

特色专题

视觉图像驱动的多模型协同超表面逆设计方法

司黎明^{1,2,3,4,5}, 牛荣^{1,2,3}, 党晨阳^{1,2,3}, 王钊睿^{1,2,3}, 马天宇^{1,2,3}, 李琰⁵, 朱卫仁⁶, 孙厚军^{1,2,3}

摘要 随着超表面在电磁波精确调控领域的快速发展,复杂功能器件的高效、自动化设计已成为当前研究的核心挑战。提出了一种视觉图像驱动的多模型协同超表面逆设计方法,以视觉图像作为输入,实现电磁超表面阵列结构的自动生成。该方法采用多模型协同的策略,构建了从图像特征到物理结构参数的端到端映射框架。首先引入条件生成对抗网络(image-to-image translation with conditional adversarial networks, Pix2Pix)将输入视觉图像转换为目标全息纹理;随后利用U型卷积神经网络(convolutional networks for biomedical image segmentation, U-Net)实现高保真相位分布的精准预测。在此基础上,进一步结合变分自编码器(variational autoencoder, VAE),将目标相位需求映射为可制造的超表面单元几何结构,有效解决了传统设计流程中视觉图像输入求解困难且与结构建模相互割裂的问题。数值仿真结果表明,所设计的全息超表面在指定工作频点下能够准确重构目标全息图案,其重构保真度和设计稳定性均显著优于传统单一网络方法。为图像驱动的智能超表面逆设计提供了一种高效、自动化的新范式,对复杂电磁功能器件的快速设计具有重要参考价值。

关键词 人工智能;多模型协同框架;超表面;逆设计;视觉图像;神经网络

电磁超材料作为一种由人工结构单元按特定空间序构排列而成的功能性人工材料,其物理特性主要取决于单元的几何形态、周期排列方式及构成材料参数^[1-4]。通过对这些结构特征的精确设计,超材料能够展现出天然材料所不具备的特殊物理性质,从而突破自然界材料的物理响应限制。这种人工调控机制赋

予了研究者对电磁波振幅、相位、偏振态及传播路径进行跨尺度、全方位灵活调控的能力^[5-9]。目前,电磁超材料已在宽带通信系统、雷达隐身技术、可重构波束成形,以及高分辨率成像探测等关键领域展现出巨大的应用潜力。这些研究不仅推动了诸如高性能超表面透镜、智能反射面等新型器件的研制,也为新一代全息成像系统的集成化发展奠定了基础。然而,随着现代电磁系统对高性能、轻量化及动态可编程特性的需求日益迫切,如何实现满足特定复杂功能需求的超表面阵列精准设计已成为制约该领域发展的核心挑战^[10-13]。在传统设计流程中,研究者通常需要构建复杂的物理模型,并依赖大规模的全波电磁仿真与反复的经验参数扫参。这种“试错式”的方法不仅计算

1. 北京理工大学集成电路与电子学院, 北京 100081

2. 临近空间环境特性及效应全国重点实验室, 北京 100081

3. 北京理工大学唐山研究院, 唐山 063000

4. 毫米波全国重点实验室, 南京 210096

5. 深圳北理莫斯科大学工程系, 深圳 518172

6. 上海交通大学集成电路学院, 上海 200240

收稿日期: 2026-01-21; 修回日期: 2026-03-30

基金项目: 国家自然科学基金项目(62271056, 62571049, 62171186, 62201037); 国家市场监督管理总局专项(AKYKF2423); 北京市自然科学基金-海淀原始创新联合基金项目(L222042); 河北省自然科学基金项目(F2025105029); 东南大学毫米波国家重点实验室开放课题(KN20250214); 中国教育部高等学校学科科研创新引智计划项目(B14010)

作者简介: 司黎明, 教授, 研究方向为电磁场与微波技术, 电子信箱: lms@bit.edu.cn

引用格式: 司黎明, 牛荣, 党晨阳, 等. 视觉图像驱动的多模型协同超表面逆设计方法[J]. 科技导报, 2026, 44(9): 88-97; doi:10.3981/j.issn.1000-7857.2026.01.00116

资源消耗大、设计周期冗长,且在面对多目标耦合与高维参数搜索时往往难以寻优,已难以满足现代电磁工程对实时响应与海量结构快速反演的需求^[14-16]。

近年来,人工智能(artificial intelligence, AI)技术,特别是深度学习与生成式模型在图像理解、语义编码、结构生成与复杂非线性映射学习方面取得显著进展,为超材料逆设计提供了新路径^[17-18]。已有研究表明,神经网络能够在无须显式物理求解的情况下学习结构-响应之间的深层映射,实现模式识别、参数反演、相位预测及全息图设计等任务,并显著减少对高耗时仿真的依赖^[19-20]。同时,生成对抗网络(generative adversarial networks, GAN)^[21]、变分自编码器(variational autoencoder, VAE)^[22]、扩散模型^[23]等生成框架,可直接输出满足目标电磁功能的结构图案,有效提升结构创新性与设计自由度。由此,AI正迅速从单纯的“辅助优化工具”演进为涵盖功能定义、图案构建与结构生成的全流程设计范式。这种转型不仅推动了智能可编程超表面的发展,也为基于视觉图像或语义描述驱动的电磁器件自动化设计奠定了技术基础^[24-27]。

深度学习与物理原理的深度融合显著提升了反演问题求解的保真度,而跨域生成技术的进步则为异构模态间的数据转换提供了稳健的支撑^[28-30]。然而,直接的端到端映射在处理电磁散射任务中固有的强非线性与域偏移问题时,往往面临模型收敛困难及泛化性不足的瓶颈。值得注意的是,深度学习在医学CT(computed tomography)图像重建领域已取得巨大成功,通过引入多阶段解耦策略将复杂的伪影去除与解剖结构恢复任务进行拆分,显著提升了医学影像的重建质量与鲁棒性^[31-33]。这种多阶段处理的方法为处理高维电磁场数据提供了重要启示,即通过解耦中间物理变量可以显著降低复杂映射的学习难度。但是,将这种分阶段处理的策略应用于视觉驱动的超表面逆设计的研究还处于起步阶段,如何克服从直观视觉图像到微观电磁结构之间巨大的信息鸿沟,并解决因缺乏有效中间过渡步骤而导致的相位预测失真及结构不可制造的问题还有待深入研究。

本研究提出一种由视觉图像驱动的多模型协同超表面逆设计框架。该框架创新性地引入中间态全息图像作为物理约束的桥梁,利用条件生成对抗网络

(image-to-image translation with conditional adversarial networks, Pix2Pix)实现从视觉特征到干涉图像的跨域映射,从而建立图像语义与电磁物理场之间的初步关联。在此基础上,结合U型卷积神经网络(convolutional networks for biomedical image segmentation, U-Net)的相位预测能力与VAE的结构生成能力,构建了从全息干涉图像到空间相位分布,再到可制造几何结构的完整映射链路。该方法有机融合了生成模型的跨域图像映射能力、多尺度网络的高精度相位重建能力,以及潜空间模型的结构表征与采样优势,实现了超表面阵面的全自动化逆设计。通过分阶段的协同建模策略,本研究有效缓解了传统设计流程中图像处理、相位求解与单元建模相互割裂的问题,最终形成了一种可直接面向工程实现的端到端智能逆设计范式,为新一代视觉驱动的智能电磁器件设计提供了可行路径。

1 多模型协同 AI 模型逆设计方法

图1全面展示了多模型协同AI赋能全息超表面逆设计方法的总体架构。该框架以原始视觉图像为起始输入,通过多模型间的深度协同,实现了从图像抽象特征到超表面具体物理结构的端到端映射。在流程起始阶段,输入图像首先进入条件生成对抗网络进行预处理,旨在完成从“视觉空间”向“电磁干涉空间”的跨域转换。该网络通过对海量图像-全息纹理

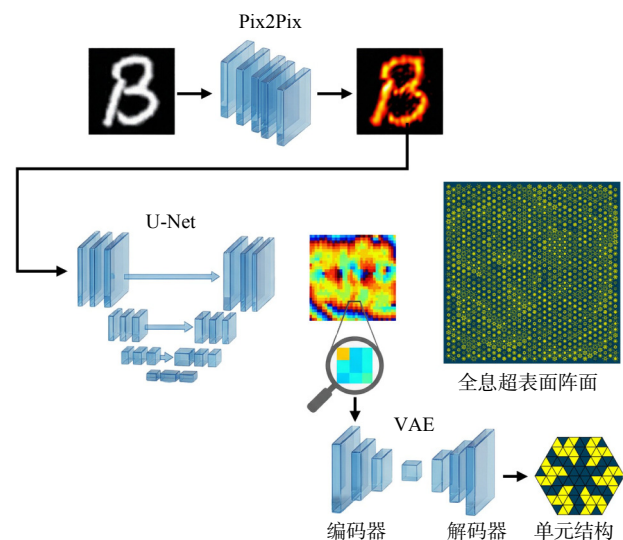


图1 多模型协同 AI 网络全息超表面逆设计示意

配对样本的对抗性训练,能够精准捕捉原始目标的边缘细节、纹理走向及全局拓扑关系。Pix2Pix的生成器负责将提取到的语义特征重构为具备干涉特性的全息纹理模式,而判别器则通过统计分布约束,确保生成结果在视觉上与物理规律保持高度一致。这种基于生成对抗的思想不仅提升了全息图的图像逼真度,而且为后续的相位恢复任务奠定了具备物理合理性的数据基础。

获取全息纹理后,系统将其输入 U-Net 相位预测网络,利用其对称的编码器-解码器结构执行复杂的相位回归任务。该网络通过多尺度特征提取机制捕捉全息图中的电磁传播特征,并利用跳跃连接将高层语义信息与底层几何细节进行深度融合,从而输出高分辨率的二维相位矩阵。在本实验配置中,设计目标设定为在 10 GHz 线极化电磁波入射下,于距离阵面 200 mm 处的目标平面内精确重构全息图像。随后,该二维相位矩阵作为物理约束条件输入 VAE 生成模型,开启从数字相位到几何结构的逆向生成过程。在单元设计中,仅对图案区域的 1/6 进行编码。编码区域被离散化为等边三角形晶格,其中“1”和“0”分别代表铜层的存在与否。通过改变编码图案,总共可生成 2^{16} 种单元,覆盖不同的相位性能。顶层的每个晶格边长设定为 1 mm,并且编码区域与单元边界之间保持 0.1 mm 的间距,以降低相邻单元间的互耦效应。此外,介质中间层采用 Rogers 4003C 基板,其厚度为 1.524 mm,相对介电常数为 3.55,损耗角正切为 0.0027。VAE 通过在潜空间中学习结构参数与电磁响应的映射分布,能够自动化地输出满足特定相位需求的离散六边形单元参数,实现了“相位到结构”的无缝衔接。最终,生成的单元根据相位排布按照 30 行

34 列交错排列的拓扑结构进行空间拼接,构建出完整的超表面阵面。这一设计贯通了从视觉输入到物理输出的完整设计链路,在实现高保真成像的同时,显著提升了全息超表面设计的自动化水平与工程实用性。

1.1 条件生成对抗网络实现跨域生成

二维平面图像的空间强度分布与电磁波前相位分布之间存在内在的数学物理关联:图像灰度梯度与所需相位梯度的变化方向具有对应关系,清晰的图像边界往往需要相位的剧烈变化来实现,这源于两者共同遵循的波动传播规律。然而,从视觉图像直接映射到电磁相位面临非线性复杂度高、逆问题非适定等挑战。为此,本研究创新性地引入中间态全息纹理作为物理桥梁,全息纹理既保留了原始视觉特征的拓扑信息,又蕴含衍射过程的干涉物理特征。由于全息纹理的强度分布梯度已反映潜在的相位变化趋势,相位预测由此转化为约束优化问题而非无约束反演,显著降低了网络的学习难度。图 2 展示了以输入图像为例,经 Pix2Pix 生成全息图像的流程。在该阶段, Pix2Pix 用于将视觉图像映射为带有电磁干涉纹理的全息图。Pix2Pix 由生成器 G 与判别器 D 构成:生成器负责生成全息图像,判别器判别生成结果的真实性,并通过对抗训练优化生成器,使输出分布尽可能接近真实全息结果。

其目标函数为 Pix2Pix 损失与 L1 重建损失的联合优化。对抗损失如下

$$\mathcal{L}_{\text{cGAN}}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_x[\log(1 - D(x, G(x)))] \quad (1)$$

式中, $\mathcal{L}_{\text{cGAN}}(G, D)$ 为条件生成对抗网络损失函数, x 为输入视觉图像, y 为对应的目标全息纹理图, $\mathbb{E}_{x,y}$ 是 (x, y) 联合分布的数学期望算子。

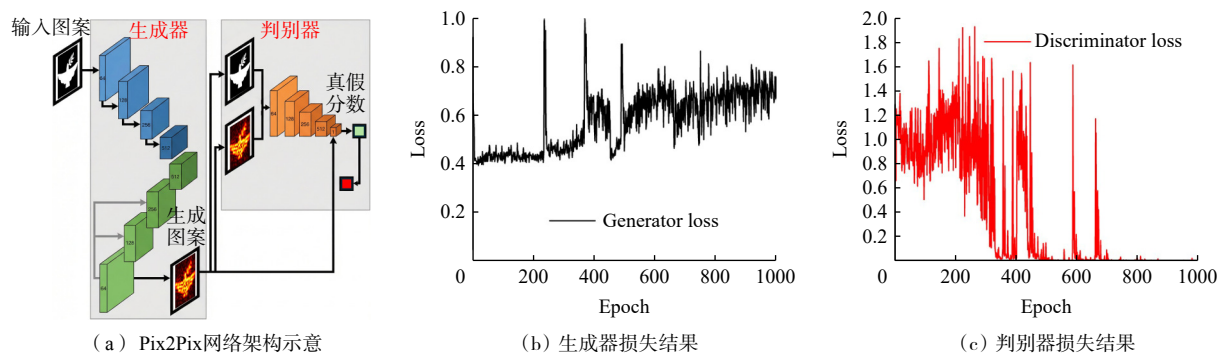


图2 Pix2Pix 网络模型结构与结果

而 L1 损失为

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y}[\|y - G(x)\|_1] \quad (2)$$

式中, $\mathcal{L}_{L1}(G)$ 代表 L1 范数重建损失函数, $\|y - G(x)\|_1$ 代表生成结果与目标之间的 L1 范数。

综合损失为

$$G^* = \arg \min_G \max_D (\mathcal{L}_{GAN}(G, D) + \lambda \mathcal{L}_{L1}(G)) \quad (3)$$

式中, G^* 代表最优生成器, $\arg \min_G \max_D$ 为对 G 求最小值、对 D 求最大值的参数配置, λ 控制对 L1 损失的权重, 确保生成图在像素级上与目标图接近, 同时保持全局分布的一致性。

在网络结构上, 生成器采用 U-Net 架构, 共包含 5 层下采样编码器与 5 层上采样解码器。编码器通过卷积将特征通道数由 64 逐步增加至 512; 解码器采用对称结构逐步恢复空间分辨率, 以保留低层细节信息。判别器引入谱归一化以稳定训练。其输入为图像与条件对拼接后的特征图, 网络包含 4 层卷积结构, 通道数依次为 64、128 和 256, 最终输出单通道分数结果, 对局部图像块进行真伪判断, 从而增强生成全息纹理的高频细节逼真度。

为了训练 Pix2Pix 网络将视觉图像映射为全息衍射图, 需要构建高质量的图像-全息图配对数据集。训练集的生成基于 Gerchberg-Saxton 相位迭代算法 (Gerchberg-Saxton algorithm, GS), 该算法通过在输入图像域和输出衍射域之间迭代计算, 实现目标图像强度的精确重建。在本研究中, 将重建距离设定为 200 mm, 并要求生成的电磁全息图与相位矩阵均达到 100×100 的空间分辨率。设输入目标图像振幅为 $A(x, y)$, 衍射面复振幅为 $U(u, v)$, GS 算法在每次迭代中的更新过程可表示为

$$U_n(u, v) = \mathcal{F}[A(x, y)e^{i\phi_n(x, y)}] \quad (4)$$

$$\phi_{n+1}(x, y) = \arg\{\mathcal{F}^{-1}[B(u, v)e^{i\arg(U_n(u, v))}]\} \quad (5)$$

式中, $U_n(u, v)$ 为第 n 次迭代的衍射面复振幅 $\mathcal{F}$ 和 $\mathcal{F}^{-1}$ 分别表示傅里叶变换与逆傅里叶变换, $A(x, y)$ 为输入目标图像的振幅分布, $\phi_n(x, y)$ 为第 n 次迭代的输入相位分布, $B(u, v)$ 为衍射截面期望振幅, $\arg(\cdot)$ 表示复数的相位角提取操作。

在训练集构建过程中, 每张目标图像初始相位设为随机分布。每次迭代后, 系统通过计算 200 mm 处衍射重构图像与目标图像的相似度指标来判断重建

效果。如果相似度在连续 10 次迭代中未能进一步提升, 即认为迭代达到收敛状态, 停止计算并保存最终生成的 100×100 像素电磁全息图及其对应的相位分布结果。这种策略确保了训练集中的每一对样本均经过充分优化, 为后续 Pix2Pix 网络的监督训练提供了具备物理真实性的输入。最终, 通过上述方法生成约 10000 对覆盖字符、图案及纹理的视觉图像与全息结果数据集, 保证了网络能够学习广泛的从视觉图像到电磁全息的映射关系。

在模型训练阶段, 本研究利用预先构建的 1 万对视觉图像与电磁全息配对样本, 对 Pix2Pix 网络进行了 1000 个 Epoch 的深度迭代优化。为了确保复杂异构映射下的参数更新稳定性, 优化器选用 Adam 算法, 并将学习率与批大小分别设定为 2×10^{-4} 与 8。训练过程中生成器与判别器的损失收敛曲线如图 2(b)~(c) 所示, 直观记录了模型从特征摸索到映射建立的学习轨迹。图 2(b) 显示随着网络对全息纹理细节捕捉能力的增强, 损失值在经历动态博弈后逐渐趋于平稳, 最终稳定在 0.626 左右。与此同时, 图 2(c) 中的判别器损失在前期剧烈波动后迅速下降并收敛至 4.5×10^{-4} , 表明判别器已难以区分生成的全息图与真实物理样本。这种生成对抗损失的协同收敛, 有力证明了模型在训练中达到了理想的纳什平衡状态, 确保了输出结果能够精准遵循电磁干涉的物理分布。通过这种深度学习策略, 生成器不仅能够还原图像的全局语义特征, 更显著提升了对全息图局部细节的重构精度。这一训练成果充分验证了本研究提出的多模型协同 AI 网络在处理电磁逆设计任务时, 具备极高的收敛效率与强大的特征表征能力。

1.2 U-Net 网络实现全息相位映射

在 Pix2Pix 网络生成高保真全息图像后, 本研究构建基于 U-Net 架构的回归网络, 用于实现从全息强度分布到空间相位信息的预测。如图 3(a) 所示, 该网络采用对称的编码器-解码器 (Encoder-Decoder) 结构。编码器由 4 个下采样模块构成, 每层均集成双层 3×3 卷积、批量归一化 (batch normalization, BN) 及修正线性激活函数 (rectified linear unit, ReLU); 特征通道数自 32 起始, 经由最大池化层逐级扩展至 512 维瓶颈层, 旨在深度挖掘图像的多尺度语义特征。解码器端通过转置卷积逐步恢复空间分辨率, 并引入跳跃

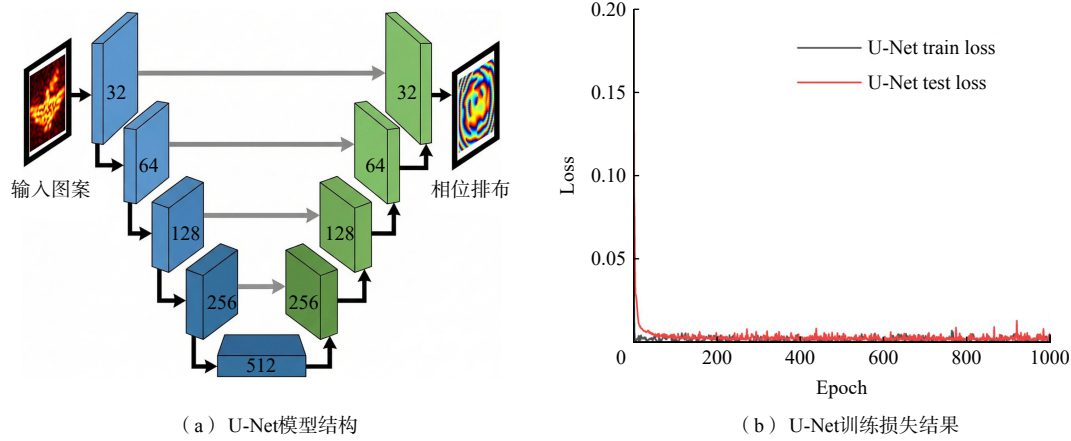


图3 U-Net网络模型结构

连接机制以实现多尺度特征融合。与传统通道拼接不同,本模型在跳跃连接中采用了中心裁剪与逐元素相加策略,将浅层空间细节精准注入深层语义特征中。网络末端经由 1×1 卷积层输出单通道特征,最后通过双线性插值将特征图映射至 100×100 的目标物理维度。

为确保相位重建的准确性,模型训练由双重约束引导,首先是基于欧氏距离的相位重建误差

$$\mathcal{L}_{\text{phase}} = \mathbb{E}_I[\|\Phi - \hat{\Phi}\|_2^2] \quad (6)$$

式中, $\mathcal{L}_{\text{phase}}$ 为相位重建误差损失函数,衡量网络输出与真实相位的欧氏距离, $\mathbb{E}_I$ 为对所有输入图像 I 的期望算子, Φ 为真实相位矩阵, $\hat{\Phi}$ 为网络预测相位矩阵, $\|\Phi - \hat{\Phi}\|_2$ 为两者差值的L2范数,范数平方项用于强化大误差的惩罚;另一方面是相位梯度平滑正则项,约束相邻单元的相位连续性,表达为并加入平滑正则项

$$\mathcal{L}_{\text{smooth}} = \sum_{ij} (\|\nabla_x \hat{\Phi}_{ij}\|_1 + \|\nabla_y \hat{\Phi}_{ij}\|_1) \quad (7)$$

式中, $\mathcal{L}_{\text{smooth}}$ 为相位梯度平滑正则损失, $\nabla_x \hat{\Phi}_{ij}$ 为预测相位矩阵在 x 方向的梯度,该损失是为了约束相邻单元的相位连续性,避免出现物理上不合理的相位断裂。最终联合损失函数定义为

$$\mathcal{L}_{\text{U-net}} = \mathcal{L}_{\text{phase}} + \alpha \mathcal{L}_{\text{smooth}} \quad (8)$$

式中, $\mathcal{L}_{\text{U-net}}$ 为U-Net模型总损失函数, α 为平滑项权重系数,用于平衡2个损失项的贡献程度。训练数据采用前阶段Pix2Pix输出的全息图像与对应的真实相位矩阵,数据集包含10000幅图像样本。训练迭代1000次epoch,优化器使用Adam,学习率为 2×10^{-4} ,

批大小为8。如图3(b)所示,随着迭代的深入,网络展现出卓越的收敛特性与泛化能力:训练集最终损失收敛至0.00138,测试集损失稳定在0.00311。结果表明,该U-Net回归网络能够建立全息纹理与相位分布之间的非线性映射,并在保持相位空间连续性的同时,为后续VAE单元结构生成提供可靠输入。

1.3 VAE网络实现单元逆设计

在U-Net输出的相位矩阵基础上,本研究引入VAE实现单元结构的逆向生成,将相位需求精准映射为可制造的几何参数(图4)。

如图4(a)所示,VAE通过编码器将输入的二维相位块映射至潜在空间,预测其概率分布的均值 μ 和标准差 σ ,编码器模型定义为

$$q_{\phi}(z|x) = \mathcal{N}(z; \mu, \sigma^2) \quad (9)$$

式中, $q_{\phi}(z|x)$ 为编码器的条件概率分布, x 为相位块输入, z 为潜变量, ϕ 为编码器参数, $\mathcal{N}(z; \mu, \sigma^2)$ 为多元高斯分布,参数为均值 μ 和方差 σ^2 。解码器接收潜变量 z 并生成对应的单元几何结构 x_{unit} ,实现条件重建

$$x_{\text{unit}} = G_{\text{VAE}}(z) \quad (10)$$

式中, x_{unit} 为生成的单元几何结构参数向量, G_{VAE} 为VAE解码器函数。

为提升潜空间的解耦能力,训练目标设定为最大化变分下界。在实际优化中,引入权重系数 $\beta = 0.5$ 来平衡重构精度与潜在分布的正则化程度,总损失函数定义为

$$L_{\text{total}} = L_{\text{recon}} + \beta \cdot D_{\text{KL}}(q_{\phi}(z|x) \| p(z)) \quad (11)$$

式中, L_{total} 为VAE总损失函数,分为2项,第1项 L_{recon} 为二元交叉熵(binary cross entropy, BCE)重构损

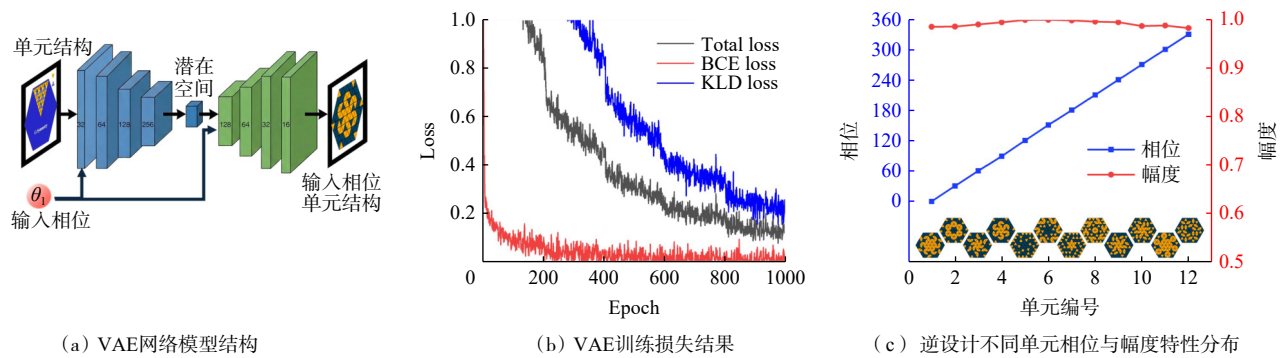


图4 VAE网络模型理论与结构

失,确保生成的单元参数与真实 CST 仿真结构高度一致;第2项 $D_{KL}(q_\phi(z|x) \parallel p(z))$ 为 KL 散度 (Kullback-Leibler divergence, KLD), 约束潜变量服从标准正态分布以保证空间的连续性与可采样性。 $q_\phi(z|x)$ 为后验分布, 真实数据驱动的潜空间分布, $p(z)$ 为先验分布, 标准正态分布。 β 为 KL 散度权重系数。

在网络配置上, 编码器由4层卷积层构成, 特征通道数依次为32、64、128和256; 解码器采用对称的转置卷积结构, 通道数经128、64、32逐步缩减, 最终通过16维全连接层输出具体的几何参数向量。训练数据集涵盖了由CST仿真得到的10000组单元结构及其对应的相位响应。实验设定迭代轮次为1000 epoch, 采用Adam优化器, 学习率为 2×10^{-4} 。如图4(b)所示, 模型展现了良好的收敛特征: 总损失最终稳定在0.13857, 其中BCE重构损失收敛至0.02681。结果表明, 该VAE模型能够高效学习结构与相位间的深层非线性映射, 实现基于目标相位需求的单元结构自动化、高精度逆设计。

为直观展示VAE模型对超表面单元结构的生成能力, 图4(c)给出了模型逆设计的12个示例单元的相位与幅度分布。结果表明, 生成的结构(如图中底部拓扑示意图所示)在 $0^\circ \sim 360^\circ$ 范围内表现出极其优异的线性相位梯度(步长 30°), 且幅度响应始终保持在0.95以上的高水平。这证明了VAE模型不仅能够潜空间中精准捕获复杂的物理约束, 实现相位到几何结构的端到端逆向映射, 还能确保生成单元在全相位覆盖范围内具有高度的幅相一致性。在实际布阵过程中, 由于GS算法输出的 100×100 原始相位矩阵与 30×34 的物理阵列存在分辨率差异, 系统首先根据物理单元的排布位置对对应的局部相位区域进行空间平均化处理。随后, 将平滑后的目标相位作为条件

输入VAE模型, 通过其解码器检索潜空间分布, 从而为阵面上的每一个格点生成对应的几何参数。最终, 生成的六边形单元按照预设的交错拓扑结构进行拼接, 构建出完整的全息超表面阵面, 显著提升了从视觉输入到工程实体的设计精度与自动化水平。

2 全息超表面实验结果

图5给出了数字“19”和“40”的端到端逆设计可视化结果, 用于验证本研究多模型协同AI网络在不

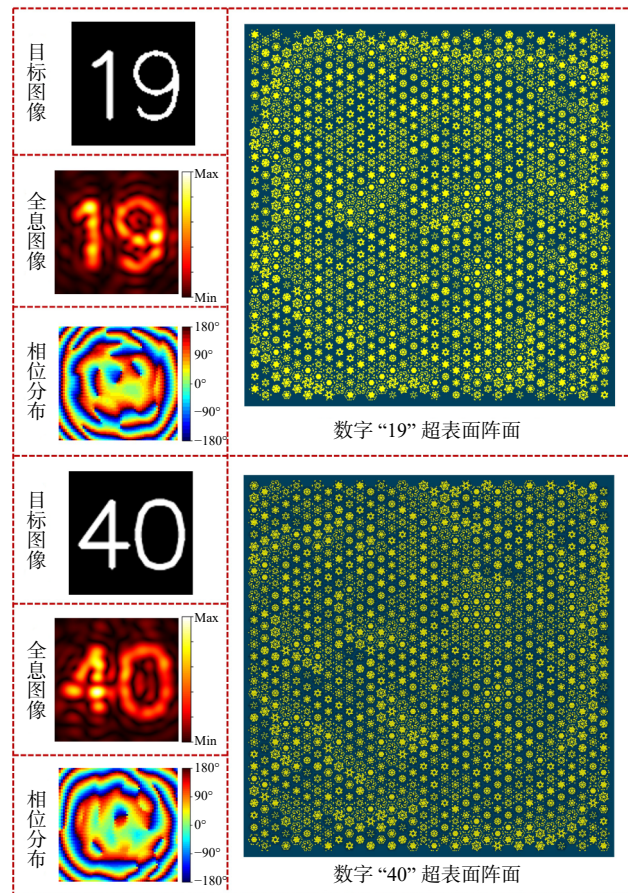


图5 数字“19”和“40”的全息重构结果与超表面设计结果

同目标形状上的可迁移性与稳定性。

对于每个样例,结果从左到右依次对应:目标强度图像、由 Pix2Pix 生成的全息强度图、U-Net 预测得到的连续相位分布,以及 VAE 映射生成组阵排布的超表面阵列结构。可以观察到,重构强度图在数字边缘处具有较高对比度,主体区域能量集中且轮廓完整,说明网络能够将目标结构转化为满足衍射成像要求的干涉纹理;同时,相位图呈现连续平滑的空间变化,而不是随机噪声状分布,这表明相位预测网络在学习全息纹理与相位之间的非线性关系时保持了必要的空间连续性约束。总体而言,图 5 的 2 组结果共同说明:该分阶段逆设计流程能够在保证结构可实现性的前提下,实现目标图像到超表面阵列的一致映射,并在典型数字图案上获得良好的重构质量与背景抑制能力。

图 6 对比展示了不同网络架构在数字(“19”“40”)、字母(“B”“I”“T”),以及复杂图形(如校徽)上

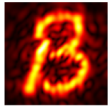
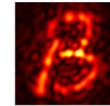
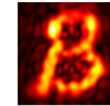
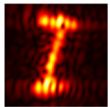
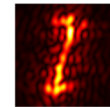
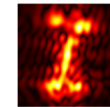
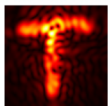
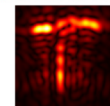
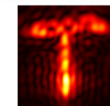
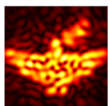
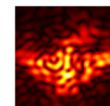
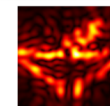
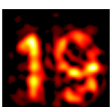
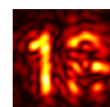
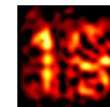
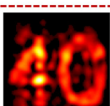
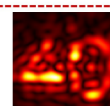
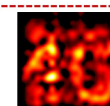
输入图像	多模型协同	GS算法	卷积网络
			
			
			
			
			
			
平均相关系数	0.8729	0.7618	0.7137
结构相似指数	0.8865	0.8469	0.8094

图 6 多模型协同 AI 网络方法、GS 算法与传统卷积网络的仿真结果比较

的全息仿真结果。左列为输入目标图像,中列为同样参数大小的传统卷积网络的仿真结果,右列为本研究多模型协同 AI 网络的仿真结果。可以看出,传统卷积网络在多数组合场景中存在较明显的能量扩散与形态塌缩现象:目标笔画边缘出现模糊、断裂或“糊团”状聚集,局部高亮区域与真实笔画位置发生偏移,同时背景中伴随较强的散斑状杂波,导致目标与背景对比度下降。相比之下,多模型协同 AI 网络能够更稳定地恢复目标几何结构,表现为:笔画边缘更清晰、内部强度分布更均匀,细长结构(如“1”“T”的竖画与横画)仍能保持连续性与方向一致性;在复杂轮廓图形上,主体区域能量更集中且背景伪影明显减弱,说明该网络在抑制旁瓣与非目标散射方面具有更强能力。上述差异表明,多模型协同 AI 网络通过更有效的特征编码与解码约束,提升了对“目标形状保持”和“背景杂波抑制”2 类矛盾指标的兼顾能力,从而获得更高的视觉可辨识度与更稳定的重构质量。

为了定量评价重构图像与标准基准图像在全局强度分布上的一致性,本研究采用皮尔逊相关系数(Pearson correlation coefficient, PCC)作为核心评价指标。PCC 通过衡量像素强度的线性相关程度反映重构图对目标整体形态与能量分布的保持能力;当背景杂波较强或目标能量外泄时,相关性会显著下降。其计算公式如下

$$Corr(I_{std}, I_{res}) = \frac{\sum_m \sum_n (I_{std, mn} - \bar{I}_{std})(I_{res, mn} - \bar{I}_{res})}{\sqrt{\left(\sum_m \sum_n (I_{std, mn} - \bar{I}_{std})^2\right) \left(\sum_m \sum_n (I_{res, mn} - \bar{I}_{res})^2\right)}} \quad (12)$$

式中, $Corr(I_{std}, I_{res})$ 代表标准图像与重构图像的相关性系数, $I_{std, mn}$ 为标准图像在位置 (m, n) 处的像素强度, $\bar{I}_{std}$ 标准图像的平均像素强度, $I_{res, mn}$ 重构图像在位置 (m, n) 处的像素强度, $\bar{I}_{res}$ 重构图像的平均像素强度。

结构相似指数(structural similarity index measure, SSIM) 综合考虑亮度、对比度和结构信息,提供更符合人眼感知的评估。其计算公式为

$$SSIM(x, y) = \frac{(2\mu_x \mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (13)$$

式中, μ_x 和 μ_y 分别表示标准图像与重构图像的平均亮

度, σ_x^2 和 σ_y^2 为对应的方差, σ_{xy} 为协方差, c_1 和 c_2 为稳定化常数。

传统卷积网络的平均相关系数和 SSIM 分别为 0.7137 和 0.8094, 而传统物理约束 GS 相位迭代算法分别为 0.7618 和 0.8469, 多模型协同 AI 网络的相关指标达到 0.8729 和 0.8865, 分别相较卷积网络提升约 22.3% 和 9.5%, 相较 GS 算法提升约 14.6% 和 4.7%。该结果与视觉对比一致: 多模型协同 AI 网络在不同类型目标(高对比数字、细笔画字母与复杂轮廓图形)上均表现出更强的鲁棒性与一致性, 说明其不仅提升局部结构保真度, 也显著改善了全局能量映射的准确性, 从而验证了本研究方法在电磁全息重构任务中的有效性。

3 讨论

本研究相比传统电磁超表面逆设计方法与现有单一深度学习网络具有显著的创新性。分阶段混合 AI 模型框架通过 Pix2Pix、U-Net 和 VAE 的协同作用, 首次系统性地建立了从视觉图像语义特征至可制造几何结构的完整映射链路, 有效克服了传统端到端映射在处理电磁强非线性问题时的收敛困难。与传统物理仿真方法相比, 该方法将设计周期从数小时的全波电磁仿真缩短至实时反演, 设计自动化程度显著提升。基于测试集统计, AI 多模型协同网络的平均皮尔逊相关系数达 0.8729, 相较传统同参数卷积网络的 0.713 提升 22.3%, 在高对比数字、细笔画字母和复杂轮廓图形等多种目标上均表现出更强的鲁棒性。该结果表明, 分阶段解耦策略通过引入中间物理变量作为约束桥梁, 显著降低了复杂映射的学习难度, 并有效保证了生成结构的物理可实现性与工程可制造性, 为图像驱动的智能超表面快速设计提供了一种高效自动化的新范式。

尽管本研究主要验证了 10 GHz 线极化、单一频点的全息成像, 但该方法的核心框架具有显著的普适性, 其理论基础支持拓展至更广泛的超表面功能器件。第一, 方法的频率可扩展性源于相位-结构映射的频率独立性。如式(10)所示, VAE 解码器本质上学习的是单元几何参数 x_{unit} 与目标相位 Φ 之间的多对一非线性关系。该映射关系虽在 10 GHz 获得训练, 但

其底层物理机制在其他频段本质相同。相位到结构映射函数 $\Phi = f(x_{\text{unit}})$ 在相对带宽 $\Delta f/f_0$ 较小的频段内仍保持连续可导性。因此, 通过迁移学习或轻量级微调, 该框架可直接适用于毫米波、太赫兹等其他频段。第二, 极化维度的拓展基于相位调控的自由度独立性。目前设计采用 x 方向线极化, 相位矩阵维度为 $\Phi \in \mathbb{R}^{m \times n}$ 。对于 y 方向线极化或圆极化器件, 所需相位矩阵维度可扩展为 $[\Phi_x, \Phi_y] \in \mathbb{R}^{m \times n \times 2}$, U-Net 结构中仅需将输入和输出通道数从 1 调整为 2 或 3, 其多尺度特征融合机制仍然有效。从数学角度看, 多极化相位预测仍遵循式(6)~式(8)的联合损失框架, 仅通道维度变化而学习策略不变, 因此该方法的理论可以用于极化维度的拓展。第三, 无论是全息成像、波束赋形、偏振转换还是各向异性调控, 一般超表面功能本质上可统一描述为: 根据空间位置 (x, y) 性能来设置局部相位 $\Phi(x, y)$ 以及幅度 $A(x, y)$ 。构建的 Pix2Pix、U-Net 和 VAE 映射框架是一个与具体功能无关的通用特征转换器, 只要为不同功能器件重新生成相应的标准相位标签, 该框架就可直接迁移, 证明了多模型协同 AI 超表面设计方法的普适性。

4 结论

本研究提出了一种基于多模型协同 AI 模型的全息超表面逆设计方法, 实现了由视觉图像输入到离散六边形单元阵列结构生成的自动化映射流程。该方法采用多模型协同框架, 依次利用 Pix2Pix、U-Net 与 VAE 完成全息纹理转换、相位分布预测及几何参数生成, 有效打通了图像语义与复杂物理结构之间的映射链路。结果表明, 分阶段逆设计策略显著降低了端到端学习难度, 在复杂字符与纹理成像任务中表现出优异的稳定性与鲁棒性。在阵面构建中, 基于相位矩阵空间平均化处理, VAE 模型成功生成满足 10 GHz 入射波调控需求的可制造六边形单元结构。通过 30×34 阵列的交错排布, 在 200 mm 处实现了高保真的电磁全息成像, 较传统同参数量的卷积网络性能提升 22.3%。该高度自动化的设计范式在保证物理可实现性的同时显著缩短了设计周期, 为图像驱动的智能超表面及大规模可编程电磁器件设计提供了一种高效的新设计思路。

参考文献(References)

- [1] Yu N F, Genevet P, Kats M A, et al. Light propagation with phase discontinuities: Generalized laws of reflection and refraction[J]. *Science*, 2011, 334(6054): 333–337.
- [2] Ren H R, Li X P, Zhang Q M, et al. On-chip noninterference angular momentum multiplexing of broadband light[J]. *Science*, 2016, 352(6287): 805–809.
- [3] Ma W, Liu Z C, Kudyshev Z A, et al. Deep learning for the design of photonic structures[J]. *Nature Photonics*, 2021, 15(2): 77–90.
- [4] Sroor H, Huang Y W, Sephton B, et al. High-purity orbital angular momentum states from a visible metasurface laser[J]. *Nature Photonics*, 2020, 14(8): 498–503.
- [5] Gu Z, Liu H Y, Zhang Z P, et al. Multi-wavelength metasurface empowered cryptography for heightened security and improved fidelity[J]. *Laser & Photonics Reviews*, 2024, 18(10): 2301014.
- [6] Cheng G, Si L M, Tang P C, et al. Study of symmetries of chiral metasurfaces for azimuth-rotation-independent cross polarization conversion[J]. *Optics Express*, 2022, 30(4): 5722.
- [7] Tang P C, Si L M, Yuan Q Q, et al. Dynamic generation of multiplexed Vortex beams by a space-time-coding metasurface[J]. *Photonics Research*, 2025, 13(1): 225.
- [8] Dong L, Si L M, Liu Y Z, et al. Speech-controlled reconfigurable intelligent metasurface for real-time wireless power transfer and communication[J]. *Research*, 2025, 8: 831.
- [9] 张航, 董琳, 汤鹏程, 等. 双频太赫兹圆极化吸波与异常反射手性超表面[J]. *电子学报*, 2023, 51(10): 2690–2699.
- [10] Zhou Z, Wei Z H, Ren J, et al. End-to-end machine-learning framework for electromagnetic inverse design: From practical constraints to optimized structures[J]. *IEEE Transactions on Microwave Theory and Techniques*, 2025, 73(11): 8690–8708.
- [11] Qian C, Kaminer I, Chen H S. A guidance to intelligent metamaterials and metamaterials intelligence[J]. *Nature Communications*, 2025, 16: 1154.
- [12] Hu S G, Li M Y, Xu J W, et al. Electromagnetic metamaterial agent[J]. *Light: Science & Applications*, 2025, 14: 12.
- [13] Wu G H, Si L M, Dong L, et al. Dynamic manipulation of high-order Poincaré sphere polarization states via a transmissive space-time-coding metasurface[J]. *Photonics Research*, 2025, 13(10): 2877.
- [14] Luo Y T, Li P Q, Li D T, et al. Probability-density-based deep learning paradigm for the fuzzy design of functional metastructures[J]. *Research*, 2020, 2020: 8757403.
- [15] Zheng X Y, Zhang X B, Chen T T, et al. Deep learning in mechanical metamaterials: From prediction and generation to inverse design[J]. *Advanced Materials*, 2023, 35(45): 2302530.
- [16] Wu G H, Si L M, Xu H Y, et al. Phase-to-pattern inverse design for a fast realization of a functional metasurface by combining a deep neural network and a genetic algorithm[J]. *Optics Express*, 2022, 30(25): 45612–45623.
- [17] Si L M, Niu R, Dang C Y, et al. Advances in artificial intelligence for artificial metamaterials[J]. *APL Materials*, 2024, 12(12): 120602.
- [18] 司黎明, 马天宇, 党晨阳, 等. 无线输能超表面技术研究进展[J]. *科技导报*, 2025, 43(18): 23–40.
- [19] Yuan H, Zhong Z Q, Zhang B. Neural network assisted end-to-end design for broadband high diffraction efficiency full-color meta-holograms[J]. *Laser & Photonics Reviews*, 2025, 19(24): e00946.
- [20] Khatib O, Ren S M, Malof J, et al. Deep learning the electromagnetic properties of metamaterials: A comprehensive review[J]. *Advanced Functional Materials*, 2021, 31(31): 2101748.
- [21] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Advances in Neural Information Processing Systems: Vol. 27. New York: Curran Associates, Inc, 2014: 2672–2680.
- [22] Kingma D P, Welling M. Auto-encoding variational bayes[C]//Proceedings of the 2nd International Conference on Learning Representations (ICLR 2014). Banff: ICLR, 2014: 1–14.
- [23] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models[C]//Advances in Neural Information Processing Systems 33. New York: Curran Associates, Inc, 2020: 6840–6851.
- [24] Kang C, Seo D, Boriskina S V, et al. Adjoint method in machine learning: A pathway to efficient inverse design of photonic devices[J]. *Materials & Design*, 2024, 239: 112737.
- [25] Lee D, Chen W, Wang L W, et al. Data-driven design for metamaterials and multiscale systems: A review[J]. *Advanced Materials*, 2024, 36(8): 2305254.
- [26] Eybposh M H, Caira N W, Atisa M, et al. DeepCGH: 3D computer-generated holography using deep learning[J]. *Optics Express*, 2020, 28(18): 26636.
- [27] Zheng X Y, Shiomi J, Yamada T. Optimizing metamaterial inverse design with 3D conditional diffusion model and data augmentation[J]. *Advanced Materials Technologies*, 2025, 10(14): 2570074.
- [28] Karniadakis G E, Kevrekidis I G, Lu L, et al. Physics-

- informed machine learning[J]. *Nature Reviews Physics*, 2021, 3(6): 422–440.
- [29] Karras T, Aittala M, Laine S, et al. Alias-free generative adversarial networks[EB/OL]. [2026–03–21]. <https://arxiv.org/abs/2106.12423>.
- [30] Saharia C, Chan W, Chang H W, et al. Palette: Image-to-image diffusion models[C]//Proceedings of Special Interest Group on Computer Graphics and Interactive Techniques Conference Proceedings. New York: ACM, 2022: 1–10.
- [31] Honkamaa J, Khan U, Koivukoski S, et al. Deformation equivariant cross-modality image synthesis with paired non-aligned training data[J]. *Medical Image Analysis*, 2023, 90: 102940.
- [32] Dalmaz O, Yurt M, Ukur T. ResViT: Residual vision transformers for multimodal medical image synthesis[J]. *IEEE Transactions on Medical Imaging*, 2022, 41(10): 2598–2614.
- [33] Özbey M, Dalmaz O, Dar S U H, et al. Unsupervised medical image translation with adversarial diffusion models[J]. *IEEE Transactions on Medical Imaging*, 2023, 42(12): 3524–3539.

Image-driven metasurface inverse design via multi-model collaborative artificial intelligence

SI Liming^{1,2,3,4,5}, NIU Rong^{1,2,3}, DANG Chenyang^{1,2,3}, WANG Zhaorui^{1,2,3}, MA Tianyu^{1,2,3}, LI Yan⁵, ZHU Weiren⁶, SUN Houjun^{1,2,3}

1. School of Integrated Circuits and Electronics, Beijing Institute of Technology, Beijing 100081, China

2. State Key Laboratory of Environment Characteristics and Effects for Near-space, Beijing 100081, China

3. Tangshan Research Institute, Beijing Institute of Technology, Tangshan 063000, China

4. State Key Laboratory of Millimeter Waves, Nanjing 210096, China

5. Faculty of Engineering, Shenzhen MSU-BIT University, Shenzhen 518172, China

6. School of Integrated Circuits, Shanghai Jiao Tong University, Shanghai 200240, China

Abstract With the rapid development of metasurfaces for precise electromagnetic wave manipulation, the efficient and automated design of complex functional devices has become a critical research challenge. This paper proposes a multi-model collaborative framework for metasurface inverse design driven by visual images, aiming to automatically generate electromagnetic metasurface array structures directly from visual image inputs. The method employs a synergistic multi-model strategy to establish an end-to-end mapping from image semantic features to physical metasurface structural parameters. Specifically, a conditional generative adversarial network (Pix2Pix) is first employed to transform visual images into target holographic textures, followed by a U-shaped convolutional neural network (U-Net) for high-fidelity phase distribution prediction. Subsequently, a variational autoencoder (VAE) is introduced to map the desired phase profiles to manufacturable unit-cell geometries, effectively bridging the long-standing gap between phase synthesis and structural modeling in conventional design workflows. Numerical simulations demonstrate that the designed holographic metasurfaces can accurately reconstruct target holographic patterns at designated operating frequencies, achieving significantly improved reconstruction fidelity and design stability compared with traditional single-network approaches. This work provides an efficient and automated new paradigm for image-driven intelligent metasurface design and offers a promising solution for the rapid development of complex electromagnetic functional devices.

Keywords artificial intelligence; multi-model collaborative framework; metasurfaces; inverse design; visual images; neural networks ●



(责任编辑 徐丽娇)