第 44 卷 第 10 期/2024 年 5 月/光学学报 研究论文

光学学报

基于U-Net的压缩光场显示图案生成方法

高晨[1,2]，谭小地[3,4,5*]，李海峰[6]，刘旭[6]

[1]福建师范大学光电与信息工程学院，福建 福州 350117；
[2]福建省光子技术重点实验室，福建 福州 350117；
[3]医学光电科学与技术教育部重点实验室，福建 福州 350117；
[4]福建省光电传感应用工程技术研究中心，福建 福州 350117；
[5]福建师范大学信息光子学研究中心，福建 福州 350117；
[6]浙江大学光电科学与工程学院，浙江 杭州 310027

摘要 压缩光场显示具有结构简单紧凑、显示空间分辨率高的优点，但求解压缩光场显示图案的迭代算法存在计算量大的问题。随着人工智能技术的发展，基于深度学习的图像生成算法也被应用到三维显示中。提出一种将计算机视觉中执行图像分割任务的U-Net作为优化压缩光场显示图案的网络模型。根据给定的观看角度生成几组经过数据增强的目标光场数据集作为U-Net的训练集；在U-Net收敛后，将训练完成的U-Net用于生成重建测试目标光场的显示图案。训练和测试结果表明，相比基于堆叠CNN和迭代算法的方法，所提出的基于U-Net的压缩光场显示图案生成方法具有重建质量更高、计算资源少的优势。

关键词 物理光学；成像系统；压缩光场显示；光场渲染；深度学习

中图分类号 TN27 **文献标志码** A **DOI**：10.3788/AOS231683

1 引 言

在观看手机、平板电脑等握于手中的便携式显示设备时，人们更倾向于正对着显示系统观看，并且常常不希望自己观看的内容泄露给公共场合中的其他人。与显示有关的人体工程学研究指出，当观看5～10 inch（1 inch=2.54 cm）的显示设备时，用户偏好通过调整设备的姿态使得观看角度保持在10°以内[1]。同时，便携式显示设备为了续航需求，其计算性能的释放有限。目前，适用于便携式设备的三维显示技术可分为指向背光显示、压缩光场显示、集成成像显示和指向光场显示等[2]。根据显示原理，指向背光、集成成像和指向光场等方法都需要导光精度极高的光束偏折器件。压缩光场利用显示面板的散射特性和三维场景视点图像之间的相关性，把光场图像压缩为二维图像堆栈。压缩光场显示系统的组成结构仅为多层显示面板，无需光束偏折器件。因此，压缩光场显示具有观看角度适中、结构简单紧凑的优点，能够很好地满足便携式三维显示设备的应用场景需求[3]。并且，压缩光场显示系统还可以显示高动态范围或者超分辨的2D图像[4-5]，易于实现2D/3D可切换的便携式显示。但是，求解压缩光场显示系统的显示图案通常基于逐像素的迭代算法，庞大的计算量是其应用于便携式设备的主要障碍。

人工神经网络（ANN）基于相互连接的人工神经元，神经元之间的连接由线性和非线性运算完成。ANN在数学上被证明可以将任何连续函数近似到所需的精度[6]。随着人工智能技术的发展，基于深度学习的图像生成算法也逐渐被应用于三维显示中[7-14]并且有如下优点：基于深度学习的显示图像生成算法能够弥补实际和理想光学系统的偏差，改善显示系统的图像质量[14]；对于需要迭代计算生成显示图案的三维显示系统，深度神经网络可以通过训练拟合出迭代计算过程，利用人工神经网络快速的前向传播减少对计算资源的消耗[7]。针对压缩光场显示的图案生成问题，日本名古屋大学的Fujii团队[15]提出利用人工神经网络以极低的计算消耗生成压缩光场显示系统的图案，使得压缩光场显示应用于便携式设备成为可能。但是，其采用的堆叠卷积神经网络（CNN）存在不易收敛、容易过拟合等问题[16]。德国马克斯-普朗克信息研究所的Zheng等[17]提出一种基于神经辐射场（NeRF）的曲面压缩光场显示层生成方法。实际上，基于NeRF的方法学习的是单一场景的体素信息，而不是显示系统的结构信息[18]，一旦显示场景变化就需

收稿日期：2023-10-20；**修回日期**：2023-11-29；**录用日期**：2023-12-01；**网络首发日期**：2023-12-12

基金项目：国家自然科学基金（U22A2080）、国家重点研发计划（2018YFA0701800）、福建省科技重大专项（2020HZ01012）

通信作者：*xtan@fjnu.edu.cn

要重新训练网络，所以该方法不适用于动态显示设备。文献[19]中提出一种基于深度校准的压缩光场显示图案学习算法，但是该方法仍需要迭代算法求解的显示图案作为先验，其计算时间为迭代算法和网络训练推理过程所花费的时间总和，因此该方法同样难以用于动态实时的压缩光场显示。本文使用计算机视觉领域中执行图像分割任务的U-Net模型学习压缩光场显示系统的结构参数，生成压缩光场显示图案。U-Net相比传统堆叠CNN具有容易收敛、泛化性高等优点[20]。根据给定的观看角度生成几组经过数据增强的目标光场数据集作为U-Net的训练集；在U-Net收敛后，可将训练完成的U-Net用于直接生成重建测试目标光场的显示图案。本文所使用的方法可用于生成不同目标光场的显示图案，无需针对不同三维场景逐一训练网络。训练和测试的结果表明，相比基于堆叠CNN的方法，所提出的基于U-Net的压缩光场显示图案生成方法具有重建质量高、计算资源少的优点。

2 基于深度学习的压缩光场显示原理

2.1 压缩光场显示

压缩光场显示的基本结构如图1所示，显示层的每个像素由于被散射屏散射或者被散射背光源照明，从而具有一定的散射角，使得前层显示图案的像素能够在散射角的范围内被后层显示图案的像素所调制，最终到达视点位置形成视点图像。通过时序方式在人眼的视觉暂留时间内显示多帧图案可能会提升重建光场的质量，人眼感受到的是多帧重建光场的光强平均值。压缩光场显示按照显示层像素的调制方式可以分为乘法型、加法型、偏振型3种基本类型，下面将依次介绍其数学模型。为了避免符号混淆，压缩光场显示层的数量均由N表示，显示帧数均由M表示，视点数量均由V表示，对应的计数变量分别为n、m和i。

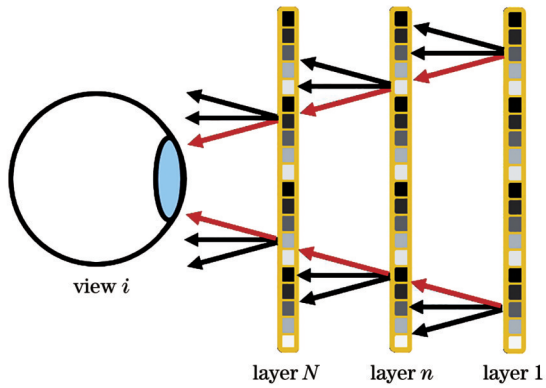

图1 压缩光场显示的原理

Fig. 1 Principle of compressive light field display

如果显示层只是以自身的透过率被动调制来自均匀散射背光源的光线，则将其称为乘法型压缩光场显示。假设背光源的光强均为1，那么发射光场L可以写为多层透过率图案相乘的多帧平均值，即

$$L=\frac{1}{M}\sum_{m=1}^{M}\prod_{n=1}^{N}t_n^m, \tag{1}$$

式中：t_n^m的范围为[0, 1]，表示第m帧第n层图像的透过率。对式(1)进行以自然常数e为底数的等价指对数变换，则有

$$L=\frac{1}{M}\sum_{m=1}^{M}\exp\left(\sum_{n=1}^{N}\xi_n^m\right), \tag{2}$$

式中：$\xi_n^m=\ln(t_n^m)\in(-\infty,0]$，表示第$m$帧第$n$层图像的衰减系数。乘法型压缩光场是最先被提出的压缩光场显示，使用堆叠的印刷胶片[4]或者液晶屏[21]即可实现，具有结构简单、易于搭建的优点。

如果所有显示层都能主动发光或者被照亮，则为加法型压缩光场显示。假设发射的光强范围为[0, 1]，那么发射光场L可以写为多层光强图案相加的多帧平均值，即

$$L=\frac{1}{M}\sum_{m=1}^{M}\sum_{n=1}^{N}l_n^m, \tag{3}$$

式中：l_n^m表示第m帧第n层图像的光强。利用投影显示的方式将显示图案投射到透明的散射屏上，即为投影型加法光场显示[22]。因为投影散射屏不具有周期性的像素结构，所以投影型加法光场显示相比乘法型的优点是，能够有效地减弱像素结构带来的衍射效应，并且显示亮度更高。

如果把乘法型压缩光场显示的多层液晶屏中间的偏振片都拆掉，只保留最前和最后两层偏振片，那么多层结构将会对入射偏振光的相位进行调制，该方式被称为偏振型压缩光场显示。偏振型压缩光场显示的出射光强L符合马吕斯定律，即服从以累加相位为变量的$\sin^2(\cdot)$分布[23]：

$$L=\frac{1}{M}\sum_{m=1}^{M}\sin^2\left(\sum_{n=1}^{N}\theta_n^m\right), \tag{4}$$

式中：θ_n^m表示第m帧第n层图像的相位值，一般液晶屏像素可调制的相位范围为[0, π/2]。移除了中间的偏振片，偏振型压缩光场显示相比乘法型提高了显示亮度，但是直接堆叠显示屏的结构仍然会使它受到衍射效应的限制。

以上3种基本类型的压缩光场显示模型的数学表达可统一为

$$L=\frac{1}{M}\sum_{m=1}^{M}f(x_n^m),n=1,2,\cdots,N, \tag{5}$$

式中：x_n^m表示第m帧第n层图像的显示层像素；$f(\cdot)$表示每个显示层像素对光线的调制方式。可以把所有层的$f(\cdot)$运算传播到目标光场平面的过程用投影矩阵$\boldsymbol{A}$和显示层相乘的形式表示，即

$$L=\frac{1}{M}\sum_{m=1}^{M}f(\boldsymbol{A}\boldsymbol{x}^m), \tag{6}$$

式中：每个显示层的像素数量为$w\times h$，第m帧的所有

显示层上的像素被拼接拉伸成长度为$w\times h\times N$的列向量$\boldsymbol{x}^m$。已知目标光场的视点数量为V，通常设置每幅视点图像的像素数量与每个显示层相同，那么投影矩阵的元素数量为$(w\times h)^2\times V\times N$。如图1所示，在显示层上只有符合直线传播定律的像素才能形成红色的有效光线，因此投影矩阵可以通过每幅视点图像向多个显示层的反向光线追迹得到。投影矩阵中的元素布局如图2所示，灰色元素表示穿过显示层的有效像素。因为每幅视点图像的每个像素只能由穿过多个显示层的一根光线确定，所以投影矩阵每行只有N个元素为非空，其余元素皆为空。非空元素的个数为$(w\times h)\times V\times N$，远小于矩阵的元素总数，因此投影矩阵是一个稀疏矩阵。

多帧重建光场之间的累加也可以写成单帧重建光场的矩阵与一个元素全为1的列向量相乘的形式[24]。引入一个维度为M且元素全为1的列向量$\boldsymbol{E}$，则式(6)可以写为

$$\boldsymbol{L}=\frac{1}{M}f(\boldsymbol{A}\boldsymbol{x}^m)\boldsymbol{E}。\tag{7}$$

最终求解显示图案的优化问题转化为求解发射光场和目标光场之间的误差最小值：

$$x=\mathrm{argmin}\left\|f(\boldsymbol{A}\boldsymbol{x})\boldsymbol{E}-M\boldsymbol{b}\right\|,\tag{8}$$

式中：b表示以列向量方式排列的目标光场；$\|\cdot\|$表示L2范数。对于上述3种基本调制方式的压缩光场显示，其优化目标和优化变量及其范围见表1。

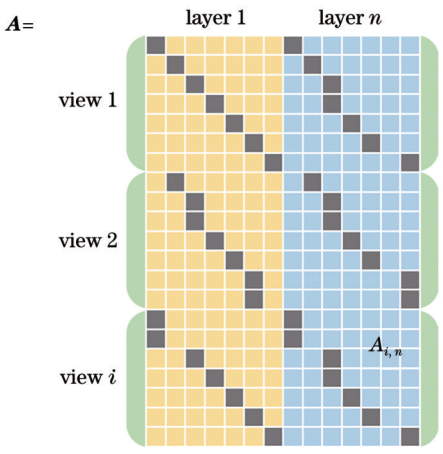

图2　投影矩阵的布局示意

Fig. 2　Schematic of projection matrix layout

表1　3种基本类型压缩光场显示的优化目标和变量范围

Table 1　Optimization objective and variable range for three basic types of compressive light field displays

Type of compressive light field displays	Optimization objective	Variable range
Multiplicative	$\mathrm{argmin}\left\|\exp(\boldsymbol{A}\boldsymbol{x})\boldsymbol{E}-M\boldsymbol{b}\right\|$	$x=\log t\in(-\infty,0]$
Additive	$\mathrm{argmin}\left\|\boldsymbol{A}\boldsymbol{x}\boldsymbol{E}-M\boldsymbol{b}\right\|$	$x=t\in[0,1]$
Polarized	$\arg\min\left\|\sin^2(\boldsymbol{A}\boldsymbol{x})\boldsymbol{E}-M\boldsymbol{b}\right\|$	$x=\theta\in[0,\pi/2]$

将压缩光场显示的图案生成过程抽象为数值优化问题后，就可以使用各种优化算法对其进行求解。受限于显示图像源的体积和刷新率，压缩光场显示系统的显示层数和显示帧数一般只有几层和几帧，而需要重建的视点图像数量往往有几十幅。不难发现，压缩光场显示图案的未知量个数为$(w\times h)\times M\times N$，远远小于作为已知量的目标光场元素数。因此，这是一个求解方程组数量远大于未知量个数的大规模超定非线性问题。经典的高斯消元法、QR分解算法无法执行如此大规模的运算，只能使用迭代算法求解。除了单纯从数值优化角度出发的迭代算法，还有基于显示系统的物理模型迭代算法，包括针对乘法型压缩光场显示的非负矩阵分解算法(NMF)[25]、针对加法型和偏振型压缩光场显示的联立代数重建技术(SART)[26]。

压缩光场显示系统的观看角度、图像质量和显示深度等性能参数之间存在折中关系。在显示深度范围一定的前提下，观看角度和图像质量相互制约。三显示层结构的压缩光场显示系统的实用观看角度一般为10°×10°。通过增加显示层数来提升显示带宽是提升压缩光场显示所有性能参数最直接的办法，具体的方式有空间复用、时间复用和偏振复用等。空间复用是直接在空间上加入更多分立的显示屏[27]，如合肥工业大学的吕国强团队[28]通过堆叠6层散射屏实现了30°×30°的较大观看角度；时间复用利用高刷新率的光学器件配合显示图像源，在人眼的视觉暂留时间内将显示图案成像到更深的范围，如浙江大学的刘旭团队[29]使用变焦透镜将不同深度的乘法型压缩光场显示以时序方式叠加；偏振复用是通过偏振选择器件将不同偏振态的显示图案成像到更深的范围，如韩国仁荷大学的Park团队[30]在乘法型压缩光场显示的基础上，增加了偏振控制板、四分之一波片、半反射镜和反射式偏振片，复用正交的偏振态照明光束将显示深度扩大了2倍。相比空间复用，时间复用和偏振复用不会导致显示系统的体积和成本明显增加，是效率更高的复用方式。本文把这种混合了多种调制方式的压缩光场显示称为混合型压缩光场显示。

2.2　基于深度学习的图像生成

基于ANN的压缩光场显示图案生成原理如图3所示，网络的训练过程分为前向传播和后向传播两个

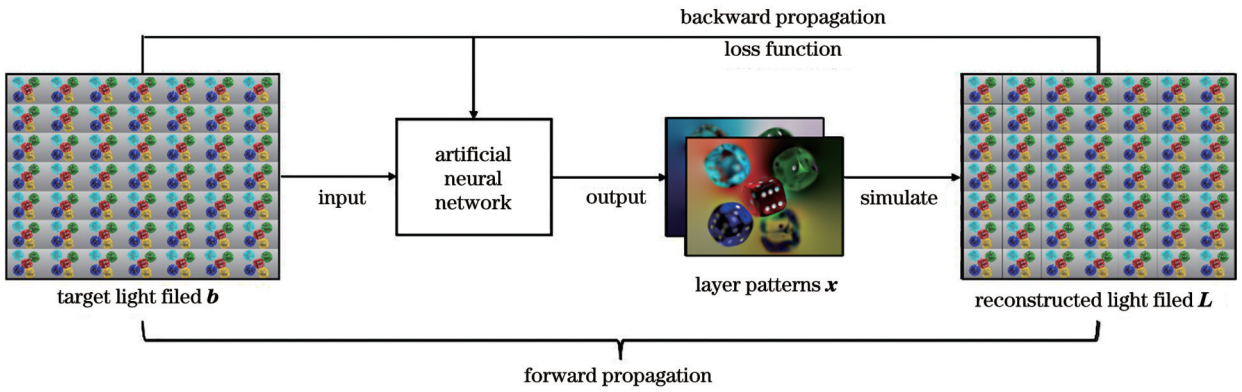

图 3　基于 ANN 的压缩光场显示图案生成原理

Fig. 3　Principle of image synthesis for compressive light field displays based on ANN

过程。前向传播的过程包括：首先，将作为训练数据的目标光场输入到人工神经网络，并输出多层显示图案；然后，利用透视投影变换仿真多层显示图案的重建光场。反向传播过程就是通过以重建光场与目标光场为变量的损失函数更新人工神经网络参数的过程。在每个训练回合（epoch）和批次（batch）都重复以上过程。在网络完成训练后，即可将作为测试数据的目标光场输入到网络并输出显示图案，上述过程即为网络的推理过程。采用不同结构的网络模型的训练和推理流程都是类似的，因此训练和推理效果的影响因素就是所选取的网络结构和训练超参数。

使用 ANN 生成压缩光场显示图案的优点是，若要达到相同的重建光场质量，训练效果好的网络模型的推理时间将远远短于迭代计算的时间。另外，基于数值优化的迭代算法只能生成光线调制方式单一的压缩光场显示图案，无法生成混合型压缩光场显示图案。比如，针对乘法型压缩光场显示亮度较低的问题，可以在双层液晶屏 t_1 和 t_2 之间插入一层自发光散射屏 l_1，那么重建光场的表达式 L 即为

$$L=(t_1+l_1)\times t_2。\tag{9}$$

对于混合型压缩光场显示图案的求解，韩国首尔大学的 Lee 团队[27]提出一种基于 NMF 和 SART 的经验迭代算法，即假设几个中间态重建光场，这些中间态重建光场只由单一调制方式的显示层构成。其迭代流程如下：先把最终重建光场分解为几个中间态重建光场，随后将中间态的重建光场代入 NMF 或者 SART 算法分解得到显示图案，再由显示图案计算最终重建光场，重复以上过程直到显示图案收敛。这种经验迭代算法需要针对混合型压缩光场显示设备的具体结构而设计，一旦显示层的调制方式或者数量发生变化，就需要重新设计迭代算法，在实际应用中十分不便。对于 ANN 的图案生成算法，只需要改变训练和推理模块中的透视投影变换仿真部分，就能适应混合型压缩光场显示的结构变化。

Fujii 团队[15]提出的用于压缩光场显示图案生成的堆叠 CNN 如图 4 所示，其主体部分由 19 个相同的模块堆叠构成，每个模块包含一个 64 通道的二维卷积操作和一个整流线性单元（ReLU）激活函数。网络输入层的通道数为目标光场的视点数量，网络输出层的通道数为显示图案的层数。输入层和输出层的通道数分别与目标光场的视点数目和显示图案的层数匹配。输入层、堆叠网络和输出层之间均由一个二维卷积操作和一个 ReLU 激活函数连接。网络中所有的卷积操作的卷积核大小均为 3×3，并在卷积操作后的图像边缘

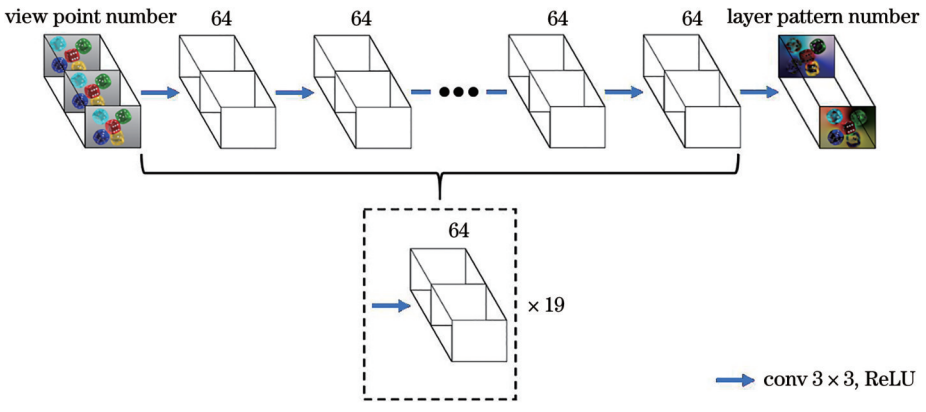

图 4　用于压缩光场显示图案生成的堆叠卷积神经网络结构

Fig. 4　Architecture of stacked convolution neural network for compressive light field display patterns synthesis

填充零像素，以保证卷积前后图像的大小不变。CNN常用于图像处理的原因在于卷积操作具有空间不变性，使得它能够提取图像中任意位置物体的空间特征。深度CNN通过对图像进行多次的降采样和卷积操作，使得网络能够学习到不同尺度下的空间特征，最后在网络的输出层聚合这些特征，对图像的所有层次进行预测，输出检测结果。这种堆叠结构的CNN通常用于图像分类，这通常意味着将图像减少到单个数字表示的“标签”或“类别”。但是这种堆叠的CNN存在梯度消失问题，即前层的参数不能通过反向传播向后层正确更新[16]，原因是ANN的反向传播是通过多元函数的链式求导法则实现的。当误差梯度反向传播到较靠前的层时，重复的乘法会使梯度变小。因此，网络中的层越多，其性能就先趋近于饱和再迅速下降。

为了解决这种堆叠CNN存在的梯度消失问题，针对各种任务目标的其他网络结构被提出，U-Net就是其中一种用于图像分割的网络[20]。图像分割是指将感兴趣的区域从图像中分割出来。压缩光场显示和CT成像具有密切的联系。CT成像利用多个方向发射的X射线环绕样本进行扫描，将获得的样本进行算法处理后得到样本不同深度平面对X射线的吸收系数。在CT成像中，U-Net被用于CT扫描切片数据的处理，根据相邻切片之间的衰减系数变化输出器官的癌变概率。器官癌变的概率是由两幅分别表示器官是否癌变的双通道概率分布图表示的。CT成像的切片数据形式和三维显示中的光场数据具有相似的形式。如果把U-Net的输入层改为目标光场，输出层改为多层显示图案，就得到一个可用于压缩光场显示图案生成的U-Net。如图5所示，U-Net的结构类似一个字母U，可分为从输入层到网络底部的下路和从网络底部到输出层的上路。U-Net的基本操作可以分为卷积、池化和转置卷积。目标光场经过输入层后通道数变为64，每次卷积池化操作后数据的通道数都会变成之前的2倍。当数据到达网络底部时，经过对称的转置卷积和卷积

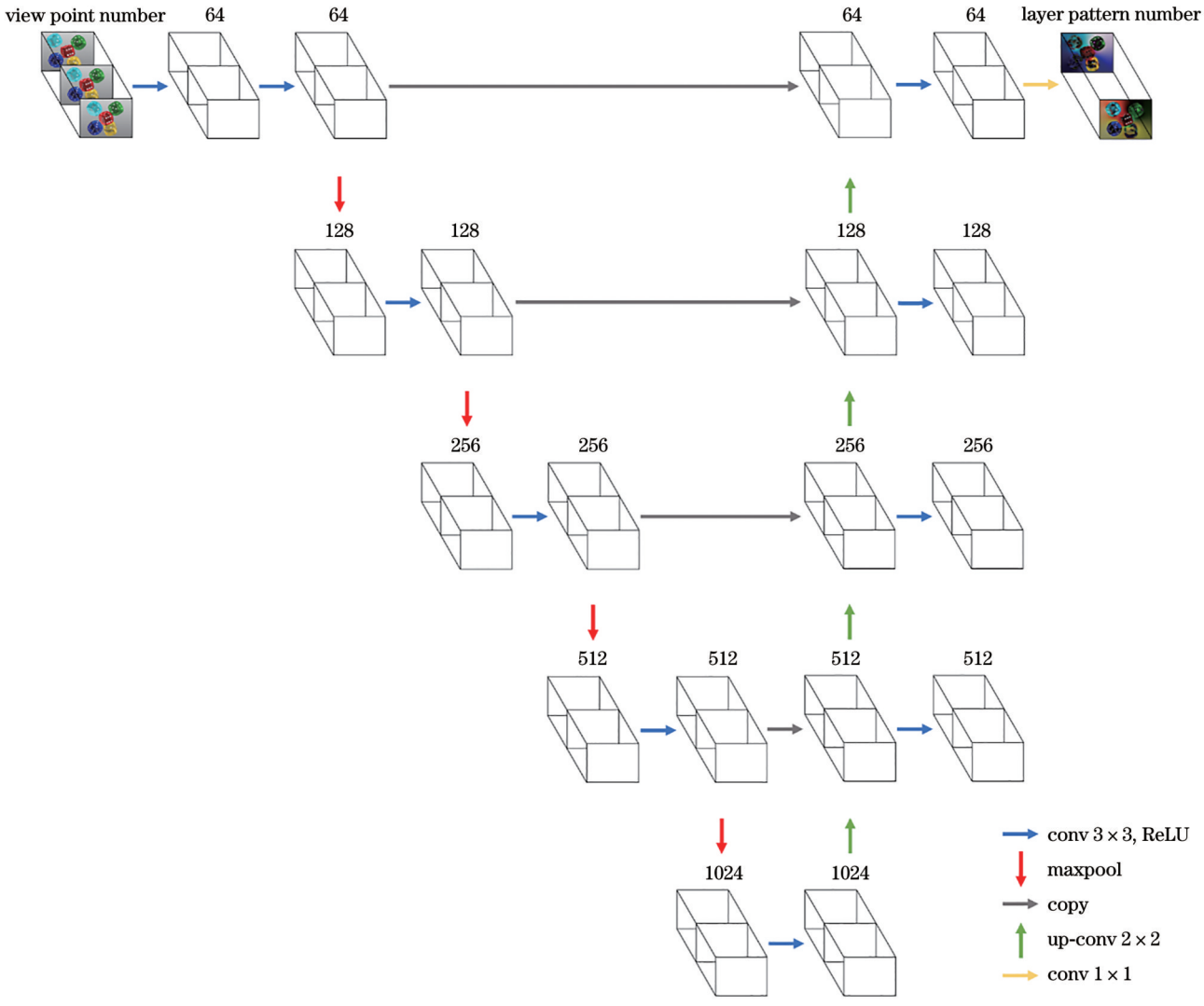

图5 用于压缩光场显示图案生成的U-Net结构

Fig. 5 Architecture of U-Net for compressive light field display patterns synthesis

操作将通道数变回64，最后用一个卷积核为1×1的卷积操作将输出通道与显示图案的层数匹配。U-Net能够避免堆叠CNN梯度消失问题的关键在于上下两路的每个对称模块之间的连接为跳跃连接，这些跳跃连接将下路的参数直接复制给对应的上路。跳跃连接的加入基于以下假设：当处理图像分割任务时，不同层次的分割结果应和不同层次的原始图像具有相似性。压缩光场显示图案和目标光场之间也具有相似性，这正是将U-Net用于压缩光场显示图案生成的理论根据。

3 网络的训练

目前，人工智能技术是由先验知识和数据共同驱动的。先验知识体现在人们根据各种任务的特点设计合适的ANN结构。在训练ANN时，训练数据集的准确性会严重影响模型的结果，这是因为网络会学习训练数据集的偏差，训练数据集的设计也依赖于对任务的先验知识。压缩光场显示图案的生成特点是：由不同角度的投影图像生成不同深度的显示层。这是一种从角度到位置的不变映射关系，还考虑了显示层之间的光线调制方式。从图3可以看到，在网络的训练过程中，对显示层之间的光线调制方式的学习已经由透视投影仿真这一部分承担。因此，在设计训练数据集时，希望网络能够尽量从中学习关于角度到位置的映射信息。U-Net的一个训练技巧是广泛使用数据增强来增加输入图像的数量，从而不需要额外的标记数据。这种数据增强主要是通过“弹性变形”来完成的，这种变形改变了图像上物体的形状，就好像物体处于不同的位置一样。但是这种数据增强方式不适用于本文的训练任务，这是因为当利用旋转、拉伸等变换改变视点图像时，角度到位置之间的映射关系也会发生变化。在网络训练过程中，如果将像素数量较大的目标光场直接作为训练数据输入到网络中，那么输出结果会收敛为灰度和场景背景接近的灰色图案，这说明网络学习的是场景的背景信息。这就启发我们对训练数据集的像素强度进行增强处理，并且减少训练数据的像素数量。另外，对于纹理缓慢变化的物体，其和场景之间的边界更能体现像素的位置变化信息。本实验生成训练数据集的过程如图6所示，其中图6(a)展示了作为训练数据集的目标光场。如图6(b)所示，截取每个视点图像中场景与物体或者物体与物体之间边界的同一空间位置为一个图像块，图像块的像素数量为64×64。将这些图像块拼接成图6(c)所示的图像，每幅图像的行数等于截取空间位置的个数，列数等于目标光场的视点数量。训练数据集由7幅类似图6(c)的图像组成，每幅图像截取的空间位置为10个。通过均匀地对红绿蓝3个颜色通道的图像块施加6种增益或者衰减，最终得到7×10×3×6=1260个训练样本数量。

人工神经网络的超参数是指在训练网络前设置的

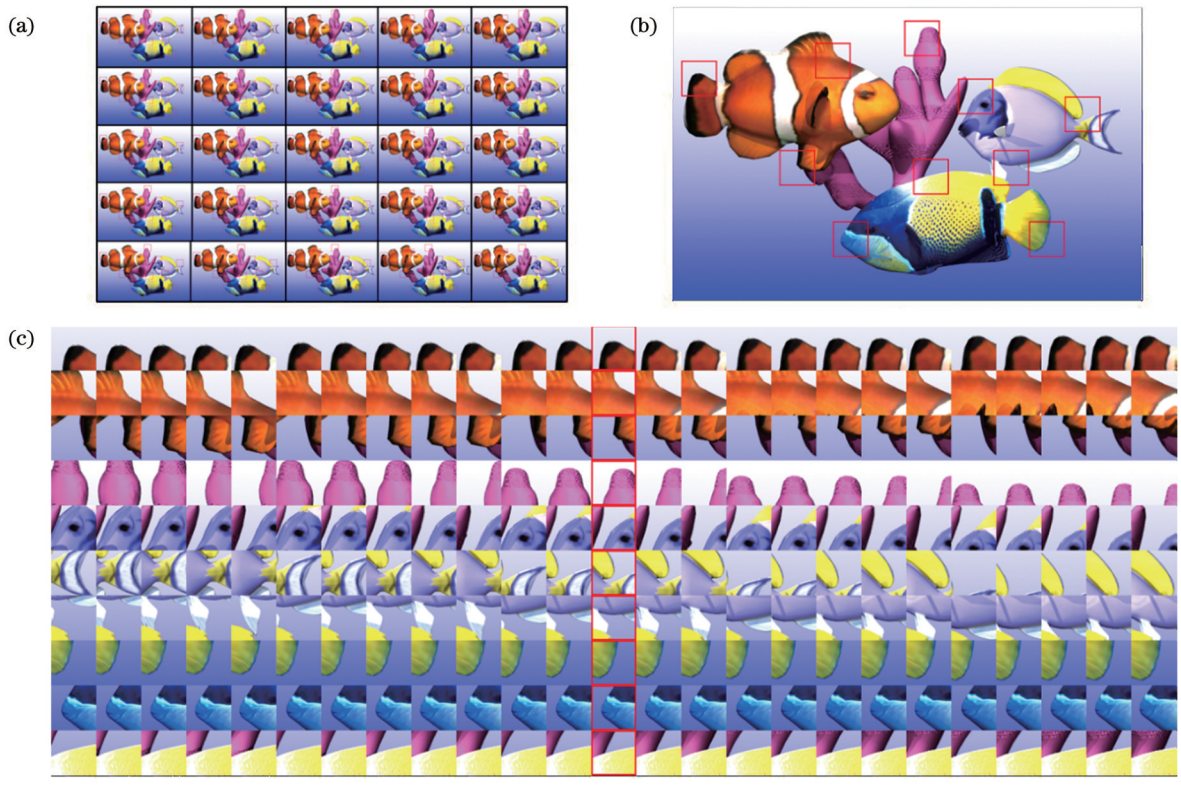

图6 目标光场的训练数据集生成。(a)用于训练的目标光场；(b)视点图像中截取的图像块；(c)图像块拼接的训练数据集

Fig. 6 Training dataset generation of target light field. (a) Target light field for training; (b) image blocks cropped from viewpoint images; (c) training dataset spliced by image blocks

除网络参数外的参数，这些参数不能通过训练获得，需要人为设置。为了公平地与Fujii团队提出的基于堆叠CNN的方法进行比较，本文将除网络结构外的其他训练超参数设为一致，具体包括：1）训练数据集均为上述1260个样本；2）激活函数都为ReLU函数；3）网络参数均使用Kaiming均匀随机初始化；4）优化器选择Adam，学习率设置为较小的10^{-4}以避免训练结果振荡，其他参数均为默认值，且在第一个训练批次结束后，Adam优化器将会自动调整学习率；5）足够多的训练回合保证网络收敛，设置训练的回合数为100，每批次训练的样本数为15；6）损失函数均由重建光场与目标光场的均方差和对显示图案像素值的正则化约束组成。

下面说明设置如上激活函数、初始化参数和损失函数的原因。使用ReLU作为激活函数有助于防止卷积神经网络的计算量呈指数式增长。ReLU激活函数还能避免梯度消失问题，这是因为ReLU的导数始终为常数1，意味着无论神经元的输入值是多少，损失函数都将沿着反向传播。从经验来看，网络的初始化参数最好接近于0但不等于0，这样训练出来的网络会具有比较好的泛化性。Kaiming均匀初始化方法就是针对ReLU激活函数设计的[31]，该方法通过严谨地对ReLU激活函数的非线性进行建模，使得超过30层的深度网络仍然能够收敛。在训练人工神经网络中的损失函数后，增加额外的正则项表示基于先验知识的约束。正则项一般以L1或者L2范数的形式表示，并且通过乘以一个系数来控制正则项的惩罚。L2范数是回归问题的常用评价指标。由于图像包含巨大的像素量，因此由输入图像生成输出图像是一个典型的回归问题。压缩光场显示图案的像素值应该具有一定的范围：对于乘法型和加法型，其像素值介于0到1之间；对于偏振型，其像素值介于0到$\frac{\pi}{2}$之间。因此，本文将网络输出的显示图案中像素值超过范围的所有像素的L2范数作为正则项，并且认为这些正则项的重要性和重建光场与目标光场的均方差相同，即损失函数中各项系数都为1。在训练网络过程中，对于网络输出的显示图案，只要其像素值的范围满足上述正则项要求，就能使得损失函数减少到只包含重建光场与目标光场的均方差这一项。

4 训练和测试结果

网络的训练和测试均在配置为Intel i7-4790中央处理单元、32 GB内存、NVIDIA GeForce RTX 2080Ti GPU的工作站上进行。所有代码均由基于Python的深度学习框架PyTorch执行，并使用NVIDIA并行计算框架CUDA加速训练。使用仿真重建光场相较于目标光场的峰值信噪比（PSNR）评估训练和测试过程。将图7所示的测试目标光场输入到

图7 用于测试的目标光场

Fig. 7 Target light field for testing

堆叠CNN和U-Net中测试网络的泛化性，测试目标光场的观看角度、视点数量等参数与训练目标光场相同：观看角度为10°×10°，视点数量为5×5；生成显示图案包括3层，层间距为1 pixel，例如，对于像素间距为0.6 mm的显示面板，层间距即为0.6 mm。基于NMF和SART的迭代算法也由CUDA加速的Python矩阵运算库CuPy实现。

图8～11分别展示了用于不同类型压缩光场显示图案生成的CNN和U-Net的训练和测试PSNR随训练次数增长的变化，其中混合型发射光场的表达式与式(9)相同。可以看到：在训练过程中CNN和U-Net都能够在一定程度上拟合训练数据集，但是U-Net的训练损失更小也更稳定；在测试过程中，由U-Net生成的显示图案重建的光场质量始终比CNN生成的高2 dB以上。以上训练和测试结果说明U-Net相比CNN能够更好地拟合和泛化压缩光场显示图案的生成过程。

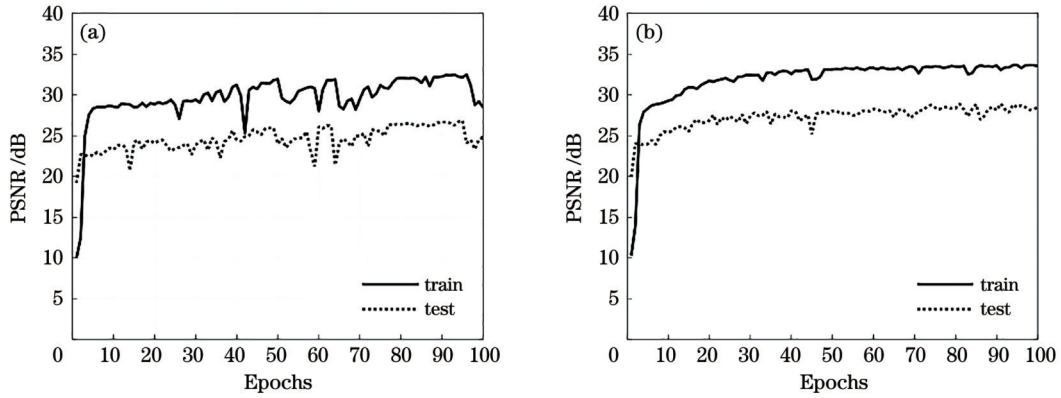

图8 用于乘法型压缩光场显示图案生成的网络训练和测试结果。(a) CNN；(b) U-Net

Fig. 8 Training and testing results of network for multiplicative-type compressive light field display patterns synthesis. (a) CNN; (b) U-Net

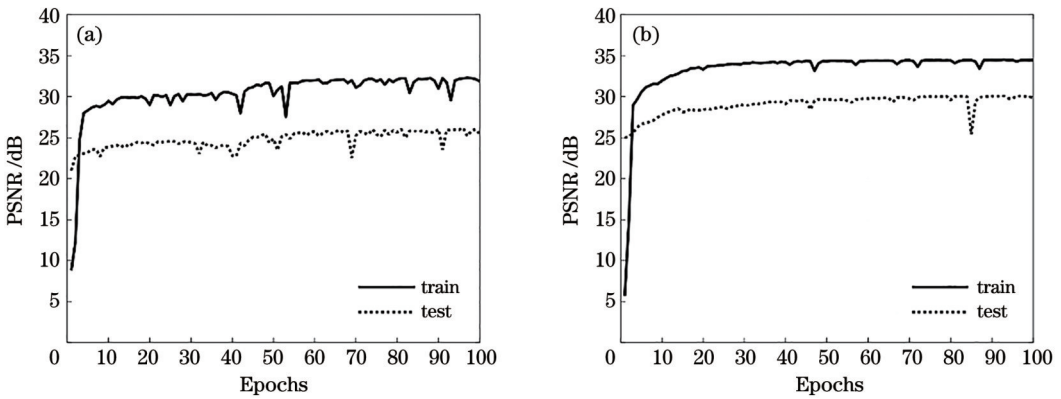

图9 用于加法型压缩光场显示图案生成的网络训练和测试结果。(a) CNN；(b) U-Net

Fig. 9 Training and testing results of network for additive-type compressive light field display patterns synthesis. (a) CNN; (b) U-Net

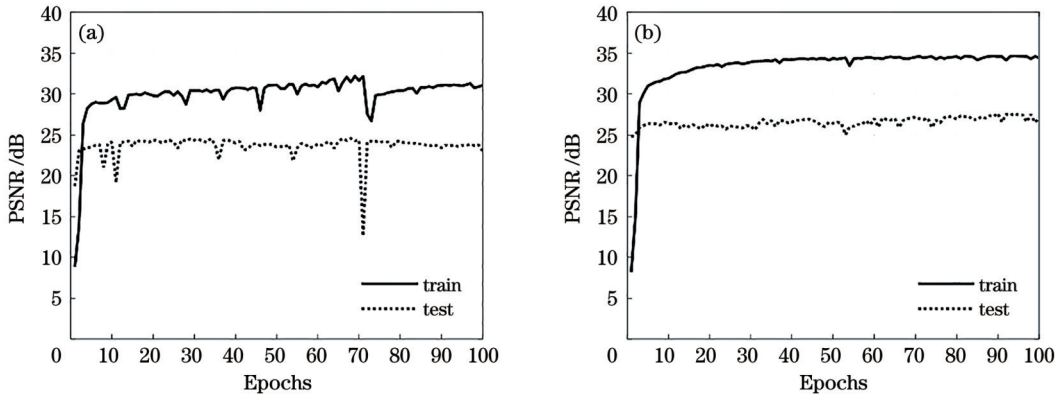

图10 用于偏振型压缩光场显示图案生成的网络训练和测试结果。(a) CNN；(b) U-Net

Fig. 10 Training and testing results of network for polarized-type compressive light field display patterns synthesis. (a) CNN; (b) U-Net

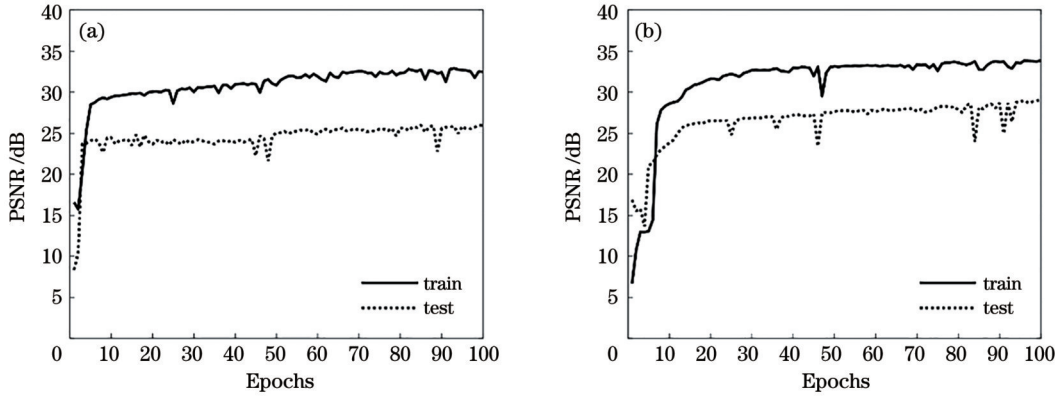

图 11 用于混合型压缩光场显示图案生成的网络训练和测试结果。(a) CNN;(b) U-Net

Fig. 11 Training and testing results of network for hybrid-type compressive light field display patterns synthesis. (a) CNN; (b) U-Net

本文选取测试结果最好的网络模型来生成显示图案、仿真重建结果,并和迭代算法迭代 100 次后生成的图案进行比较。基于 CNN、U-Net 和迭代算法生成的压缩光场显示图案如图 12~14 所示,分别对应乘法型、加法型、偏振型压缩光场显示。显示图案的像素值实际上表示各种类型压缩光场显示的调制量:对于乘法型,该调制量为透过率;对于加法型,该调制量为光强值;对于偏振型,该调制量为相位值。观察以上显示图案的特点可以发现,基于 CNN 和 U-Net 生成的显示图案具有明显差异,但是都符合正则项的像素值范围约束。CNN 倾向于将尽量多的调制量分配给第一层,其余层的调制量逐渐减少并趋于平稳。这很好地体现了 CNN 由整体到局部的特征提取过程。U-Net 倾向于将调制量均匀分配到每一层,这正是 U-Net 相比于堆叠 CNN 多出的上路和跨越连接作用——将提取的整体和局部特征恢复到原来的尺度。基于物理模型的迭代算法的思想是将重建误差平均分配到每一层,这和 U-Net 生成的显示图案类似。无论是 CNN 还是

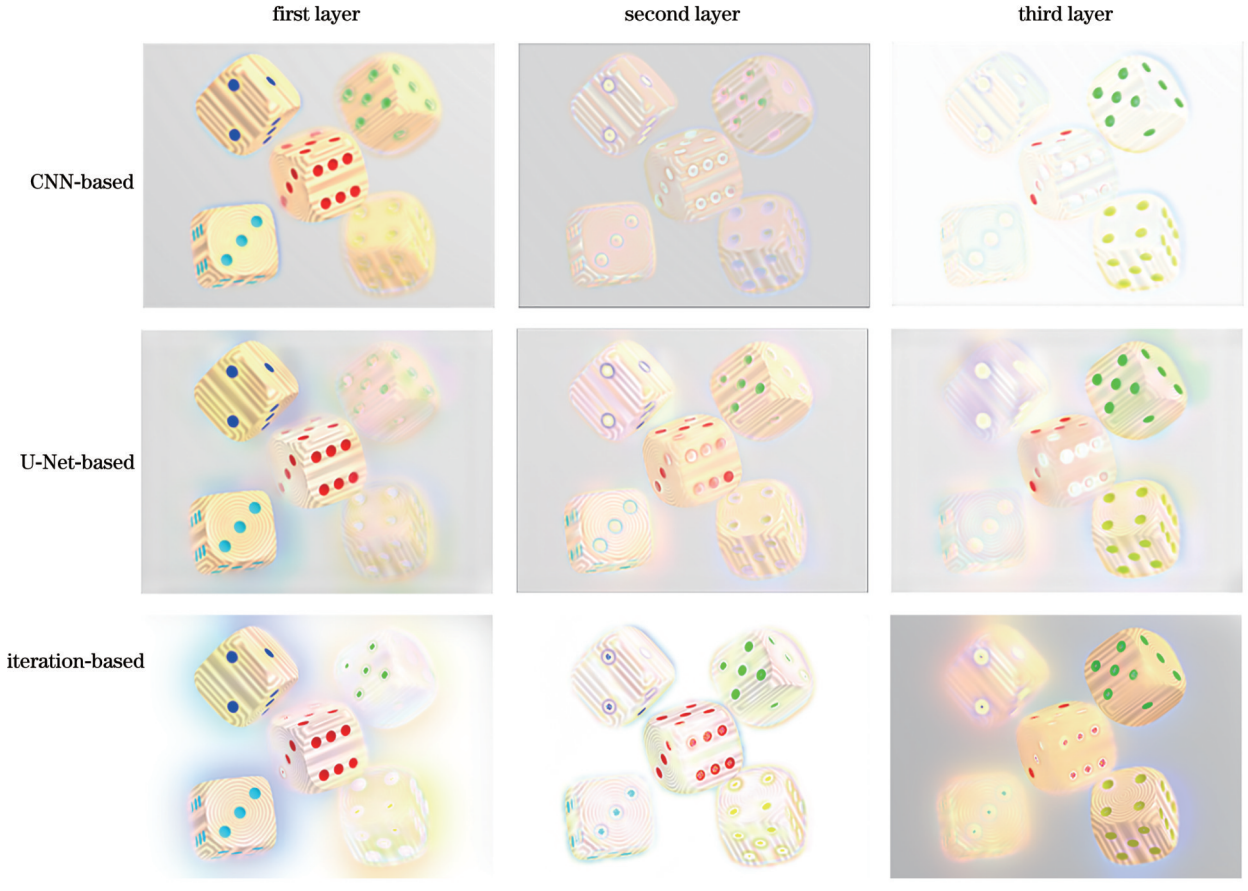

图 12 基于 CNN、U-Net 和迭代算法生成的乘法型压缩光场显示图案

Fig. 12 Multiplicative-type compressive light field display patterns synthesized by CNN, U-Net, and iterative algorithm

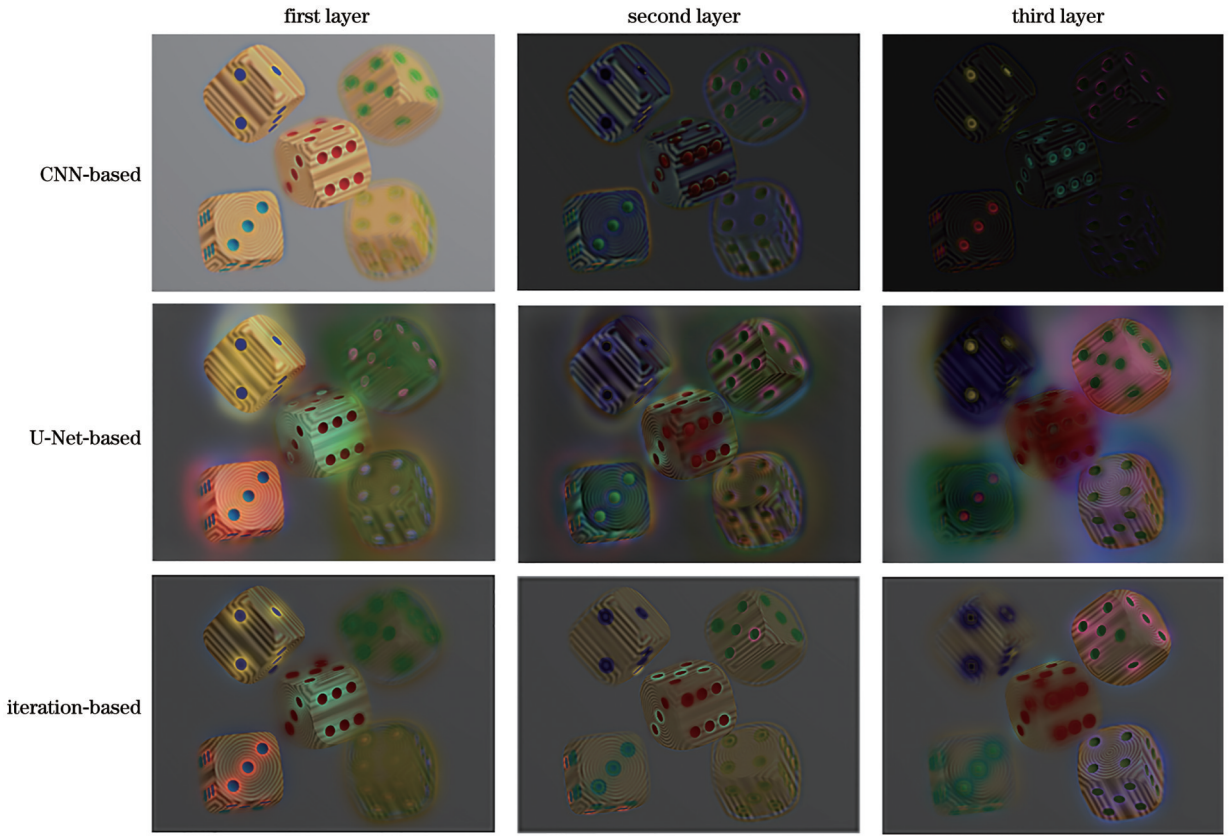

图 13　基于CNN、U-Net和迭代算法生成的加法型压缩光场显示图案

Fig. 13　Additive-type compressive light field display patterns synthesized by CNN, U-Net, and iterative algorithm

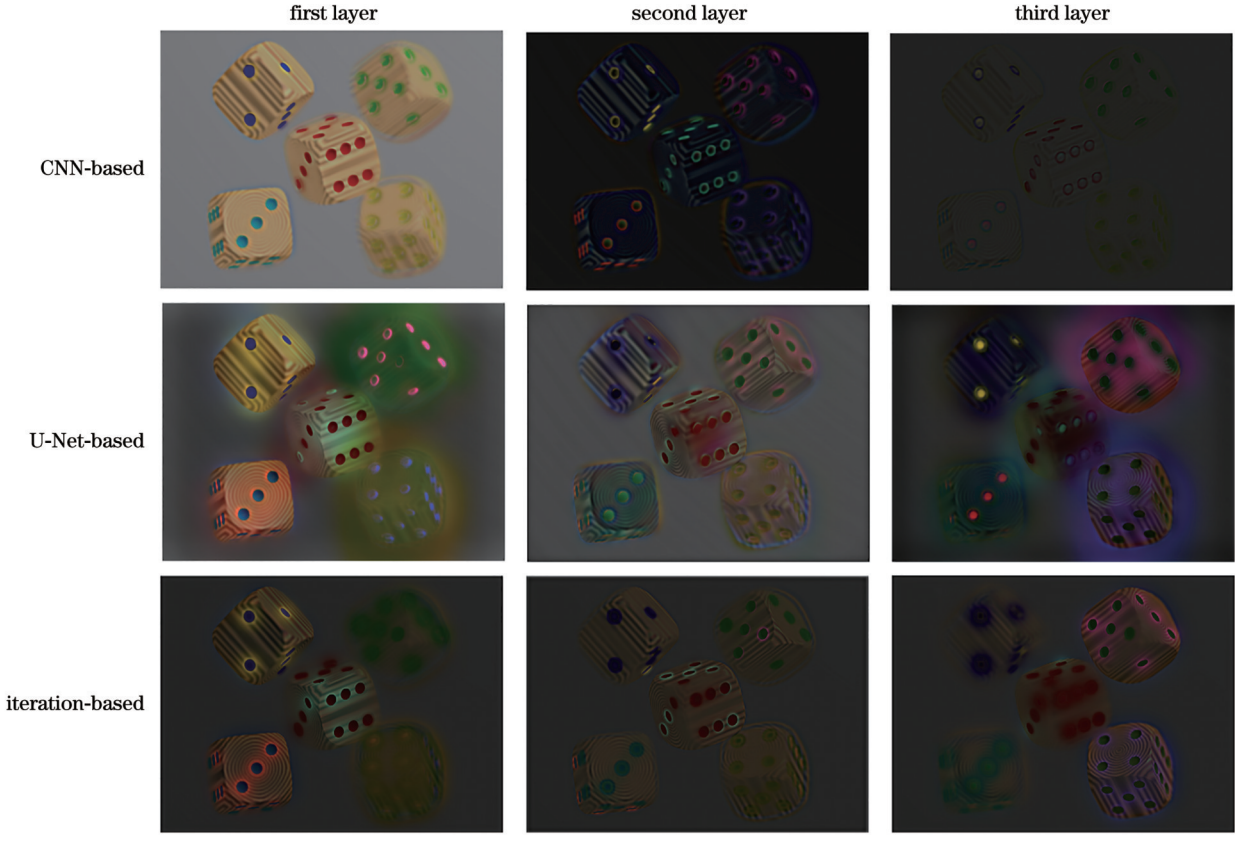

图 14　基于CNN、U-Net和迭代算法生成的偏振型压缩光场显示图案

Fig. 14　Polarized-type compressive light field display patterns synthesized by CNN, U-Net, and iterative algorithm

U-Net，所生成的显示图案都具有与迭代算法类似的按照深度分层的现象。上述现象说明所设计的数据集和损失函数有助于网络结构更好地学习从角度到位置的不变映射关系，而且抑制了网络对目标光场平均光强的学习。

图 15 展示了使用 CNN 和 U-Net 生成的混合型压缩光场显示图案，这是难以使用迭代算法求解的。根据图 12 和图 13 展示的基于 NMF 和 SART 算法求解的乘法型和加法型压缩光场显示图案特征，基于经验迭代算法求解的显示图案 t_2 的平均光强应与目标光场接近，而显示层 t_1 和 l_1 的平均光强之和应与目标光场接近，这样才能使重建光场的平均光强与目标光场接近。可以看到，使用 U-Net 生成的显示图案更符合基于物理直觉的预测，而使用 CNN 生成的每层显示图案的平均光强在各个颜色通道的分布是不均匀的。

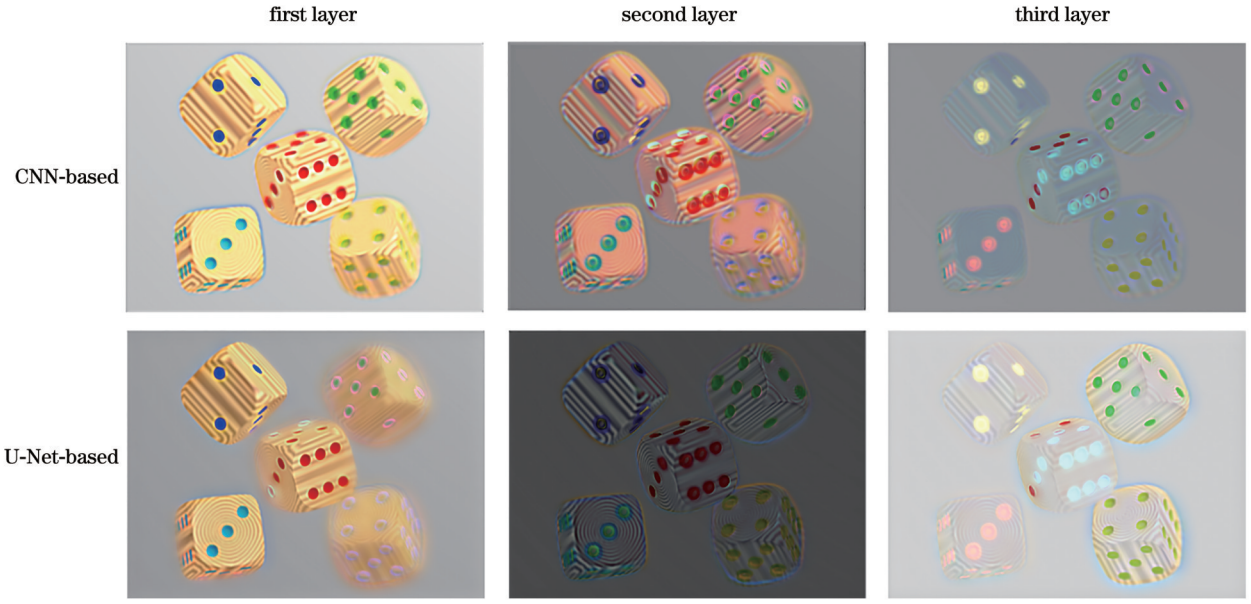

图 15　基于 CNN、U-Net 生成的混合型压缩光场显示图案

Fig. 15　Hybrid-type compressive light field display patterns synthesized by CNN and U-Net

如图 16 所示，本文选取测试目标光场中最左上角的第 1 个、中心的第 13 个和最右下角的第 25 个视点的图像对重建光场的质量进行说明，这 3 幅视点图像之间具有明显的视差，例如位于骰子间隙的方框图像。

图 17～20 分别展示了基于以上乘法型、加法型、

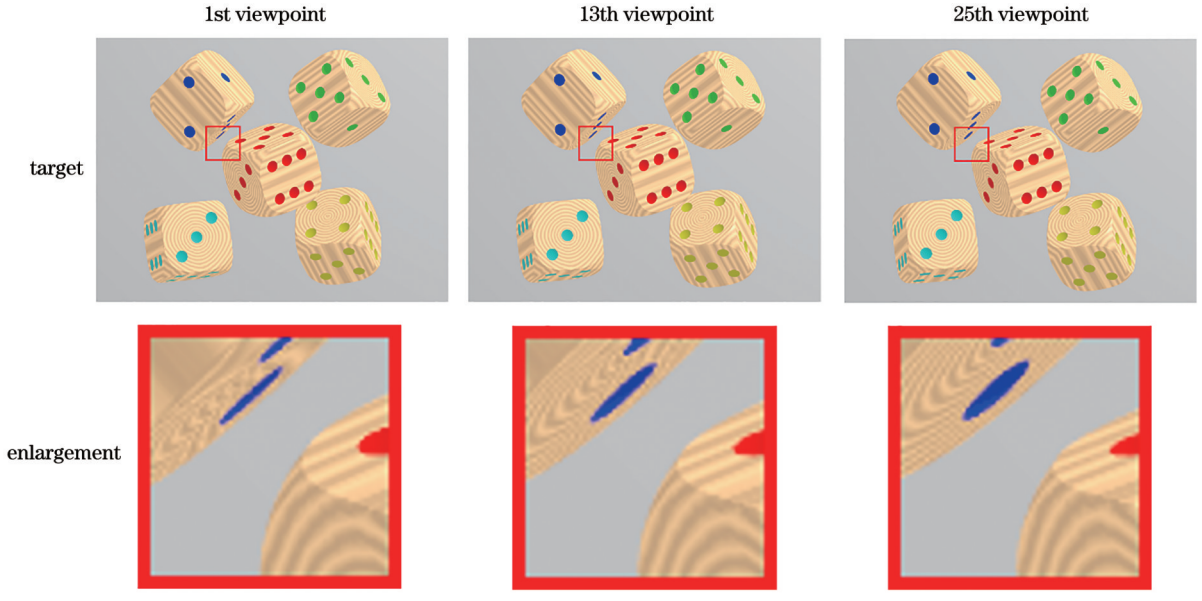

图 16　用于测试的目标光场的视点图像及其局部放大

Fig. 16　Viewpoint images of target light field and their enlargements

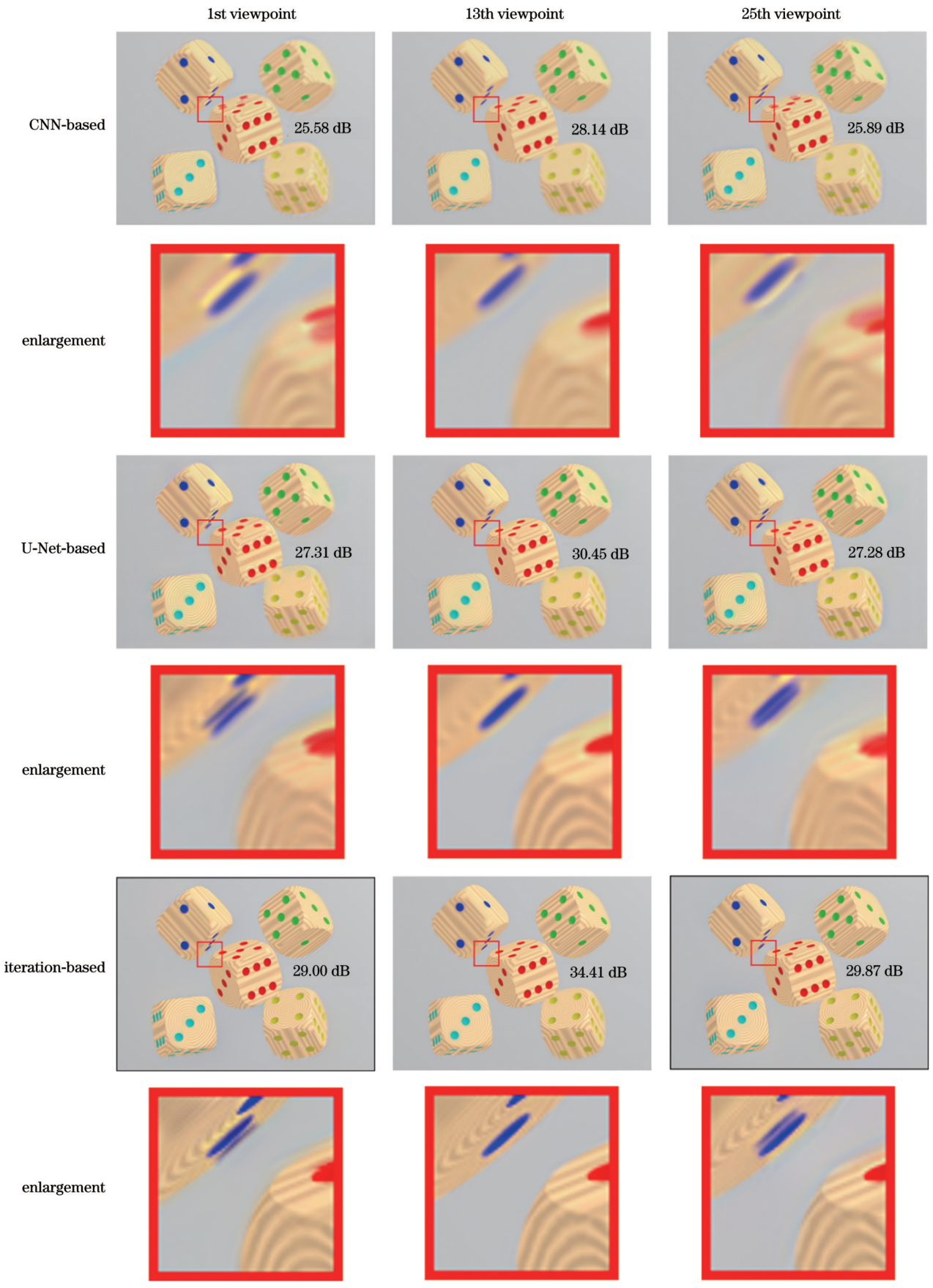

图 17　乘法型压缩光场显示的视点图像仿真重建结果

Fig. 17　Simulated reconstruction results of viewpoint images by multiplicative-type compressive light field display

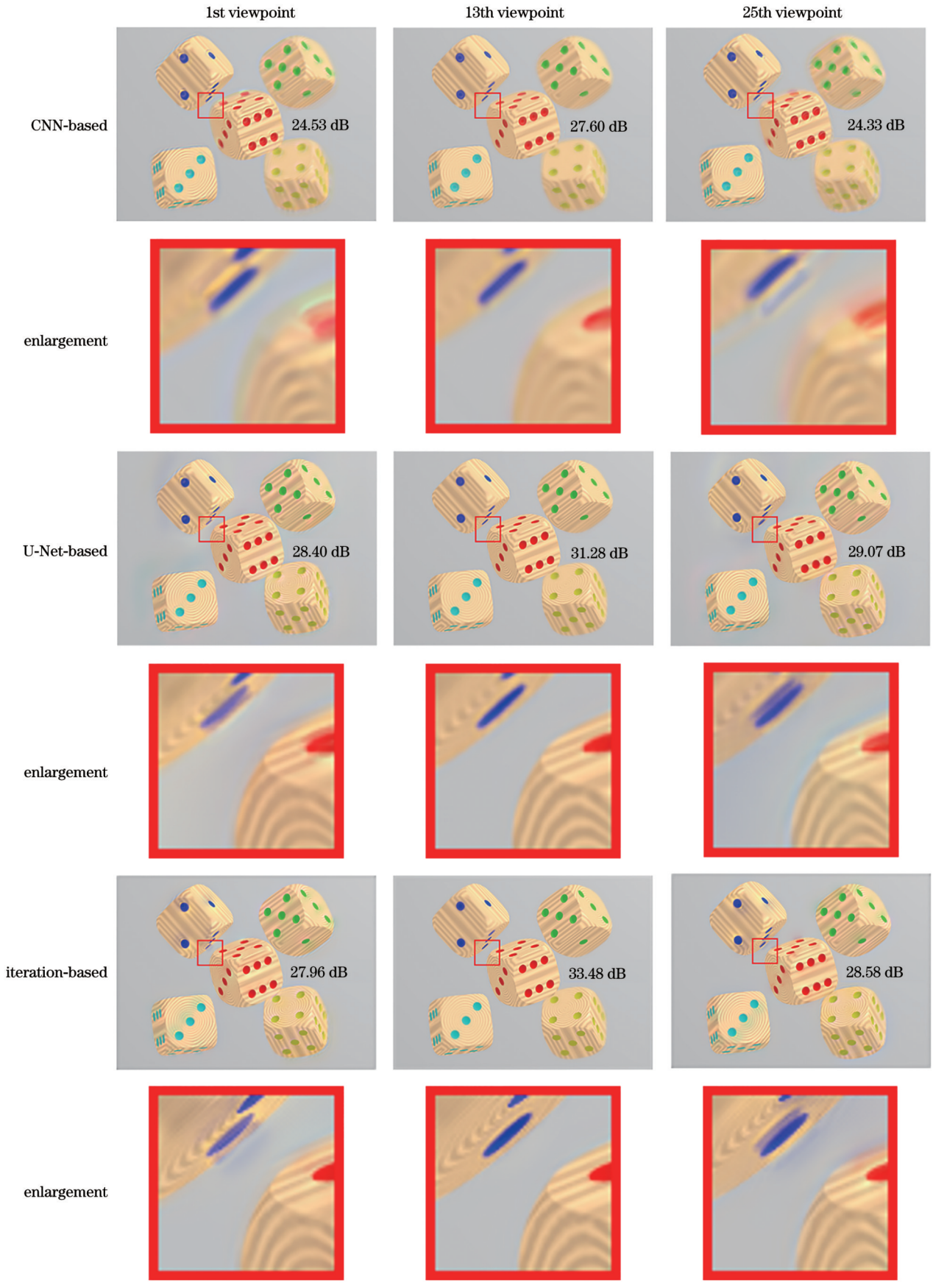

图 18　加法型压缩光场显示的视点图像仿真重建结果

Fig. 18　Simulated reconstruction results of viewpoint images by additive-type compressive light field display

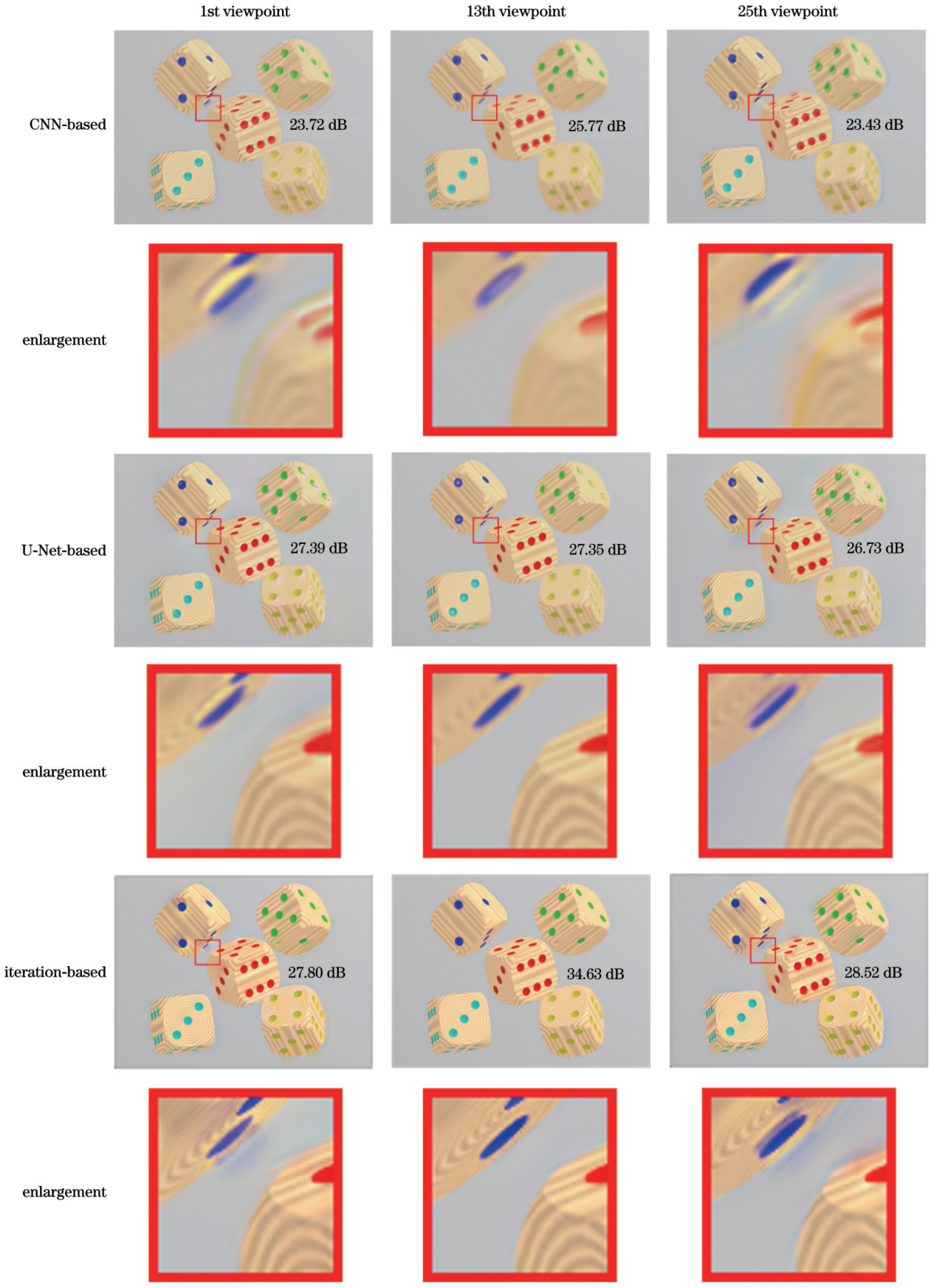

图 19　偏振型压缩光场显示的视点图像仿真重建结果

Fig. 19　Simulated reconstruction results of viewpoint images by polarized-type compressive light field display

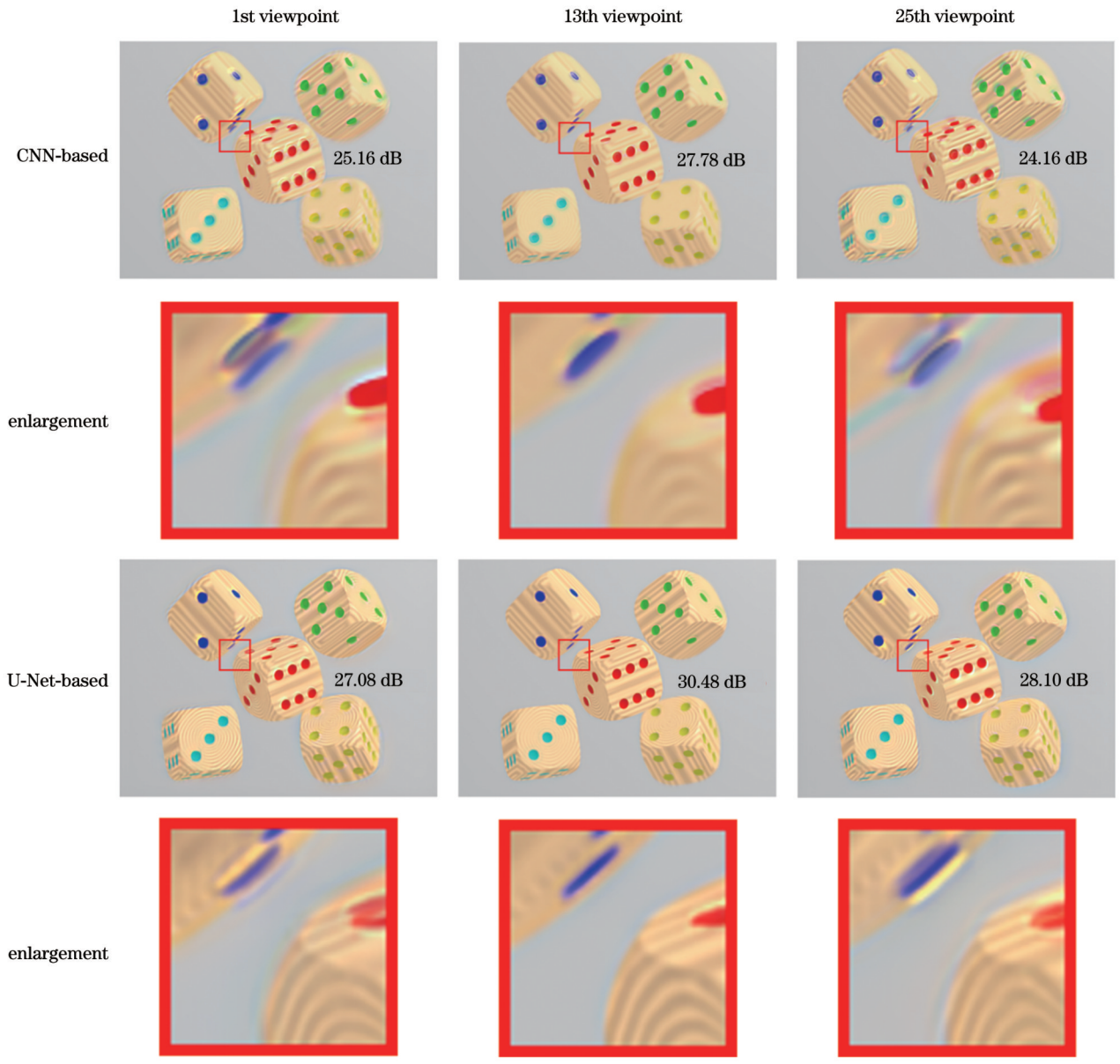

图 20　混合型压缩光场显示的视点图像仿真重建结果

Fig. 20　Simulated reconstruction results of viewpoint images by hybrid-type compressive light field display

偏振型和混合型压缩光场显示图案的仿真重建结果，其中的数值为重建视点图像的PSNR值。

可以看到，相比基于CNN的方法，由U-Net生成的显示图案重建的视点图像的质量都更高，这是因为U-Net生成的显示图案对每层图像的调制量利用率更高。相比基于迭代算法的方法，基于U-Net生成的显示图案重建的边缘视点图像质量与中心视点图像质量的差距更小。也就是说，U-Net的重建结果优于迭代算法的重建结果。但是，对于乘法型和偏振型压缩光场显示，基于U-Net方法的重建光场质量都不如迭代算法，说明所训练的U-Net的泛化性还有待提升。

在本文所使用的计算平台上，CNN和U-Net的训练过程中每个批次的耗时都约为70 ms；CNN的推理时间约为128 ms，U-Net的推理时间约为132 ms。相较于U-Net对CNN重建光场质量的提升，认为这多耗费的几毫秒时间是值得的。U-Net耗费更多推理时间的原因是它具有比较复杂的结构和数量庞大的参数，本文训练得到的U-Net网络占用的存储空间为118 MB，远远大于堆叠CNN所占用的2.6 MB。然而，相较于如今电子设备动辄上百GB的存储容量，U-Net所占用的空间还是可以接受的。但是两个网络的推理时间都没有在人眼暂留时间内，即不能实时生成显示图案，这是因为Python编程语言是逐句解释并执行的。未来，将利用针对深度学习网络的推理优化和运行加速的软件框架（如TensorRT[32]）对网络中的卷积、激活函数等运算进行合并，使得基于C++运行的网络模型能更轻量化地部署在手机、平板电脑等嵌

入式平台并实时推理。乘法型、加法型和偏振型压缩光场显示图案的每次迭代计算时间分别约为 50、15、17 ms，其图像质量与迭代次数的关系如图 21 所示。可以看到：当生成相同重建质量的加法型压缩光场显示图案时，U-Net 的推理时间远远小于迭代算法的计算时间，这是因为基于 U-Net 方法的重建光场质量相当于迭代算法迭代 50 次（750 ms）的结果；U-Net 对乘法型和偏振型压缩光场显示图案生成过程的泛化性还有待提升。因此，在 U-Net 的推理时间内，迭代算法生成图案的重建质量已经超过 U-Net 了。

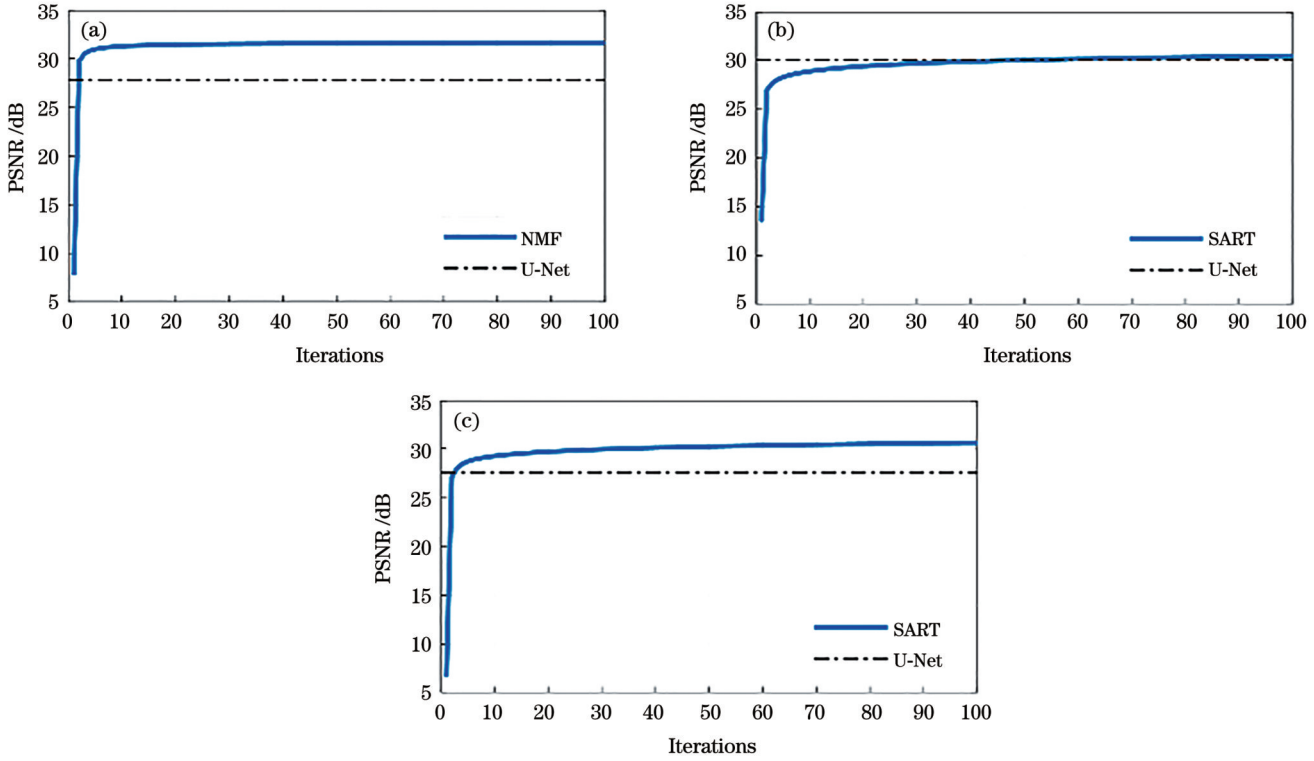

图 21　3 种基本类型压缩光场显示的图像质量和迭代次数的关系。(a)乘法型；(b)加法型；(c)偏振型

Fig. 21　Relationship between image quality and iterations of three basic types of compressive light field displays. (a) Multiplicative-type; (b) additive-type; (c) polarized-type

5　结　　论

提出一种基于 U-Net 的压缩光场显示图案生成方法，并与基于堆叠卷积神经网络的方法进行比较。训练和测试结果表明，所提出的 U-Net 结构对压缩光场显示图案的生成过程具有更好的拟合和泛化能力，并且生成的显示图案和迭代算法所生成的图案更相似。其中，所训练的针对加法型压缩光场显示图案生成的 U-Net 的推理速度明显快于迭代算法，但是所训练的针对乘法型和偏振型光场显示图案生成的 U-Net 的泛化性还有待提高。U-Net 对乘法型和偏振型压缩光场显示图案生成过程的泛化性不如加法型的可能原因是：网络所使用的 ReLU 激活函数性质更符合加法型压缩光场显示中非负光强线性叠加的特点，而乘法型和偏振型压缩光场显示的调制量在线性叠加后又分别经过 $\exp(\cdot)$ 和 $\sin^2(\cdot)$ 函数的非线性运算，其相应改进方法可能有改变网络中的激活函数、增加网络的深度等。

本文使用的训练数据集是人工标注的并且样本数量有限，未来可以考虑通过使用检测物体边缘的算法来自动生成样本数量庞大的光场数据集并公开，从而推动有关光场显示的研究。本文使用的人工神经网络并不能生成多帧压缩光场显示图案，故人工神经网络的参数规模与可生成的显示图案层数和帧数的详细关系还有待研究。此外，由于本文只是通过透视投影变换仿真了离散的视点图像，由人工神经网络生成的显示图案是否真的能实现压缩光场显示还需要搭建显示样机进行实验验证。

参　考　文　献

[1] Hancock P A, Sawyer B D, Stafford S. The effects of display size on performance[J]. Ergonomics, 2015, 58(3): 337-354.

[2] 高晨, 李子寅, 吴仍茂, 等. 便携式三维显示的发展与展望[J]. 激光与光电子学进展, 2023, 60(8): 0811009.
Gao C, Li Z Y, Wu R M, et al. Development and prospect of portable three-dimensional displays[J]. Laser & Optoelectronics Progress, 2023, 60(8): 0811009.

[3] 乔文, 周冯斌, 陈林森. 距离移动电子设备有多远？裸眼 3D 显示现状与展望[J]. 红外与激光工程, 2020, 49(3): 0303002.
Qiao W, Zhou F B, Chen L S. Towards application of mobile devices: the status and future of glasses-free 3D display[J].

Infrared and Laser Engineering, 2020, 49(3): 0303002.

[4] Wetzstein G, Lanman D, Heidrich W, et al. Layered 3D: tomographic image synthesis for attenuation-based light field and high dynamic range displays[J]. ACM Transactions on Graphics, 2011, 30(4): 95.

[5] Heide F, Lanman D, Reddy D, et al. Cascaded displays: spatiotemporal superresolution using offset pixel layers[J]. ACM Transactions on Graphics, 2014, 33(4): 60.

[6] Kreinovich V Y. Arbitrary nonlinearity is sufficient to represent all functions by neural networks: a theorem[J]. Neural Networks, 1991, 4(3): 381-383.

[7] Xiao L, Kaplanyan A, Fix A, et al. DeepFocus: learned image synthesis for computational displays[J]. ACM Transactions on Graphics, 2018, 37(6): 200.

[8] Choi S, Gopakumar M, Peng Y F, et al. Time-multiplexed neural holography: a flexible framework for holographic near-eye displays with fast heavily-quantized spatial light modulators [C]//SIGGRAPH '22: ACM SIGGRAPH 2022 Conference Proceedings, August 7-11, 2022, Vancouver, BC, Canada. New York: ACM Press, 2022: 1-9.

[9] 常琛亮, 戴博, 夏军, 等. 面向视觉舒适度的全息近眼显示研究综述[J]. 激光与光电子学进展, 2022, 59(20): 2011001.
Chang C L, Dai B, Xia J, et al. Review of holographic near-eye displays for visual comfort[J]. Laser & Optoelectronics Progress, 2022, 59(20): 2011001.

[10] 刘娟, 皮大普, 王涌天. 实时全息三维显示技术研究进展[J]. 光学学报, 2023, 43(15): 1509001.
Liu J, Pi D P, Wang Y T. Research progress of real-time holographic 3D display technology[J]. Acta Optica Sinica, 2023, 43(15): 1509001.

[11] Liu K X, Wu J C, He Z H, et al. 4K-DMDNet: diffraction model-driven network for 4K computer-generated holography[J]. Opto-Electronic Advances, 2023, 6(5): 220135.

[12] Zheng H D, Peng J C, Wang Z, et al. Diffraction model-driven neural network trained using hybrid domain loss for real-time and high-quality computer-generated holography[J]. Optics Express, 2023, 31(12): 19931-19944.

[13] Ren H, Wang Q H, Xing Y, et al. Super-multiview integral imaging scheme based on sparse camera array and CNN super-resolution[J]. Applied Optics, 2019, 58(5): A190-A196.

[14] Yu X B, Li H Y, Sang X Z, et al. Aberration correction based on a pre-correction convolutional neural network for light-field displays[J]. Optics Express, 2021, 29(7): 11009-11020.

[15] Maruyama K, Takahashi K, Fujii T. Comparison of layer operations and optimization methods for light field display[J]. IEEE Access, 2020, 8: 38767-38775.

[16] 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 (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.

[17] Zheng Q, Babaei V, Wetzstein G, et al. Neural light field 3D printing[J]. ACM Transactions on Graphics, 2020, 39(6): 207.

[18] Mildenhall B, Srinivasan P P, Tancik M, et al. NeRF: representing scenes as neural radiance fields for view synthesis [J]. Communications of the ACM, 2022, 65(1): 99-106.

[19] Sun Y F, Li Z, Wang S Z, et al. Depth-assisted calibration on learning-based factorization for a compressive light field display [J]. Optics Express, 2023, 31(4): 5399-5413.

[20] 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.

[21] Wetzstein G, Lanman D, Hirsch M, et al. Tensor displays: compressive light field synthesis using multilayer displays with directional backlighting[J]. ACM Transactions on Graphics, 2012, 31(4): 80.

[22] Lee S, Jang C, Moon S, et al. Additive light field displays: realization of augmented reality with holographic optical elements [J]. ACM Transactions on Graphics, 2016, 35(4): 60.

[23] Lanman D, Wetzstein G, Hirsch M, et al. Polarization fields: dynamic light field display using multi-layer LCDs[J]. ACM Transactions on Graphics, 2011, 30(6): 186.

[24] Zhang J H, Fan Z C, Sun D W, et al. Unified mathematical model for multilayer-multiframe compressive light field displays using LCDs[J]. IEEE Transactions on Visualization and Computer Graphics, 2019, 25(3): 1603-1614.

[25] Lanman D, Hirsch M, Kim Y, et al. Content-adaptive parallax barriers: optimizing dual-layer 3D displays using low-rank light field factorization[J]. ACM Transactions on Graphics, 2010, 29 (6): 163.

[26] Wetzstein G, Lanman D, Hirsch M, et al. Real-time image generation for compressive light field displays[J]. Journal of Physics: Conference Series, 2013, 415(1): 012045.

[27] Kim D, Lee S, Moon S, et al. Hybrid multi-layer displays providing accommodation cues[J]. Optics Express, 2018, 26 (13): 17170-17184.

[28] Zhu L M, Du G, Lü G Q, et al. Performance improvement for compressive light field display with multi-plane projection[J]. Optics and Lasers in Engineering, 2021, 142: 106609.

[29] Liu M L, Lu C H, Li H F, et al. Bifocal computational near eye light field displays and structure parameters determination scheme for bifocal computational display[J]. Optics Express, 2018, 26(4): 4060-4074.

[30] Jo N Y, Lim H G, Lee S K, et al. Depth enhancement of multi-layer light field display using polarization dependent internal reflection[J]. Optics Express, 2013, 21(24): 29628-29636.

[31] He K M, Zhang X Y, Ren S Q, et al. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1026-1034.

[32] Vanholder H. Efficient inference with tensorRT[C]//2016 GPU Technology Conference, April 4-7, 2016, San Jose, USA. Beaverton: Khronos, 2016: 1-24.

Image Synthesis of Compressive Light Field Displays with U-Net

Gao Chen[1,2], Tan Xiaodi[3,4,5*], Li Haifeng[6], Liu Xu[6]

[1]*College of Photonic and Electronic Engineering, Fujian Normal University, Fuzhou* 350117, *Fujian, China*;
[2]*Fujian Provincial Key Laboratory of Photonics Technology, Fuzhou* 350117, *Fujian, China*;
[3]*Key Laboratory of Optoelectronic Science and Technology for Medicine of Ministry of Education, Fuzhou* 350117, *Fujian, China*;
[4]*Fujian Provincial Engineering Technology Research Center of Photoelectric Sensing Application, Fuzhou* 350117, *Fujian, China*;
[5]*Information Photonics Research Center, Fujian Normal University, Fuzhou* 350117, *Fujian, China*;
[6]*College of Optical Science and Engineering, Zhejiang University, Hangzhou* 310027, *Zhejiang, China*

Abstract

Objective 3D display technology is the entrance to the realistic-feeling metaverse for tabletop, portable, and near-eye electronic devices. True 3D displays are mainly divided into light field displays and holographic displays, among which light field displays can be further subdivided into integral-imaging displays, directional light field displays, and compressive light field displays. Compressive light field displays utilize the scattering characteristic of display panels and the correlation between viewpoint images of the 3D scene. The compressive light field display is a candidate for portable 3D display owing to its compact structure, moderate viewing angle, and high spatial resolution. However, computational resources of portable electronic devices are restricted to satisfy their duration demand. Meanwhile, iterative algorithms to solve the compressive light field display patterns have the problem of heavy computation, preventing compressive light field displays from being a practical solution to portable dynamic 3D displays. With the development of artificial intelligence technology, image generation algorithms based on deep learning are gradually applied to 3D displays. Deep neural networks can be trained to fit the iterative process. Additionally, fast display image synthesis could be realized with rapid forward propagation of artificial neural networks. Previously, researchers proposed a stacked CNN-based method to generate images for compressive light field displays. However, the stacked CNN-based method suffers from convergence and over-fitting problems. U-Net is initially employed for image segmentation in computed tomography to handle slicing data and output the organ's cancer probability. The skip connection added in the U-Net architecture significantly improves its convergence compared with the stacked CNN model. Light field data are pretty similar to slicing data in computed tomography. Thus, we introduce U-Net as the network model for optimizing compressive light field display patterns for better convergence and generalization. Given a specific viewing angle, several augmented target light field datasets are generated as the training sets of U-Net. After the U-Net converges, the trained U-Net synthesizes the display patterns that reconstruct the target light field for testing. The training and testing results prove that compared to the stacked CNN-based method and iterative algorithms, the proposed U-Net-based pattern generation method for compressive light field displays features higher reconstruction quality and fewer computing resources.

Methods An artificial neural network's training procedure can be split into forward and backward propagation. The forward propagation includes the following steps. Firstly, the target light field for training is input into the network, display images are output, and then the light field is reconstructed by simulated perspective projection. The backward propagation is to update the network's parameters with the loss function and regular terms. Meanwhile, the above procedure is repeated during every epoch and batch. When the training is finished, the target light field for testing is input into the network, and display images are synthesized. This is called the inference procedure. The datasets, network architecture, and hyper-parameters are carefully designed to fit the features of compressive light field displays. The datasets contain 1260 pairs of image blocks cropped from seven scenes. The ReLU function is set as the activation function of the U-Net model initialized uniformly with Kaiming Initialization. The loss function is the mean square error between the target and reconstructed light field and the regular term is the effective range of image pixel values.

Results and Discussions Performances of the proposed U-Net-based method, the stacked CNN-based method, and iterative algorithms are compared fairly for multiplicative (Fig. 8), additive (Fig. 9), polarized (Fig. 10), and hybrid (Fig. 11) types of compressive light field displays. The training and testing results (Figs. 17–20) prove that the proposed method's light field reconstruction quality is always 2 dB higher than that of stacked CNN-based method. The reason is that the U-Net-based method utilizes the value range of image pixels more effectively than the stacked CNN-based method.

Additionally, for additive-type compressive light field displays, the proposed method takes less time to reach the same reconstruction quality than iterative algorithms (Fig. 21).

Conclusions To improve the image quality, uniformity, and computation performance of compressive light field displays, we apply an elaborate U-Net model to synthesize display images. The proposed method is compared with the stacked CNN method and iterative algorithms by simulating the perspective projection of display images with the same target light field as input. For the additive-type compressive light field display, the trained U-Net's inference speed is much faster than the speed of iterative algorithm under the same reconstruction quality. However, the trained U-Net's generalization performance still needs promotion for multiplicative and polarized-type compressive light field displays. Possible improvements are changing activation functions and increasing the network's depth.

Key words physical optics; imaging system; compressive light field display; light field rendering; deep learning