。然而，无论采用标准方法还是机器学习（ML）模型处理这些全息图，都需要大量计算资源、时间以及偶尔的人工干预。当前的机器学习模型是在基于探测仪物理模型获得的模拟全息图上训练的，因为真实全息图没有绝对的真实标签。任何试图使用其他处理方法生成标签的做法都可能引入错误，而这些错误随后会被机器学习模型继承。只有在训练期间对模拟图像进行图像损坏，以此模拟实际探测仪中的非理想条件时，训练好的模型才能在真实全息图上表现良好（Schreck 等，2022年）。优化图像损坏过程需要耗费大量精力进行人工标注。

本文展示了将神经风格迁移方法（Gatys 等，2016年）应用于模拟全息图的可行性。通过使用标准的预训练卷积神经网络（VGG-19），我们将模拟全息图进行“风格化”，使其呈现出与探针采集的真实全息图相似的外观，同时保留模拟图像的“内容”（例如粒子的位置和大小）。多项图像相似度指标一致表明，风格化图像比合成图像更接近真实全息图。使用风格化数据集训练机器学习模型来预测粒子的位置和形状后，我们发现该模型在模拟和真实全息图上的表现相当，因此无需进行人工标注。本文所述方法并不局限于全息图像，也可应用于其他领域，用来捕捉观测仪器中的噪声和瑕疵，使模拟数据更接近真实世界的观测结果。

1 引言

全息成像能够在较大的采样体积内对粒子位置和尺寸进行三维重建。通过这种技术，可以在单个瞬时体积中捕获并表征数千个粒子组成的粒子群。这项技术已成功应用于利用机载原位仪器表征云的微物理特性，例如

HOLODEC 探测器（Fugal 等，2004；Spuler 和 Fugal，2011），该探测器由美国国家大气研究中心（NCAR）运行，并已成功部署于多个野外观测活动中。

HOLODEC 仪器可捕获 15 cm^3 采样体积的内联全息图。波长为 355 nm 的激光器在两个探测臂之间发射扩束光束。光线穿过采样体积后，会成像到电荷耦合器件上，其有效像素分辨率为 3 μm。激光器以 3 Hz 的重复频率发射，因此飞机穿过云层时，可对粒径分布进行点状采集；这些采集点的空间间隔由飞机速度决定。

虽然全息成像能够在单张图像中捕获大量信息，但处理这些全息图往往需要大量计算，通常还需要人工介入来调整处理方法。因此，研究人员已开展大量研究，探索如何利用机器学习加速或改进全息粒子数据的处理（Shimob- aba 等，2019年；Shao 等，2020年；Zhang 等，2022年）。然而，这些研究中的许多都聚焦于深度分量相对较小的图像，这与 HOLODEC 的需求不同。另一个关键区别在于，待处理的全息图往往过于理想。在受控实验室环境中或使用模拟全息图表现良好的处理方案，在 HOLODEC 实际运行采集数据上的表现通常不尽如人意。对于安装在飞机外部的仪器，不能指望它始终获得同样清晰、优质的图像，因为各种运行和环境条件都可能导致捕获的图像中存在大量瑕疵。处理 HOLODEC 数据所面临的挑战，要求调整滤波器，以尽可能减少假阳性，同时避免漏检全息图中真实存在的粒子，即使这些全息图存在不可忽视的瑕疵。

这些瑕疵难以在模拟中建模，因为它们往往取决于仪器中多种部件瑕疵的组合。例如，透射光学元件产生的少量反射会在重建中形成调制图样和聚焦光斑；激光模式的振幅和相位会偏离理想高斯分布；渐晕会导致视场响应不均匀，且视场响应还取决于光束指向，而光束指向会随振动和温度变化。归根结底，开发能够补偿这些以及其他（有时未知的）影响的处理技术，对于获得高质量的科学观测至关重要。

在先前的工作中，我们提出了一种处理流程概念，通过结合硬件加速与机器学习来加速并改进 HOLODEC 图像的处理（Schreck 等，2022）。与大气传感领域的常见情况一样，实际全息图数据缺少真实标签。为了使处理能力优于现有方案，我们开发了一种方法，通过对模拟全息图进行损坏来模拟实际 HOLODEC 数据的瑕疵。我们使用模拟全息图（其标签的准确度近乎绝对）训练机器学习方案，以处理实际 HOLODEC 数据；训练时添加随机噪声、应用模糊和倾斜变形，并调整图像对比度。尽管模拟训练数据在视觉上与实际 HOLODEC 数据并不相似，但这些图像损坏足以训练出一个模型，其表现仍优于标准处理方案中使用的粒子识别方法（Schreck 等，2022）。然而，要取得成功，必须通过超参数优化来调整每一种变换，这需要人工标注数千张重建全息图。这是个棘手的问题，因为人工标注需要训练有素的人员投入大量人时，而且（正如我们在过程中发现的）人工判定的标签存在一定误差，尽管误差非零，却无法量化。此外，在优化过程中，我们只是猜测应该对图像施加哪些类型的损坏。最后，可能存在许多传感器情境，简单的损坏操作并不足以开发出性能更高的处理方案。

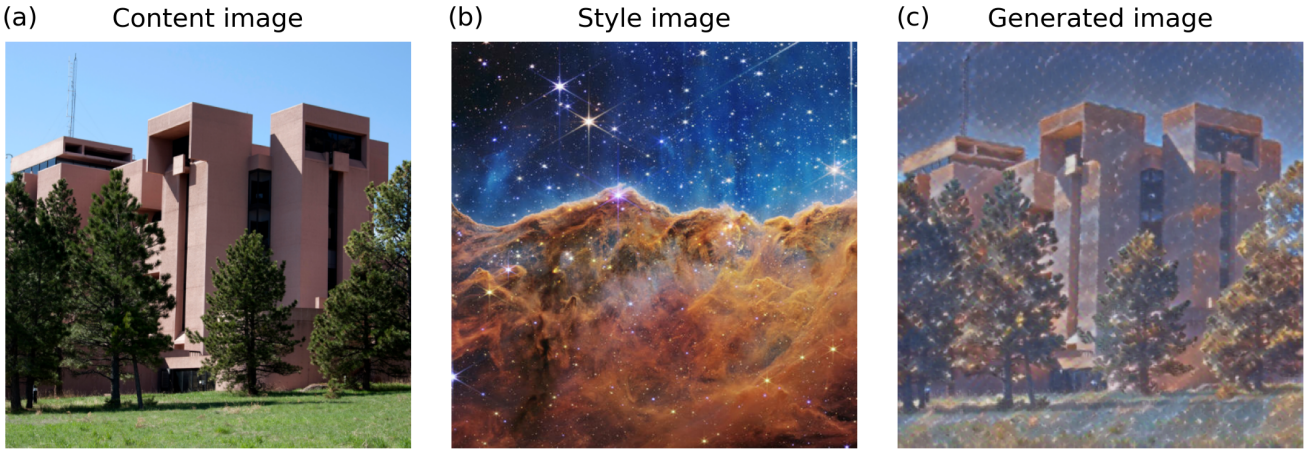

图1。 (a) 作为内容图像的美国国家大气研究中心梅萨实验室（Commons，2007年）。(b) 詹姆斯·韦布望远镜拍摄的展示恒星形成的图像，作为风格图像（美国国家航空航天局，2022年）。(c) 合成图像是利用内容图像和风格图像的神经风格迁移方法生成的产物。

为了将方案推广至更广泛的传感器信号处理领域，方案应能够应对特定传感器类型的专有行为；这要求具备更通用的能力，以表征仪器中难以建模的非理想行为。

本文介绍了一种新方法，用于弥合优化信号处理/机器学习解决方案与实际传感器数据之间的差距。我们表明，由（Gatys 等，2016年）提出的原始神经风格迁移方法利用卷积神经网络（卷积神经网络）执行图像增强，可以在无需人工标注的情况下生成外观逼真且带有标签的全息图示例，从而用于调整噪声参数。这些风格化的合成训练数据所实现的处理性能与先前方法相近，但无需人工标注，因此代表了向生成用于处理 HOLODEC 数据的可运行解决方案迈出的重要一步。

2 方法论

2.1 神经风格迁移

最初的风格迁移方法利用“内容”图像和“风格”图像的卷积神经网络表示，生成第三幅图像，将风格特征自动合成到内容特征中。例如，图 1 展示了如何运用该方法创建美国国家大气研究中心梅萨实验室的图像，如 (a) 所示；再使用詹姆斯·韦布望远镜近期拍摄的一张图像作为风格图像，如 (b) 所示，最终生成合成图像，如 (c) 所示。合成图像仍清晰地呈现了梅萨

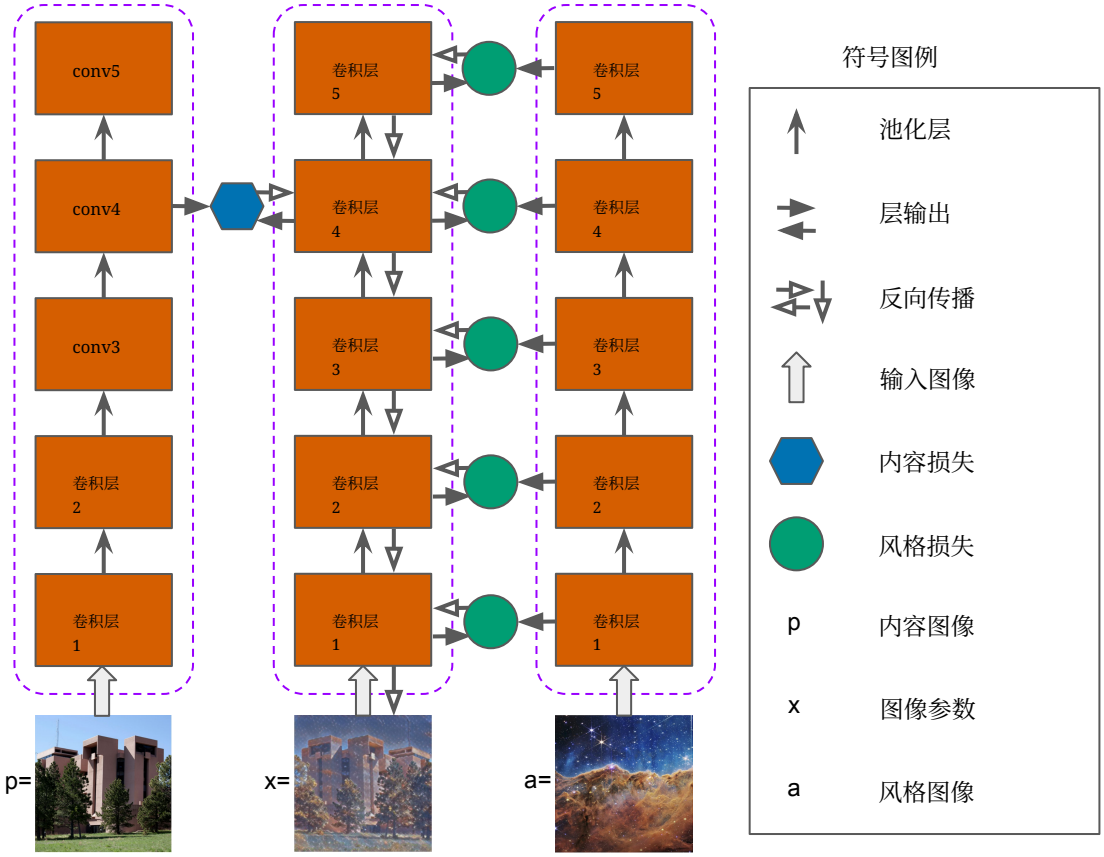

图2. Gatys等人（2016年）提出的神经风格迁移方法示意图，采用 VGG-19 神经模型，该模型的 18 个卷积层堆叠成五个模块（橙色方块）。在此示例中，使用五个模块中每个模块的第一个卷积层的输出计算风格损失和内容损失。这些模块表示输入 p、x和 a所共用的同一个模型。

实验室，不过整体而言，图像颜色、对比度和线条平滑度都相较于原图发生了变化，也呈现出风格图像的特征。

合成全息图带有标签且不含噪声，因此被视为具有我们希望保留的“内容”——也就是其中粒子的信息。HOLODEC 示例都包含不同程度的噪声/瑕疵，我们希望将这些特征有效地赋予合成示例，因此将其视为“风格”示例。通过这种方式，合成图像经过转换，可用于创建带有已知标签的逼真训练数据集。

图2示意性地展示了采用VGG-19网络作为所选卷积神经网络的风格迁移方法（Simonyan和Zisserman，2015年）。VGG模型包含19层，最初用于在ImageNet数据集上执行目标识别和定位（Russakovsky等人，2014年）。该模型的18个卷积层均采用ReLU激活，并依次连接成五个模块，各模块之间由二维平均池化层分隔，如图2所示。前两个模块各包含两个卷积层，其余三个模块各包含四个。卷积层的位置

在VGG中由整数 l表示。VGG-19中的所有卷积层均使用卷积核大小为 $3x3$ 的卷积核，步幅和填充大小均设为1。三个全连接层构成VGG-19的其余部分。因此，图2中的每一列都对应一个独立的（相同的）VGG-19模型（去除了全连接层）。目前，全息图像仅使用单个颜色通道，因此将第一个卷积层替换为颜色通道数为1的卷积层。

风格迁移方法首先为 x采样随机值，并冻结卷积神经网络中所有层的权重。对于输入内容图像 p 和生成图像 x，卷积层 l 的输出是特征矩阵表示 P^l和 F^l，其中 P^l_{ij}和 F^l_{ij}分别是卷积层位置 l处位置 j的 ith滤波器激活的响应。风格迁移的一个核心问题是如何选择损失函数，以及应使用哪些层输出来计算损失。选择将决定我们需要多大程度上使用完整的 VGG-19 模型，并会极大影响运行时性能。我们采用 Gatys等人（2016年）提出的原始方法，但也已有其他有效的损失函数（Li and Wand，2016b；Li等人，2017年；Wilmot等人，2017年）。通过使用两个特征表示的平方损失进行梯度下降，更新 x 的值，即可找到一幅与内容图像 p 的响应相匹配的新图像

$$\mathcal{L}_{content}(p,x,l) = \frac{1}{2}\sum_{i,j}(F^l_{ij} - P^l_{ij})^2 \tag{1}$$

对于位置为 l 的卷积层，如图2中的蓝色六边形所示。输入图像 x 的“风格”表示定义在一个用于捕捉纹理信息的特征空间中。它取为第 l 层在空间域上各输出之间的相关性，因此，风格特征图 i 和 j 的预期相关性由 Gram矩阵给出

$$G^l_{ij} = \sum_k G^l_{ik}G^l_{jk}. \tag{2}$$

现在，可以通过梯度下降更新图像 x，使其风格特征 G^l 与图像 a 在第 l 层的风格特征 A^l 相匹配。来自第 l 层的贡献为

$$E_l = \frac{1}{4N_l^2M_l^2}\sum_{i,j}(G^l_{ij} - A^l_{ij})^2 \tag{3}$$

而总风格损失（图2中以绿色圆圈表示）为

$$\mathcal{L}_{style}(a,x) = \sum_l w_l E_l \tag{4}$$

其中， w_l 表示来自第 l 层的各项贡献的权重。尽管 Gatys等人（2016年）并未使用这一方法，预测图像有时可能包含我们希望平滑掉的高频噪声。这可以通过总变分项实现，其定义为

$$\mathcal{L}_{TV}(x) = \sum_{i,j}|x_{i,j} - x_{i+1,j}| + |x_{i,j} - x_{i,j+1}| \tag{5}$$

对于像素 i 和 j。对于内容图像 p、风格图像 a和生成图像 x，总风格迁移损失为

$$\mathcal{L}_{total}(p,a,x) = \alpha\mathcal{L}_{content}(p,x) + \beta\mathcal{L}_{style}(a,x) + \gamma\mathcal{L}_{TV}(x) \tag{6}$$

其中，参数 α、 β和 γ 为恒定权重，且仅使用单个层的输出来计算内容损失（见图2）。通常，如果选择多个层输出来计算内容损失，则需在公式6中将各项求和。我们设定 $\alpha = \gamma = 1$，并在下文重点讨论 β 与预测图像之间的关系。

除 β 外，内容表示和风格表示所采用的层也作为超参数留待选择。对于内容特征层的选择，距离输入图像较远的卷积层输出，相较于那些靠近输入图像的模块，应能捕捉全局特征；而靠近输入图像的层则应能捕捉更多输入细节。风格特征也有类似的规律：距离更远的卷积层能够捕捉更全局的风格，反之亦然。Gatys等人（2016年）选取第四个模块中的第二个卷积层输出（例如 VGG-19 中第 12 个卷积层）来计算内容损失，并选取全部五个模块中第一个卷积层的输出（位于第 1、3、5、9 和 13 个位置的卷积层）来计算风格损失分量。这些选择会形成一个较大的卷积神经网络，在与我们相对较大的全息图配合使用时，会显著影响训练性能。我们发现，一个小得多的模型也能胜任 Gatys等人（2016年）所用模型的工作，因此运行速度快得多。具体而言，我们选取 VGG-19 中第四个卷积层的输出用于内容贡献，并选取 VGG-19 中前五个卷积层的输出用于风格贡献（公式 4 中所有权重均设为 1），因此只需使用 VGG-19 五个模块中的三个模块。或许存在一个更优、更快且能取得类似结果的模型，但我们没有进一步探索其他组合。

图2显示，内容图像 p 和风格图像 x 分别通过卷积神经网络，并在每种情况下保存选定特征层的输出。随后，初始白噪声图像 x 通过卷积神经网络，并使用内容层和特征层的输出来根据公式6计算总损失。接着，可以通过反向传播更新 x 中的图像像素；经过连续迭代， x 便能学习匹配 p的内容和 a的风格特征。

2.2 用于比较图像的指标

α 和 β 的选择将决定生成图像与内容图像和风格图像的相似程度。有许多指标可用于估算两幅图像之间的相似度。对于图像 x_0 和 x_1，我们选择结构相似性指数测量（SSIM）（Wang等，2004年）（公式A1），这是一种标准的基于像素的指标，可提取图像的结构、对比度和亮度信息，并以此为比较两幅图像的依据。此外，我们还采用一种基于 AlexNet 卷积神经网络模型（Krizhevsky等，2017年）的相似度指标，称为 $d_{\text{Alex}} \equiv d_{\text{Alex}}(x_0, x_1)$，用于捕捉两幅图像 x_0 和 x_1在深度特征空间中的距离（Zhang等，2018年）。SSIM的取值范围为-1至1，数值越大表示图像之间越相似。AlexNet指标的取值范围为0至1，数值越小表示越相似；我们使用1减去该值，使两种指标的数值越大都表示越相似。对于图像 x_0和 x_1，两种指标均通过对局部差异取平均值来计算：SSIM所用图像块的尺寸为 11×11，AlexNet指标所用图像块的尺寸为 3×3（即所用滤波器尺寸）。

2.3 使用 U-Net 进行图像分割

图 3 展示了一种神经分割模型，例如“U-Net”（Ronneberger 等人，2015年；Shelhamer 等人，2017年），该模型用于本研究中处理全息图，也曾用于其他研究（Shao 等，2020年）。该模型包含两个

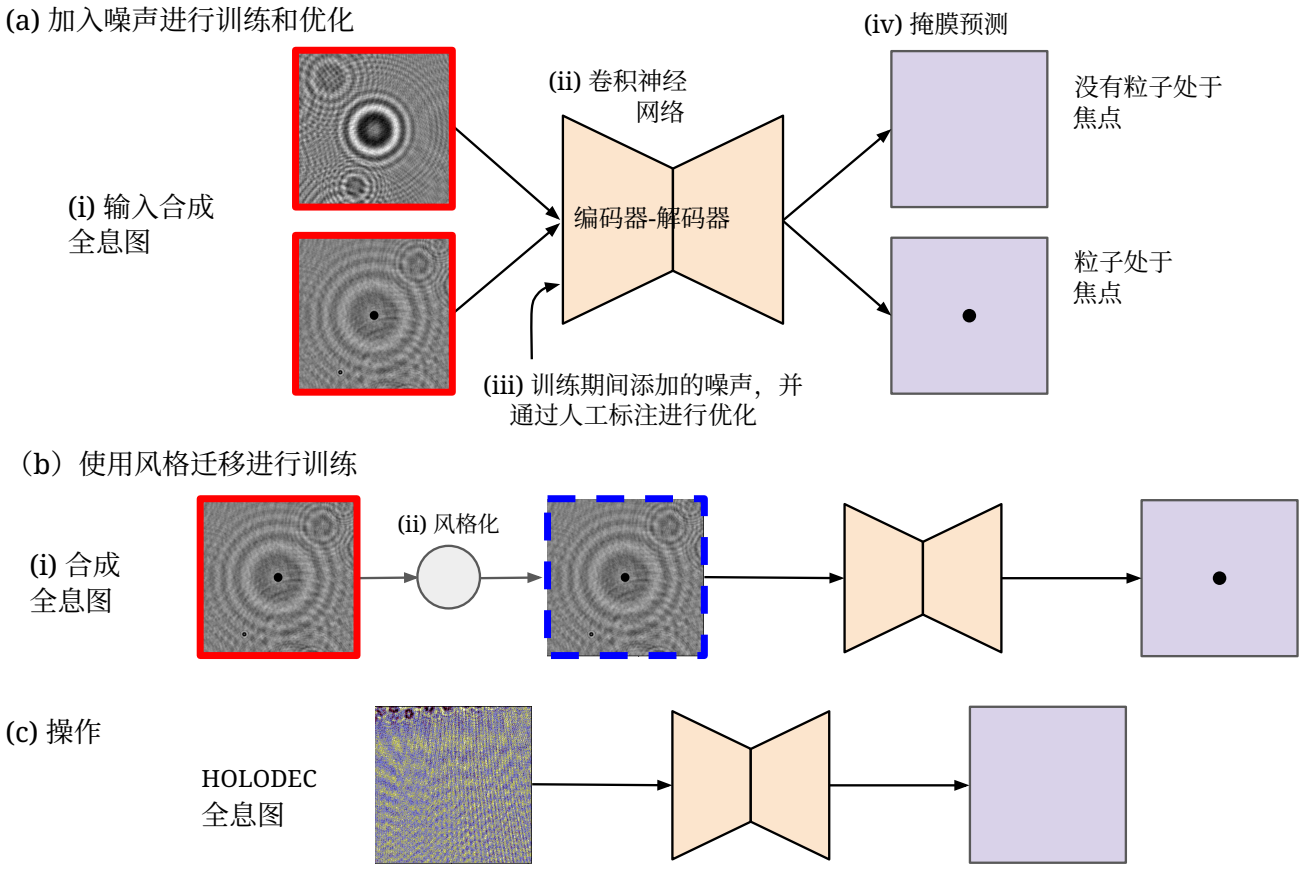

图 3. (a) U-Net分割模型。(i) 将合成全息图（实线红色）输入 (ii) 所示 U-Net 的编码器头部。(iii) 训练期间，可以选择向全息图添加噪声。(iv) 输入示例的预测掩膜，深色圆圈圈出了焦内粒子。(b) (i) 合成图像经 HOLODEC 全息图（蓝色虚线）(ii) 风格化，然后 (iii) 用于训练 U-Net 模型。(c) 实际运行时，将 HOLODEC 全息图输入经风格迁移图像训练的 U-Net，以获得掩膜预测。此示例中未识别出任何粒子。

如图所示的卷积神经网络“头部”。编码器头接收输入图像，并将其压缩为尺寸固定、维度更低的表示。解码器头对图像的潜在表示进行上采样，并输出一幅与输入图像尺寸相同的预测图像。这里，我们在解码器的最后一层使用 Sigmoid激活函数，将所有像素的预测输出重新缩放到 (0, 1) 范围内。换句话说，模型经过训练后，会预测某个像素被焦内粒子占据的概率。如果该概率超过某个阈值（对于二元任务，通常为 0.5），则将该像素标记为 1，否则标记为 0。U-Net 编码器头中的卷积层还可以利用其他卷积模型（例如在 ImageNet数据集上训练的 ResNet（He 等，2016））各层的预训练权重，这通常有助于加快训练速度并提升整体性能。

2.4 数据集

本研究中使用的数据集基本上有四个。前两个来自 HOLODEC 拍摄的全尺寸图像，或对 HOLODEC 进行模拟得到的图像。这些初始图像的尺寸为 4872×3248 像素，经过处理和分割后，生成可供 U-Net 直接处理的图像。在 U-Net 的训练过程中，还会对这些图像进行抽样，以确保焦内粒子和离焦颗粒的数量均衡。随后，利用来自这两个来源的训练数据进行风格迁移，生成逼真的生成数据集（内容来自合成图像，风格来自 HOLODEC 图像），并用于训练 U-Net。这里，风格迁移使用在 z 不同位置重建的合成图像和 HOLODEC 图像进行（见下文）。最后，使用一组经过人工标注的 HOLODEC 全息图来评估 U-Net 的性能，进而评估风格迁移方案的有效性。

2.4.1 HOLODEC 和合成数据

我们使用仪器的物理模型生成了 120 张合成全息图，其光学设置与 HOLODEC 仪器完全相同（Fugal等人，2009）。随机选取其中 100 张全息图用于训练，10 张用于验证，其余 10 张用于测试。每张全息图包含 500 个球形粒子，这些粒子随机分布在 x 和 y 方向上，沿 z 方向的位置分别介于最小值 14.072 毫米和最大值 158.928 毫米之间。颗粒直径按照伽马分布进行采样。

HOLODEC 数据选自数百张全息图，这些全息图来自 Cloud Systems Evolution in the Trades（CSET）项目的 RF07 子集，最初采集于2015年6月1日至8月15日（Albrecht等人，2019）。其中主要包含数量和形状各异的液态水颗粒，基本不含冰，但含有物理模型未能捕捉的其他未识别“伪影”和图像扰动。总体而言，已知粒子数量从零到数百不等，平均颗粒直径小于 $50\,m$。

按照（Schreck 等，2022）中描述的方法，首先通过沿 z轴（激光传播轴）重建 1,000 个平面来处理每幅全息图。由此生成 1,000 幅 1,200 万像素大小的图像，需要对这些图像进行扫描以查找焦内粒子。这些图像太大，无法用卷积神经网络高效处理。因此，按照我们之前的方法（Schreck 等人，2022年），将每幅全息图的重建结果分割成 38×28 个图块，每个图块为 512×512 像素。这些图块的重叠步幅为 128 像素，因此每幅完整图像共包含 828 个图块。这意味着任意一个像素都会出现在多个图块中。

使用合成全息图进行训练时，对每幅图像进行选取，然后沿 z 轴对其中全部 500 个粒子重新聚焦；随后选取并保存一个包含焦内粒子的随机网格图块，最终得到 50,000 幅图像（有关重新聚焦的详情，请参见 Schreck 等人（2022））。此外，还选取了 25,000 个图块，其中的粒子略微失焦，失焦距离不超过 120 微米。最后，随机选取了 25,000 个不包含焦内粒子的图块。随后采用类似流程生成验证集和测试集的图块图像，其中一半图像包含焦内粒子。

对于此处使用的 RF07 子集，从数百张全息图中随机抽取一张，排除先前留作人工检查的 20 张（并将在下文介绍）。这些原始数据可通过 https://doi.org/10.5281/zenodo.6347222 获取。

由于真实的粒子位置未知（而且无需知道，因为这些全息图仅用于风格迁移），因此在 z 方向上 14 至 160 毫米之间随机选取一个位置，对全息图重新聚焦，然后从重新聚焦后的全息图中随机选取并保存 10 个图块。重复这一过程，直到选出约 100,000 个图像切块用于训练、10,000 个用于验证，以及 10,000 个用于测试掩膜预测性能。每个数据集划分中的大多数图块都没有粒子直接处于焦点上，但偶尔也会非常接近焦点。

2.4.2 使用风格迁移生成数据集

带标签的合成训练集划分用作内容图像，而没有标签的 RF07 全息图则用作风格图像。我们针对 β 的 20 个不同取值创建了训练、验证和测试集划分，这些取值在约 10^{-5} 至 10^{15}, 的范围内均匀采样，并搭配 $\alpha = 1$ 和 $\gamma = 1$，以找到 β 的最优值。确定最优值后，我们使用 $\beta \equiv \beta_0 = 10^9$ 创建了一个包含训练、验证和测试集划分的数据集；我们将在下文中将其视为最优选择，并令 α 和 γ 均等于 1。创建该数据集是为了与仅使用合成训练集划分训练的模型，以及 Schreck 等人（2022）所述的模型进行比较；后者在训练期间加入了基于人工标注数据的优化噪声。

对于每个 β 取值（包括 β_0），我们随机选取一个内容（合成）图像图块，并与随机选取的风格（HOLODEC）图像图块配对。随后使用风格迁移方法，将内容图像和风格图像共同用于生成训练样本；生成样本的掩膜标签继承自内容图像。对于 β_0 数据集，我们生成了 100,000、20,000 和 20,000 张风格生成图像，分别用于训练、验证和测试 U-Net 模型复现掩膜标签的能力。对于其余 20 个 β 取值，每种情况下都生成了总计 20,000 个训练样本、2,000 个验证样本和 2,000 个测试样本，用于相同目的。对于后面这些数据集，我们还使用了相同的随机种子，以确保在不同的 β 选项下（不包括 β_0），每个数据集划分中选取的所有图像对都相同。

2.4.3 手动标注的 HOLODEC 数据集

这 20 个手动标注的 HOLODEC 示例共包含 2,356 张图像，每张图像的尺寸为 512 × 512 像素，待确认的焦内粒子位于图像中心。这些示例经人工标注，注明是否包含焦内粒子，并给出 0 到 5 的审阅者置信度评分（分数越高表示越有把握）。U-Net 的二值输出采用简单规则：只要任一预测值大于 0.5，就记为 1，否则记为 0。许多示例具有主观性，而且我们能够识别的粒子总数受限于此前的建模尝试，因此这些全息图中可能还有尚未发现的粒子。

标注样本总数被分为验证数据集和测试数据集。验证集共包含1,204张图像，其中367张至少包含一个焦内粒子。测试集包含1,154张图像，其中874张至少包含一个焦内粒子。验证集用于指导我们此前研究中的噪声优化（本文将与之进行比较）（Schreck 等，2022），而测试集则代表由人工标注样本组成的留出集。

3 结果

3.1 全息图风格迁移

图4展示了一张包含单个焦内粒子的合成全息图（内容图像），以及用于为 α = 1和 $\beta = 10^9$生成风格迁移图像的不同 HOLODEC 全息图示例（风格图像）。图 4(a) 所示的 HOLODEC 示例之所以被选中，是因为其中包含一个焦内粒子。请注意，虽然风格图像中的细尺度纹理似乎有所迁移，但这些风格图像中的粒子并未迁移到生成图像中。因此，无需为风格迁移图像寻找空白全息图，也无需在挑选适合这一过程的图像时进行严格的质量控制。

图4(b-c)中的图像是典型示例：仅使用合成图像训练的 U-Net（左列）在实际 HOLODEC 图像上的表现不佳，假阳性率很高。图4(b)中的 HOLODEC 图像可见干涉条纹，这些条纹常会产生类似焦内粒子的图案；而(c)则展示了许多细小伪影呈圆形的示例。在这三个示例中，合成示例中的焦内粒子仍保留在生成图像中，而且生成图像的整体对比度与 HOLODEC 图像更为接近。

图5(a)展示了图4(a)所示生成图像的训练曲线。图5(b)展示了若干选定训练轮次下预测图像的变化过程。以 x 的随机噪声图像作为初始输入，风格损失在最初几个训练轮次中急剧下降。内容损失和总变差损失也在最初几个训练轮次后迅速下降，随后曲线大致趋于平缓；约经过 10 个训练轮次后，内容损失的贡献开始主导总损失。在这些示例中，α/β 的选择决定了生成图像与用于生成它的内容图像和风格图像之间的相似度。根据 AlexNet指标，所有生成图像与 HOLODEC 图像的相似度都高于与合成图像的相似度。

图6(a)和(b)展示了 SSIM 和 d_{Alex} 指标随 β 变化、并采用 α = 1时的情况。这些指标分别针对 HOLODEC-合成、HOLODEC-生成以及合成生成图像对的测试数据计算。请注意，对于 HOLODEC-合成图像对，这些指标保持不变，因为这些图像不受 β影响。

在两幅图中，均用垂直线大致标出了四个感兴趣区域。区域 A 对应约为 $\beta \leq 10$ 的取值，此时内容损失在总损失中占主导地位。因此，这两项指标都表明，风格迁移图像与原始合成图像非常相似（另见图4(a,b)）。这两项指标的主要区别在于，与 SSIM 相比，d_{Alex} 对 HOLODEC-生成图像对的评分高于对 HOLODEC-合成图像对的评分。不过，对于 β的所有取值，这两项指标的曲线在定性上都相似。

图6中各区域的图像示例见图7；这些示例采用了图4(a)中的内容图像和风格图像，使用 $\alpha = 1$，并改变 β。

大约在 $10^2 < \beta \leq 10^7$时可识别出B区，此时合成图像与HOLODEC生成图像对的两个指标均显著下降。图4(c,d)中的示例表明，随着风格项开始发挥作用，生成图像开始与原始合成图像呈现出明显差异。尽管这些示例可能给人一种生成图像仍然像合成图像的印象，但实际上它们是一种混合表示

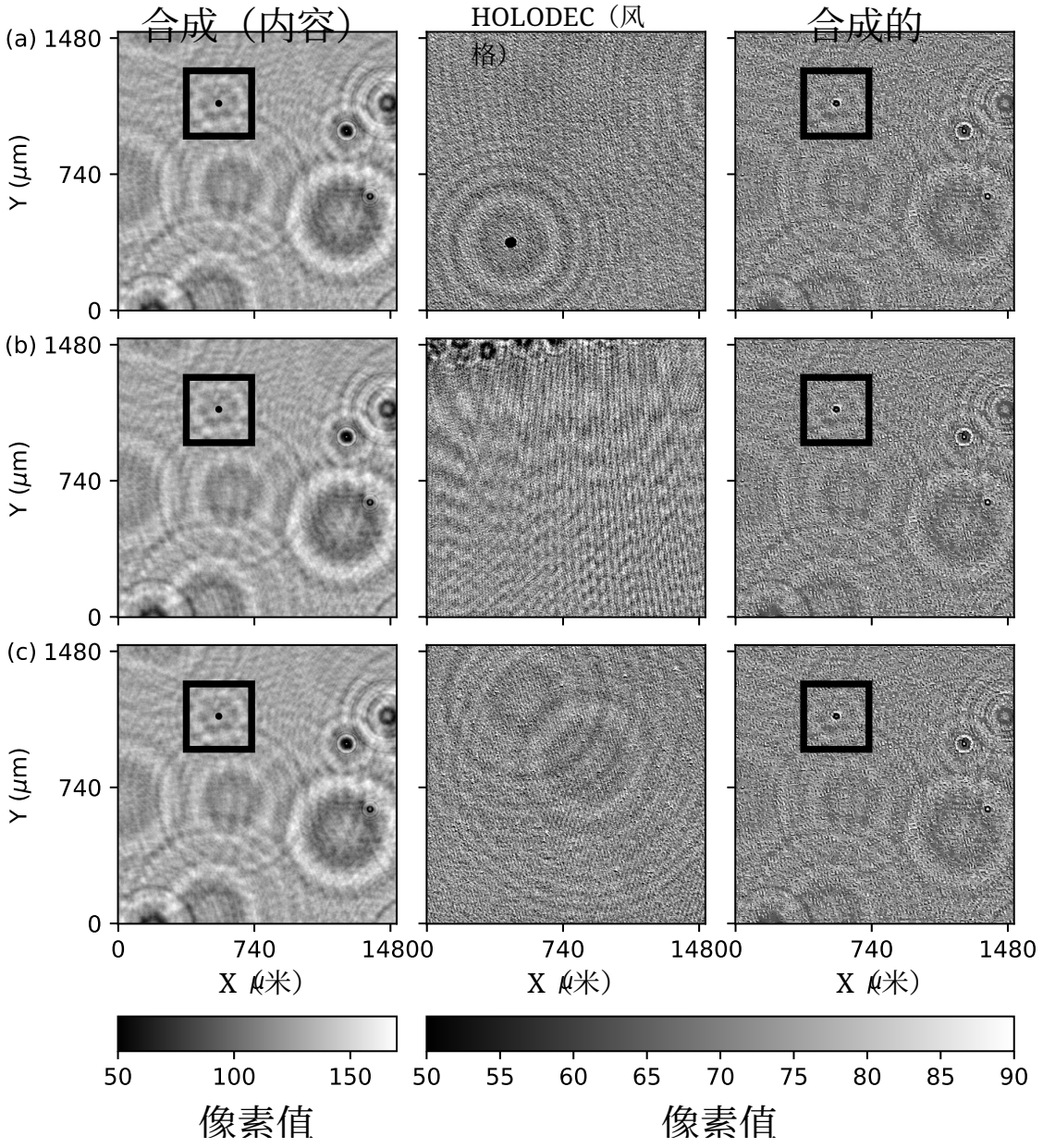

图4. 在 (a-c) 中，将同一张合成全息图分别与不同的 HOLODEC 全息图配对，通过风格迁移生成全息图。HOLODEC 示例展示了 (a) 一个焦内粒子，(b) 干涉图样和连续的暗斑，以及 (c) 小型伪影。研究发现，仅使用合成图像训练时，U-Net 网络在 (b) 和 (c) 上的表现较差。内容权重 $\alpha = 1$，而风格权重 $\beta = \beta_0 = 10^9$。

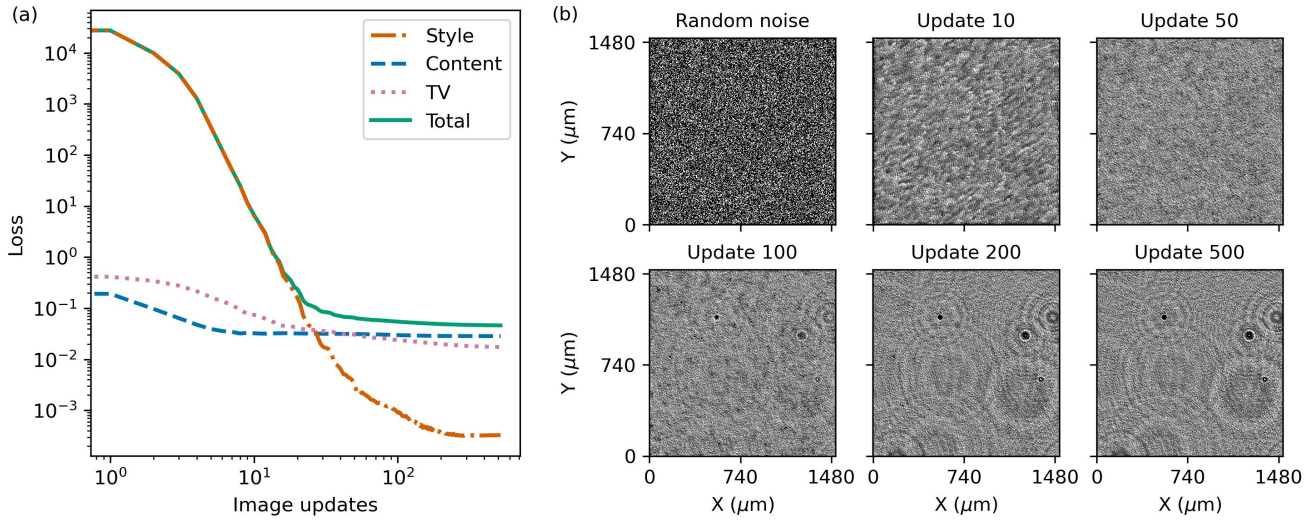

图5. (a) 风格、内容、总变差和总损失随训练轮次（图像更新的次数）变化的曲线。(b) 展示了初始随机图像以及不同训练更新阶段选取的图像。与 HOLODEC 和生成图像在图4中的情况相同，所示像素值范围为 50 到 90。

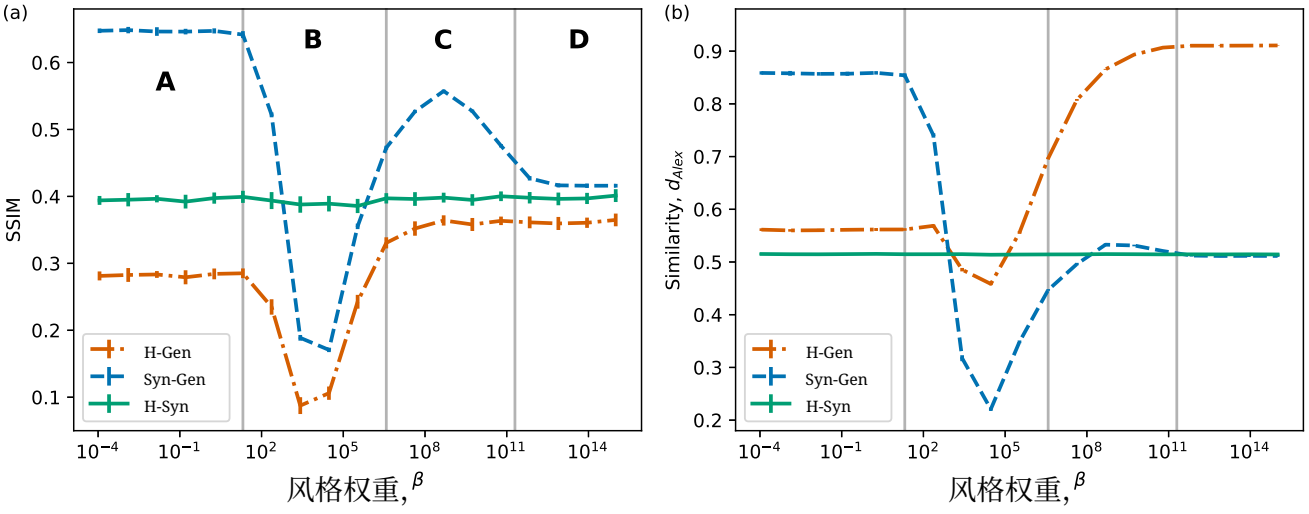

图6. 使用测试数据集计算的平均 SSIM（a）和平均 d_{Alex} 分数（b），随公式6中的风格权重 β 变化的曲线，其中 α 和 $\gamma = 1$。

平均而言，并不类似于合成全息图或 HOLODEC 全息图（例如，图4(c,d)中这类示例的平均对比度差异小于合成图像，但仍大于 HOLODEC 示例中的观测值）。在 $\beta = 10^4$ 附近，两个指标都达到最小值，随后随着 β 增大而开始上升。

当 $10^7 < \beta \leq 10^{12}$ 时，大致可判定为区域 C。对于合成-生成图像对，SSIM 和 d_{Alex} 均继续上升，并在 $\beta = 10^9$ 附近达到峰值，不过 SSIM 的峰值更明显。HOLODEC-生成图像对的两项得分也有所上升并达到平台值，其数值高于区域 A 和区域 B，表明图像在一定程度上已被风格化；而合成-生成图像对的得分则表明图像仍保留着内容特征。显然，神经网络 d_{Alex} 判定生成的全息图与 HOLODEC 全息图的相似度高于 SSIM 的判定结果，但总体而言，两个指标都表明预测图像在保留内容与习得风格特征之间达到了某种平衡。图 4(e-g) 显示，焦内粒子仍保持原位，形状也大致不变；与此同时，例如，图像对比度经过调整，变得更接近 HOLODEC 图像。

最后，区域 D 由 $\beta > 10^{12}$ 确定，此时风格项在损失中占主导地位。风格生成图像对的 d_{Alex} 分数仍然很高，而内容生成图像对的分数则收敛至与内容-风格图像对相同的值，也就是说，该指标无法区分生成图像与风格图像。然而，图 4(h,i) 显示，合成图像中的焦内粒子及其他内容大多已被移除。此外，生成图像中开始出现看似伪粒子的东西。换言之，原始图像的风格化程度过高。

区域 A 表示输出中的内容图像几乎没有被赋予风格，因此最接近理想化的合成数据。我们此前已在 Schreck 等人（2022）中指出，当目标是处理真实的 HOLODEC 数据时，这类数据并不适合作为 U-Net 训练数据集。区域 D 中的图像则表示输出内容基本缺失的情况。显然，我们无法期望训练 U-Net 来识别输入图像中不存在的粒子。若不采用人工标注的 HOLODEC 图像，显而易见的选择是将 β 设在区域 C 中 $\beta_0 = 10^9$ 附近，此时合成生成图像对的两项指标均达到峰值，而且根据 d_{Alex}，生成示例看起来与 HOLODEC 示例更加相似。

因此，当 β 接近 $\beta_0 = 10^9$时，预期在生成图像上训练的 U-Net 在掩膜重现方面表现最佳，因为区域 C 中的预测图像与内容图像最相似。此外，与区域 A 和 B 相比，生成图像在区域 C 中与 HOLODEC 图像更为相似，因此我们可能会预期，在这些生成图像上训练的 U-Net 模型在 HOLODEC 图像上的表现最佳。如果确实如此，可以根据 d_{Alex} 和 SSIM 找出生成图像与 HOLODEC 图像更相似时对应的 β 值，以此选定 β。这样选出的生成数据集应最适合用于训练 U-Net。

3.2 在合成图像和 HOLODEC 图像上的表现

为了检验能否根据 d_{Alex} 和 SSIM 以解析方式确定最优 β 这一假设，我们使用不同 β 值下的生成图像（及其对应标签）来训练 U-Net 模型，如图 3(b) 所示。对于每个不同的 β 值，模型架构和训练超参数均完全相同（更多详情见附录 A1）。训练完成后，在生成图像测试集和人工标注的真实 HOLODEC 图块测试集上，以运行模式运行每个 U-Net 模型。图 8(a) 展示了 U-Net 对生成图像所附掩膜标签的复现情况的若干二元性能指标，而图 8(b) 展示了真实 HOLODEC 图块中粒子检测的类似指标。由于这些人工标注的图块样本均未用于创建风格生成图像，因此 2,356 个样本全部

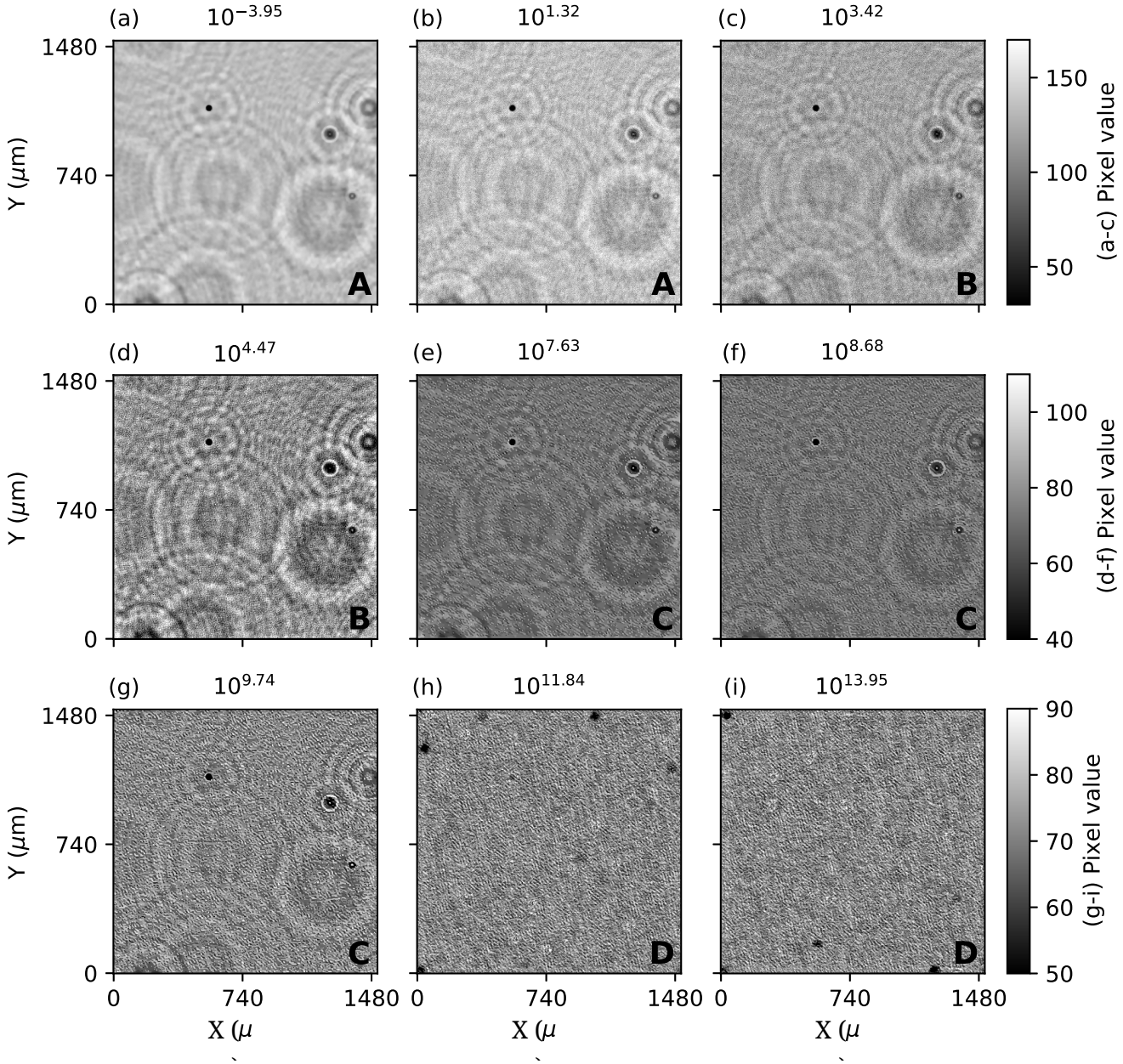

图7. (a-i) 使用图4(a)中的内容和风格图像，以及不同风格权重生成的全息图示例。区域标示在右下角。

均用于图中。所用指标包括 F1分数、ROC 曲线下面积（曲线下面积，AUC）、检出概率（POD）和虚警率（FAR）。

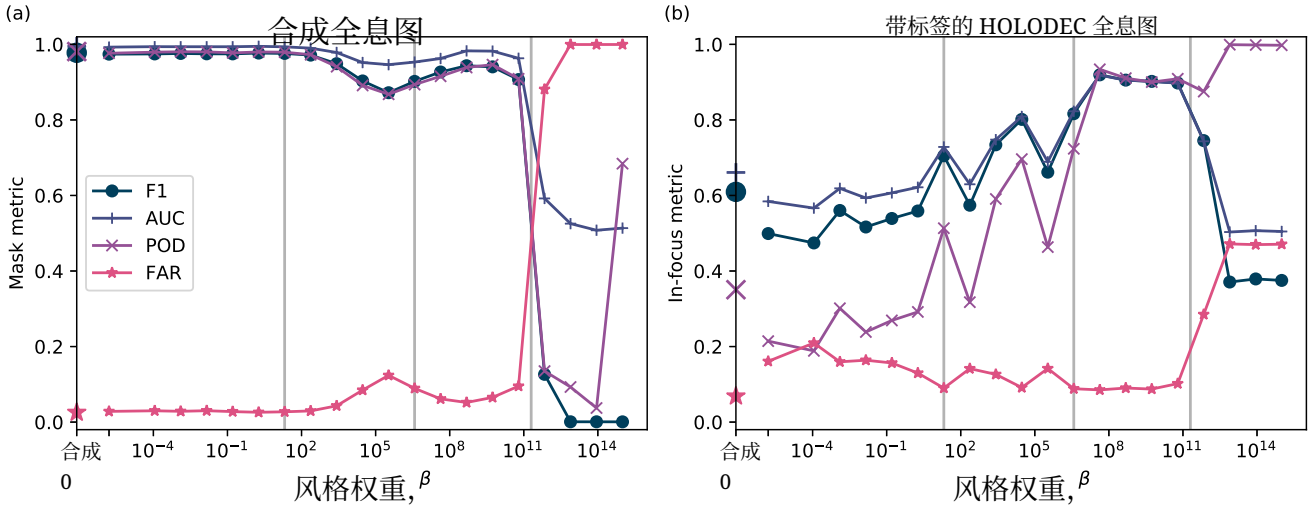

图 8。 (a) 绘制了 U-Net 掩膜预测性能的若干二元指标随风格权重 β 变化的情况，适用于合成图像或风格化生成图像。(b) 展示了人工标注的 HOLODEC 图像中粒子检测任务的相同二元指标随 β变化的情况。由于模型在训练期间未使用人工样例，因此图中使用了全部 2,356 个样例。横轴标签“Syn”和“0”分别表示使用合成图像训练的 U-Net 网络，以及使用 $\beta = 0$生成图像训练的 U-Net 网络。

图 8(a) 显示，随着 β增加，整体训练性能逐渐下降。这是预期结果，因为理想模拟经过越来越强的风格化处理，加入了非理想仪器噪声和伪影，使粒子检测更加困难。在区域 A 中，生成的全息图与合成样例相似，掩膜预测性能非常高且总体平稳。但在 B区，随着生成的全息图受到风格图像更大程度的影响，曲线下面积和检出概率明显下降，而虚警率则上升。随着 β 继续增大，生成图像开始同时呈现内容（合成图像）和风格（HOLODEC）样例的特征，性能随之上升，并在 $\beta = \beta_0$ 附近达到峰值，之后开始下降，起初较为缓慢，进入 D区后则骤然下降。我们预计，区域 C 中的峰值会低于区域 A 中的峰值，因为前者的图像缺少仪器的典型噪声。在 D区，曲线下面积约为 0.5，这是可以预料的，因为过度风格化图像及其掩膜标签保留的真实粒子内容非常少。使用过度风格化图像训练时，U-Net 会输出全部或过多的 1，因此检出概率会迅速趋近于 1。我们还计算了预测掩膜的直径，发现对于测试集，除 D区中的 β 取值外，所有取值下的直径都与真实值均值没有显著偏差。

总体而言，图 8(b) 表明，可根据我们的假设选择 β 的相关范围。与仅使用合成图像训练的 U-Net（在图 8(b) 的横轴上标记为「Syn」）相比，在区域 A 中，模型在 HOLODEC 示例上的表现体现为较低的检出概率和较高的虚警率，尽管曲线下面积值都约为 0.6。即使是使用带有 $\beta = 0$ 的生成图像训练的模型，其虚警率也高于「Syn」，这表明使用理想化合成全息图训练的 U-Net 倾向于少检粒子，从而获得较低的虚警率。需要注意的是，仅使用合成图像训练的情形与 $\beta = 0$ 情形之间的性能差异表明，设置 $\beta = 0$ 并不能完美

指标	F1		AUC		POD		FAR		CSI	
	合成/生成	HOLO	合成/生成	HOLO	合成/生成	HOLO	合成/生成	HOLO	合成/生成	HOLO
合成	0.977	0.220	0.992	0.448	0.979	0.106	0.025	0.387	0.956	0.100
优化噪声	0.968	0.866	0.985	0.808	0.967	0.922	0.031	0.096	0.937	0.840
风格化	0.966	0.864	0.988	0.805	0.968	0.922	0.035	0.097	0.935	0.839

表1。 列出了每个经过超参数优化的模型的若干指标，这些指标分别针对掩膜和粒子检测的二元预测，在合成或生成（合成/生成）以及 HOLODEC（HOLO）测试数据集上计算得出。带标注的 HOLODEC 图像验证集用于优化优化噪声模型中的噪声参数。

地复现内容图像，导致 U-Net 性能略有下降。然而，这两种情形都与本研究的目标无关，因为它们在 HOLODEC 数据上的表现都不佳。在区域 B 中，即使掩膜性能下降，粒子检测性能总体上仍有所提升。不过，粒子检测性能的波动相当显著，因此在没有人工标注样例的情况下，难以可靠地选择 β。在区域 C 中，检测性能达到大致的平台期，曲线下面积值较高，且虚警率始终较低。当风格权重接近区域 D 的取值时，性能似乎略有下降。在区域 D 中，模型无法有效训练，导致输出全为 1。因此，检出概率和虚警率都很高，曲线下面积为 0.5。总体而言，区域 C 中的模型在 HOLODEC 图像上的表现也优于区域 A、B 和 D。正如预期，区域 D 中的模型在人工样例上的表现完全失效，曲线下面积接近 0.5。

3.3 优化模型的比较

现在，无需依赖人工样例进行优化即可选择风格权重，我们选择了 $\beta = \beta_0 = 10^9$，并使用风格迁移图像训练了一个模型。表1比较了该模型（“风格化”）、仅使用理想化合成图像训练的模型（“合成”），以及在训练期间引入噪声、使用合成图像训练的模型（“优化噪声”，其优化过程需要人工标注）。在这三种情况下，我们都进行了广泛的超参数优化，以寻找最佳架构和训练参数；优化指标是生成图像验证集上的掩膜性能。训练期间利用噪声的模型还使用了人工标注的 HOLODEC 验证样例，以指导所添加噪声的优化。此外，表中还列出了图 8 中相同的二元指标，以及测试图像（合成或生成）上的掩膜性能和测试 HOLODEC 图像上的检测性能所对应的临界成功指数（CSI）。由于使用人工标注样例验证集来选择所添加的噪声，从而得到优化噪声模型，因此表1中仅使用测试划分。

表1显示，尽管风格化模型不需要人工标注，但在两个数据集的所有指标上，风格化模型与优化噪声模型的表现几乎相同。优化噪声模型在每个指标类别上的表现都比风格化模型高出百分之零点零几。合成模型在合成数据上的掩膜性能更好。然而，与另外两个模型相比，合成模型在真实 HOLODEC 数据上的粒子检测表现差得多

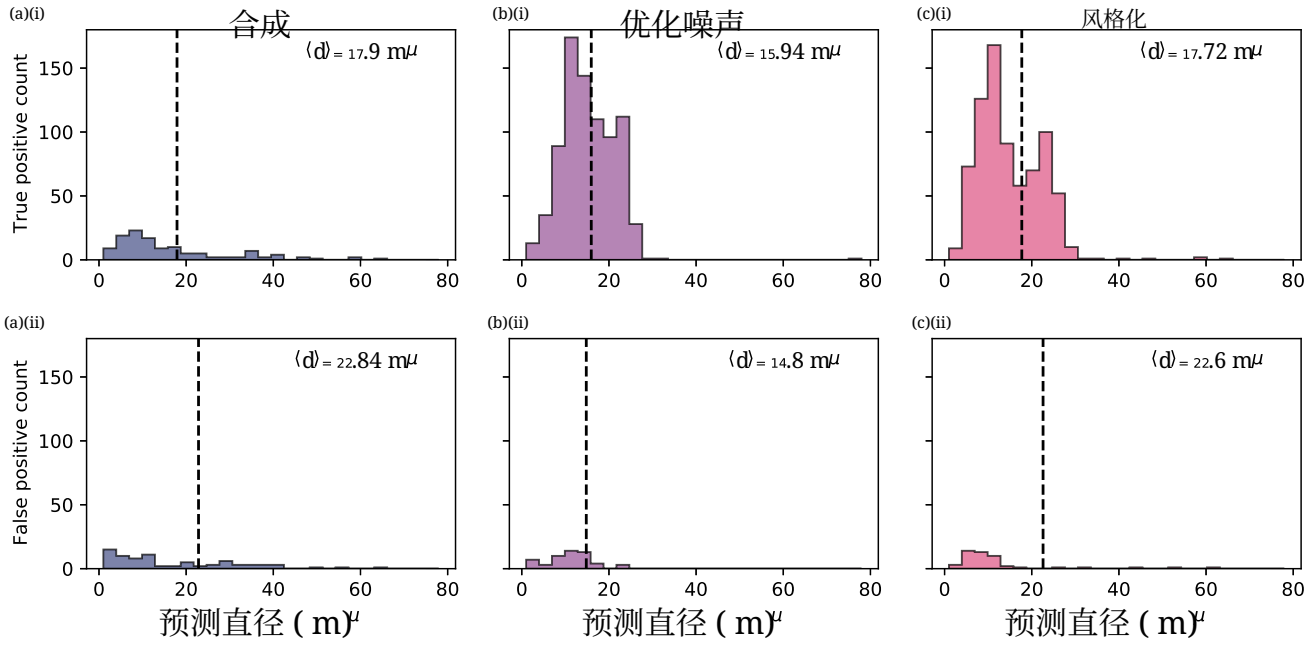

图9. (a-c) 测试集中人工标注的 HOLODEC 图像对应的三个模型所计算出的粒子直径直方图。每个面板中的虚线表示预测的平均粒径 $\langle d \rangle$。

。例如，表1显示其虚警率（FAR）高得多，临界成功指数（CSI）比其他模型低了 70% 以上。

最后，由于优化噪声模型和风格化模型的表现几乎相同，我们计算了人工标注的 HOLODEC 样例测试集中预测粒子的直径，以探究模型如何描述粒子形状。图9展示了三个模型预测直径的直方图。若预测掩膜中至少存在一个焦内粒子，则计算其粒子直径。在预测掩膜中，任何标记为 1 的像素，若其沿 x 或 y 方向（但不包括平面内的对角方向）相邻像素也标记为 1，则将其归为同一组。因此，组与组之间的间隔定义了多个粒子。组的直径取其在 x 或 y 方向上的最大跨度。

图9(a)(i)显示，与另外两个模型相比，合成模型如预期般低估了真实粒子数量；如图9(a)(ii)所示，其假阳性预测数量几乎与真阳性预测数量相当。同时，优化噪声模型和风格化模型预测出的分布相似，但与合成模型的分布不同。这两个模型预测出两个峰值，中心分别约为15和22 μm，其中风格化模型的第二个峰值更为明显。与优化噪声模型相比，风格化模型预测出的较大粒子数量也略多，而合成模型有时也能观察到这些较大粒子。因此，风格化模型和优化噪声模型预测的平均直径分别约为17.8 μm和15.9 μm。图9(b)(ii)和(c)(ii)进一步显示，这两个模型预测的虚假直径大多为15 μm或更小。

4 讨论

总体而言，优化噪声模型和风格模型在合成图像和HOLODEC 图像上的表现相当，而且这两个模型的表现均优于仅使用合成全息图训练的模型。这表明，将风格迁移方法用于全息图像增强具有优势：它无需人工标注来优化 U-Net 模型训练期间添加的噪声。此外，我们也无需像优化模型那样，含糊地选择对图像进行的增强类型。再者，由于风格迁移应用于用于训练的图像，而非在运行期间应用，因此其计算性能也与优化噪声模型相当。

涉及风格迁移方法的主要计算瓶颈出现在创建风格迁移训练数据集的过程中。许多所需的生成图像必须逐一从独特的内容-风格图像对中创建，并且都需要采用上述迭代优化方法。所使用的图像尺寸也相当大。这进一步延长了训练时间，使得创建用于探索不同权重组合的大型数据集变得十分繁琐。改进方法可以着重于用经过训练的神经网络替代优化过程，由该网络预测风格化图像。在采用前馈架构时，这能够使速度比原始方法提升几个数量级（Li and Wand，2016年a；Ulyanov 等，2016年；Huang and Belongie，2017年）。其中可能包括 StyleGAN 方法（Karras 等，2021年），该方法将风格迁移的概念与生成对抗网络相结合（Goodfellow 等，2020年），并能够精细调节不同尺度下图像特征的相对强度。对于全息图而言，StyleGAN 可能可以用于按需更有选择地控制特定的内容或风格特征（例如，添加更多与伪影相关的特征）。

此外，我们仅使用了 CSET 全息图，特别是 RF07 数据划分，其中粒子密度和粒子尺寸范围都有限。利用某次野外观测活动中的风格图像训练的模型，在其他野外观测活动中的表现如何，仍有待未来研究检验。正如我们在 Schreck 等人（2022）中所了解到的，风格生成的全息图数据集与运行输入并不相似，因此很可能无法作为训练目标。在这种情况下，可以将最近一次野外观测活动中的全息图用作风格示例，以生成相关的训练数据集。StyleGAN 方法同样有望利用野外观测活动中获得的各种全息图（以及每张图中包含的所有独特物体），因为它或许能够利用数据集中出现的多种风格和内容特征。

最后，由于本研究的主要目标是在不牺牲性能的前提下去除人工标注步骤，因此我们没有尝试评估生成全息图的物理真实性，因为实际运行时只使用真实的 HOLODEC 全息图。不过，它们在物理上是否合理仍是一个有待回答的问题。检验这一假设的一种方法是对全尺寸全息图进行风格迁移；这在计算上颇具挑战，但借助现有图形处理器并非不可能，然后再通过沿焦平面进行波传播来重新聚焦这些全息图。这需要忠实地重建电场的实部和虚部，而不仅仅是强度；本研究采用后者，是因为风格迁移应用于从全尺寸全息图传播后选取的图块。

参数	合成	优化噪声	风格化（β_0）	$10^{-5} < \beta \leq 10^{15}$
学习率	3.86×10^{-4}	2.46×10^{-4}	8.5×10^{-5}	1×10^{-3}
训练损失	焦点-Tversky	焦点-Tversky	IoU	焦点-Tversky
分割模型	U-Net	LinkNet	U-Net	U-Net
编码器模型	EfficientNet-b0	Xception	DenseNet121	VGG11
分块变换	255	归一化	标准	对称
L2正则化	0.0	0.0	2.1×10^{-7}	6.0×10^{-6}
高斯模糊 σ	-	2.125	-	-
高斯噪声	-	0.326	-	-
亮度因子	-	1.270	-	-

表 A1. 针对三种物种的神经分割模型进行优化研究所得的最佳超参数值。批量大小固定为 16。

5 结论

总之，我们证明，风格迁移算法能够有效地将使用仪器理想化物理模型创建的合成全息图，转换为与仪器实际观测结果相似的全息图。原则上，利用风格迁移进行图像增强不应局限于全息图，因为仪器的非理想行为是许多领域共同面临的问题。像本研究这样，使用合成全息图训练机器学习模型，以预测焦内粒子周围的掩膜时，必须在训练期间向图像注入噪声，才能让模型在合成数据和原始数据上都表现良好。然而，只有在对原始图像进行耗时的人工标注并完成超参数优化之后，才能确定噪声变换及其参数。风格迁移方法将干净的全息图数据集转换为更接近观测全息图的数据集，无需繁琐的人工标注要求，便实现了同等的掩膜预测性能。此外，在一小组人工标注的 HOLODEC 示例上，两个模型的表现相当，并且预测出的粒子直径分布相似。

A1 模型和训练超参数

表 A1列出了所研究的分割模型找到的最佳参数。关于使用全息图数据集对分割模型进行超参数优化的更多详情，请参阅 Schreck 等人（2022年）。表 A1 中发现的最优分割模型为 U-Net（Ronneberger 等人，2015年）和 LinkNet（Chaurasia 和 Culurciello，2017年）；所考察的预训练编码器模型权重则包括 DenseNet-121（Huang 等人，2017年）、Xception（Chollet，2017年）、EfficientNet-b0（Tan 和 Le，2019年）以及 VGG-11（Simonyan 和 Zisserman，2014年）。有关分割模型和编码器模型的更多详情，请参阅位于 https://github.com/qubvel/segmentation_models.pytorch 的 segmentation-models-pytorch 软件包。找到的最优训练损失为交并比（IOU）和焦点-Tyversky 损失（Lin 等人，

2017年；Salehi等，2017年）。有关各损失函数的更多定义，请参阅位于 https://github.com/NCAR/holodec-ml 的 Holodec-ML 软件包。我们还在所有训练模型中使用了从 ImageNet 数据集获取的预训练权重。在将图块输入分割模型之前应用的图块变换包括：将所有像素除以 255；重新缩放所有像素，使其取值介于 0 和 1 之间（归一化）或介于 -1 和 1 之间（对称）；或者对每个图块进行重新缩放，即减去像素均值并除以方差的平方（标准）。

A2 结构相似性指数（SSIM）

用于比较图像 x 和 y 的全局结构相似性定义为

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{A1}$$

其中， C_1 和 C_2 为固定常数， μ和 σ 定义为

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \sigma_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)^2 \qquad \sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y). \tag{A2}$$

实际应用中，我们使用平均结构相似性指数，该指数通过对 11×11 的圆对称高斯加权函数取平均得到（Wang等，2004年）。

A3 数据集

HOLODEC 和合成数据集可通过 https://doi.org/10.5281/zenodo.6347222 获取。带标签的 HOLODEC 示例是 RF07 数据集的一个子集（验证集 ID 0-9，测试集 ID 10-19），而合成全息图则通过模拟生成。RF07 数据集中的所有全息图都用于创建风格迁移数据集。更多详情请参见第 2.4 节。风格生成数据集可使用位于 https://github.com/NCAR/holodec-ml 的 Holodec-ML 软件包构建。

致谢。本材料基于美国国家大气研究中心开展的工作，该中心是由美国国家科学基金会根据合作协议第1852977号资助的主要设施。我们感谢由美国国家科学基金会资助、美国国家大气研究中心计算与信息系统实验室提供的Cheyenne和Casper计算与信息系统实验室（CISL，2020年）高性能计算支持。本文所述神经网络以及用于训练和测试模型的模拟代码已存档于 https://github.com/NCAR/holodec-ml。为本研究创建的所有HOLODEC和合成全息图数据集均可在 https://doi.org/10.5281/zenodo.6347222 获取。

参考文献

Albrecht, B.、Ghate, V.、Mohrmann, J.、Wood, R.、Zuidema, P.、Bretherton, C.、Schwartz, C.、Eloranta, E.、Glienke, S.、Donaher, S.等：《热带信风区云系统演变（CSET）：利用NSF–NCAR GV跟踪边界层云系统的演变》，《美国气象学会公报》，100，93–121，2019年。Chaurasia, A.和Culurciello, E.：《Linknet：利用编码器表示实现高效语义分割》，载于：2017年IEEE视觉通信与图像处理会议（VCIP），第1–4页，IEEE，2017年。Chollet, F.：《Xception：采用深度可分离卷积的深度学习》，载于：IEEE计算机视觉与模式识别会议论文集，第1251–1258页，2017年。Commons, W.：《美国国家大气研究中心》，科罗拉多州博尔德，https://upload.wikimedia.org/wikipedia/commons/3/31/ National_Center_for_Atmospheric_Research_-_Boulder%2C_Colorado.jpg，2007年。计算与信息系统实验室，CISL：《Cheyenne：HPE/SGI ICE XA系统（NCAR社区计算）》，技术报告，美国国家大气研究中心，https://doi.org/10.5065/D6RX99HX，2020年。Fugal, J. P.、Shaw, R. A.、Saw, E. W.和Sergeyev, A. V.：《用于云粒子测量的机载数字全息系统》，《应用光学》，43，5987–5995，2004年。Fugal, J. P.、Schulz, T. J.和Shaw, R. A.：《数字同轴全息图中粒子的自动重建与表征实用方法》，《测量科学与技术》，20，075501，2009年。Gatys, L. A.、Ecker, A. S.和Bethge, M.：《使用卷积神经网络进行图像风格迁移》，载于：2016年IEEE计算机视觉与模式识别会议，CVPR 2016，美国内华达州拉斯维加斯，美国，2016年6月27日至30日，第2414–2423页，IEEE计算机学会，https://doi.org/10.1109/CVPR.2016.265，2016年。Goodfellow, I. J.、Pouget-Abadie, J.、Mirza, M.、Xu, B.、Warde-Farley, D.、Ozair, S.、Courville, A. C.和Bengio, Y.：《生成对抗网络》，《ACM通讯》，63，139–144，https://doi.org/10.1145/3422622，2020年。He, K.、Zhang, X.、Ren, S.和Sun, J.：《用于图像识别的深度残差学习》，载于：2016年IEEE计算机视觉与模式识别会议（CVPR），第770–778页，https://doi.org/10.1109/CVPR.2016.90，2016年。Huang, G.、Liu, Z.、Van Der Maaten, L.和Weinberger, K. Q.：《密集连接卷积网络》，载于：IEEE计算机视觉与模式识别会议论文集，第4700–4708页，2017年。Huang, X.和Belongie, S. J.：《利用自适应实例归一化实现实时任意风格迁移》，载于：IEEE国际计算机视觉会议，ICCV 2017，意大利威尼斯，2017年10月22日至29日，第1510–1519页，IEEE计算机学会，https://doi.org/10.1109/ICCV.2017.167，2017年。Karras, T.、Laine, S.和Aila, T.：《用于生成对抗网络的基于风格的生成器架构》，《IEEE模式分析与机器智能汇刊》，43，4217–4228，https://doi.org/10.1109/TPAMI.2020.2970919，2021年。Krizhevsky, A.、Sutskever, I.和Hinton, G. E.：《采用深度卷积神经网络进行ImageNet分类》，《ACM通讯》，60，84–90，https://doi.org/10.1145/3065386，2017年。Li, C.和Wand, M.：《采用马尔可夫生成对抗网络的预计算实时纹理合成》，载于：计算机视觉——ECCV 2016——第14届欧洲计算机视觉会议，荷兰阿姆斯特丹，2016年10月11日至14日，论文集，第III部分，由Leibe, B.、Matas, J.、Sebe, N.和Welling, M.编辑，《计算机科学讲义》卷9907，第702–716页，施普林格，https://doi.org/10.1007/978- 3-319-46487-9_43，2016年。

Li, C. 和 Wand, M.：结合马尔可夫随机场与卷积神经网络进行图像合成，载于：2016年IEEE计算机视觉与模式识别会议，CVPR 2016，美国内华达州拉斯维加斯，2016年6月27日至30日，第2479–2486页，IEEE计算机学会，https://doi.org/10.1109/CVPR.2016.272，2016b。Li, Y.、Wang, N.、Liu, J. 和 Hou, X.：揭秘神经风格迁移，载于：第二十六届国际人工智能联合会议论文集，IJCAI 2017，澳大利亚墨尔本，2017年8月19日至25日，由 Sierra, C. 编，第2230–2236页，ijcai.org，https://doi.org/10.24963/ijcai.2017/310，2017。Lin, T.-Y.、Goyal, P.、Girshick, R.、He, K. 和 Dollár, P.：用于密集目标检测的焦点损失，载于：IEEE国际计算机视觉会议论文集，第2980–2988页，2017年。美国国家航空航天局：船底座恒星形成区，https://www.nasa.gov/sites/default/files/styles/full_width/public/thumbnails/image/main_image_star-forming_region_carina_nircam_final-1280.jpg，2022年。Ronneberger, O.、Fischer, P. 和 Brox, T.：U-net：用于生物医学图像分割的卷积网络，载于：国际医学图像计算与计算机辅助干预会议，第234–241页，施普林格，2015年。Russakovsky, O.、Deng, J.、Su, H.、Krause, J.、Satheesh, S.、Ma, S.、Huang, Z.、Karpathy, A.、Khosla, A.、Bernstein, M. S.、Berg, A. C. 和 Fei-Fei, L.：ImageNet大规模视觉识别挑战赛，CoRR，abs/1409.0575，http://arxiv.org/abs/1409.0575，2014年。Salehi, S. S. M.、Erdogmus, D. 和 Gholipour, A.：使用三维全卷积深度网络进行图像分割的Tversky损失函数，载于：医学影像机器学习国际研讨会，第379–387页，施普林格，2017年。Schreck, J. S.、Gantos, G.、Hayman, M.、Bansemer, A. 和 Gagne, D. J.：全息图像的神经网络处理，《大气测量技术》，15，5793–5819，2022年。Shao, S.、Mallery, K.、Kumar, S. S. 和 Hong, J.：用于三维粒子场成像的机器学习全息术，《光学快报》，28，2987–2999，https://doi.org/10.1364/OE.379480，2020年。Shelhamer, E.、Long, J. 和 Darrell, T.：用于语义分割的全卷积网络，《IEEE模式分析与机器智能汇刊》，39，640–651，https://doi.org/10.1109/TPAMI.2016.2572683，2017年。Shimobaba, T.、Takahashi, T.、Yamamoto, Y.、Endo, Y.、Shiraki, A.、Nishitsuji, T.、Hoshikawa, N.、Kakue, T. 和 Ito, T.：使用深度神经网络进行数字全息粒子体积重建，《应用光学》，58，1900–1906，https://doi.org/10.1364/AO.58.001900，2019年。Simonyan, K. 和 Zisserman, A.：用于大规模图像识别的超深卷积网络，arXiv预印本 arXiv:1409.1556，2014年。Simonyan, K. 和 Zisserman, A.：用于大规模图像识别的超深卷积网络，载于：第三届国际学习表征会议，ICLR 2015，美国加利福尼亚州圣迭戈，2015年5月7日至9日，会议论文集，由 Bengio, Y. 和 LeCun, Y. 编，http://arxiv.org/abs/1409.1556，2015年。Spuler, S. M. 和 Fugal, J.：用于机载云测量的内联数字全息成像系统设计，《应用光学》，50，1405–1412，2011年。Tan, M. 和 Le, Q.：Efficientnet：重新思考卷积神经网络的模型缩放，载于：国际机器学习会议，第6105–6114页，PMLR，2019年。Ulyanov, D.、Lebedev, V.、Vedaldi, A. 和 Lempitsky, V. S.：纹理网络：纹理和风格化图像的前馈合成，载于：第33届国际机器学习会议论文集，ICML 2016，美国纽约市，2016年6月19日至24日，由 Balcan, M. 和 Weinberger, K. Q. 编，第48卷《JMLR研讨会与会议论文集》，第1349–1357页，JMLR.org，http://proceedings.mlr.press/v48/ulyanov16.html，2016年。Wang, Z.、Bovik, A.、Sheikh, H. 和 Simoncelli, E.：图像质量评估：从误差可见性到结构相似性，《IEEE图像处理汇刊》，13，600–612，https://doi.org/10.1109/TIP.2003.819861，2004年。

Wilmot,P.、Risser, E. 和 Barnes, C.：利用直方图损失实现稳定且可控的神经纹理合成与风格迁移，CoRR，abs/1701.08893，http://arxiv.org/abs/1701.08893，2017年。Zhang, R.、Isola, P.、Efros, A. A.、Shechtman, E. 和 Wang, O.：深度特征作为感知度量具有非凡效能，载于：2018年IEEE计算机视觉与模式识别会议，CVPR 2018，美国犹他州盐湖城，2018年6月18日至22日，第586–595页，计算机视觉基金会 / IEEE计算机学会，https://doi.org/10.1109/CVPR.2018.00068，2018年。Zhang, Y.、Zhu, Y. 和 Lam, E. Y.：使用单阶段网络进行全息三维粒子重建，《应用光学》，61，B111–B120，2022年。