引用格式：WU Jiawei, MIAO Yongqiang, ZHANG Zhiguang, et al. Wavefront Reconstruction Using Phase Diversity Based on the MobileViT Model[J]. Acta Photonica Sinica, 2026, 55(1):0101001
吴佳伟,苗永强,张之光,等.基于MobileViT模型的相位差法波前复原[J].光子学报,2026,55(1):0101001

基于MobileViT模型的相位差法波前复原

吴佳伟[1],苗永强[1],张之光[1],杨慧珍[2]
(1 江苏海洋大学 电子工程学院,连云港 222005)
(2 金陵科技学院 网络与通信工程学院,南京 211169)

摘 要:基于MobileViT模型的相位差波前复原方法实现高效、非迭代的波前重构。以不同湍流条件下的波前像差作为复原对象,分析基于MobileViT的波前重构方法的复原精度。为便于对比,同时给出了CNN、ResNet-50和EfficientNet-B0的复原结果。实验结果表明,随着湍流强度的增大,MobileViT的复原精度显著优于其它三种模型,表现出强湍流条件下的适应性,同时兼顾了轻量化与高精度,为强湍流条件下的无波前探测自适应光学系统波前校正提供了新的解决方案。

关键词:自适应光学;MobileViT模型;相位差法;Zernike系数;波前复原
中图分类号:O436 **文献标识码**:A **doi**:10.3788/gzxb20265501.0101001

0 引言

自适应光学(Adaptive Optics, AO)是一种通过测量和校正波前畸变以提升光学系统性能的技术[1],广泛应用于显微成像[1]、激光大气传输[2]等领域。AO系统可以分为两大类:有波前传感器AO系统和无波前传感器AO系统[3]。AO系统中,波前畸变通常由Zernike多项式表示,其系数量化了诸如倾斜、离焦和像散等像差的大小。因此,精确估计Zernike系数是实现高质量成像和波前校正的核心。传统AO系统中,Zernike系数通常由Shack-Hartmann波前传感器测量局部波前斜率来估算,然后由波前校正器进行实时补偿。然而,这种方法依赖复杂的硬件设备,且在强湍流条件下易受测量噪声和计算复杂度的限制。无波前传感器AO系统一般通过优化控制算法来校正波前,典型的控制方法有随机并行梯度下降算法[5]、模拟退火法[7]、遗传算法[8]等。这种方法收敛速度慢,难以用于实时波前校正系统。

随着人工智能的快速发展,深度学习技术[9]引起了广泛关注,许多神经网络算法应用到AO系统控制中[10]。2018年,PAINE S W等提出利用卷积神经网络(CNN)从单帧点扩散函数(Point Spread Function, PSF)图像预测出Zernike系数的初始估计以辅助非线性优化,降低大型空间望远镜部署和校准失败的风险[11]。针对当前无波前传感器自适应光学系统中波前像差检测速度慢、测量精度低的问题,LI Yushuang等提出一种基于深度学习的自适应光学波前畸变预测方法[12],仅使用1帧焦平面退化图像预测出Zernike系数。尽管上述方法在研究中取得了良好的性能,但从单帧强度图像预测Zernike系数理论上存在一定局限性。由于傅里叶变换的对称性,即使Zernike系数的符号相反,也会生成相同的PSF图像,导致预测的Zernike系数存在符号不唯一的问题。SIDDIK A B等研究了从单帧强度图像预测“修正后的”Zernike系数[13],一定程度上缓解了符号模糊性带来的影响。

相位差法(Phase Diversity, PD)通过同时获取焦平面与离焦平面的强度图像,引入额外相位信息,从而提升波前重构的准确性。GONSALVES R A于1982年首次提出该方法[14],并结合极大似然估计构建迭代优化模型,实现波前重建。尽管PD方法有效缓解了符号模糊性,但其迭代优化过程仍存在计算复杂度高、

基金项目:国家自然科学基金(12473081, U2141255),自适应光学全国重点实验室基金(FNLAO-24-MS-001)
第一作者:吴佳伟,wujiawei@jou.edu.cn
通讯作者:杨慧珍,yanghz@jit.edu.cn
收稿日期:2025-07-27;**录用日期**:2025-10-28

收敛速度慢等问题。为提升波前复原效率，研究者们将深度学习与PD技术相结合，利用神经网络在非迭代框架中快速预测Zernike系数。马慧敏等[15][16]分别基于CNN与GoogLeNet构建了Zernike系数预测方法，指出较深的网络结构有助于提高湍流恢复的精度。WU Yu等提出一种相位分集卷积神经网络（PD-CNN）方法[17]，建立了强度图像与对应像差系数的非线性映射，无需复杂的迭代变换和优化过程，极大缩短了推理时间。2024年，KOK Y E等[18]利用ResNet-50模型从三帧相位差异图像直接预测Zernike系数，为像差校正提供了更高效的解决方案。2025年，许元斋等[19]提出基于深度学习的波前探测器，利用cGAN模型建立光强-相位映射反演波前，并将其运用到自适应光学系统中。但上述方法在大湍流条件下，预测精度有限。MobileViT模型采用了卷积和Transformer混合架构，使得MobileViT模型既具有卷积的性质，又能进行全局的建模，以该架构作为特征提取逻辑，MobileViT模型能够高效处理复杂视觉任务，并在资源受限场景下保持高精度。目前，MobileViT模型已成功应用于医学图像处理[20]，农业监测[21]等工程领域。

鉴于现有方法在强湍流条件下波前复原精度低的问题，本文在PD方法理论基础上，提出一种基于MobileViT的波前重构方法，该方法基于焦面和离焦面两帧光强图像，自动提取光强图像特征实现非迭代相位复原。以不同湍流条件下的波前像差作为复原对象，分析基于MobileViT的波前重构性能。便于对比，同时给出了CNN、ResNet-50和EfficientNet-B0的复原结果，为实际应用提供参考。

1 理论与方法

基于深度学习模型的无波前探测自适应光学系统原理如图1所示，畸变波前经变形镜补偿后，由成像透镜聚焦至焦点和离焦位置处的两个图像传感器，分别生成焦面图像和离焦面图像。将两幅图像收集之后，输入到模型中，模型根据输入的图像输出对应的Zernike系数，然后由波前控制器根据预测得到的Zernike系数计算出变形镜变形所需要的驱动信号，从而实现系统闭环控制。

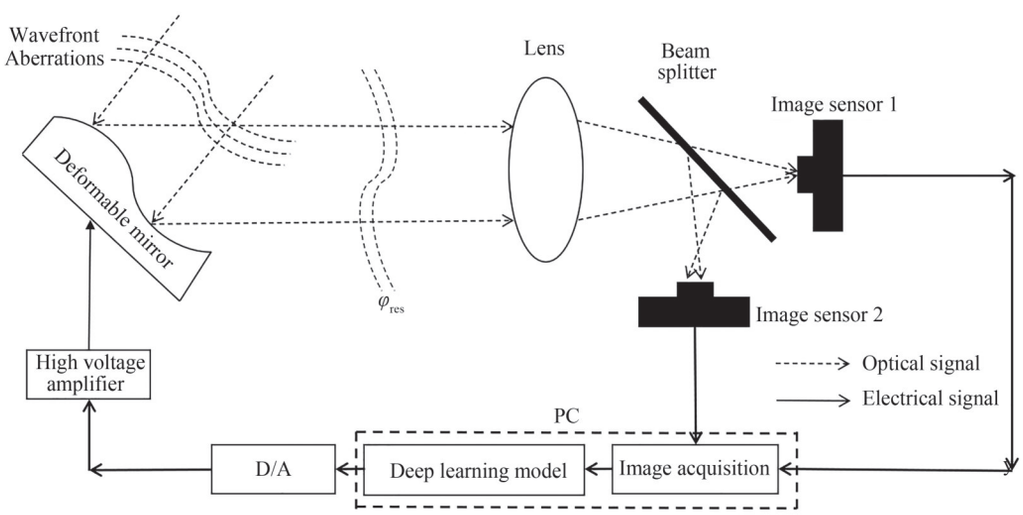


图1 基于深度学习的无波前探测自适应光学系统

Fig. 1 Sensorless adaptive optics system based on deep learning

1.1 MobileViT模型

MobileViT是一种轻量化深度学习模型，由MEHTA S和RASTEGARI M于2022年提出[22]，旨在结合卷积神经网络（CNN）和视觉变换器（Vision Transformer, ViT）的优势，为资源受限设备提供高效的视觉任务处理方案。相较于传统深度学习模型（如ResNet-18至ResNet-152）通过增加网络深度或宽度提升性能，MobileViT创新性地采用混合架构，在保持低计算复杂度的同时显著提升精度。

MobileViT的核心在于其混合架构，融合了CNN的局部特征提取能力和Transformer的全局建模能力。模型包含两个关键模块：MV2Block和MobileViT Block，其结构分别如图2和图3所示。MV2Block借鉴MobileNetV2[23]的倒残差结构，通过深度可分离卷积，以较低的计算成本提取局部空间特征。其基本单元由

卷积层（Convolution，Conv）、批量归一化（Batch Normalization，BN）以及SiLU激活函数（Sigmoid Linear Unit，SiLU）组成，旨在保持低计算复杂度的同时保证特征提取能力。

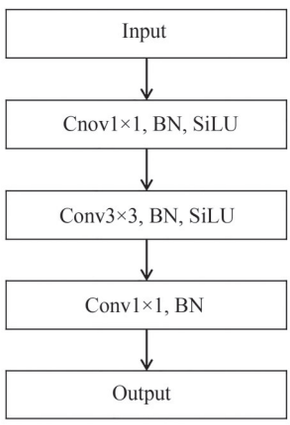


图2　MV2Block网络结构

Fig. 2　MV2Block model structure

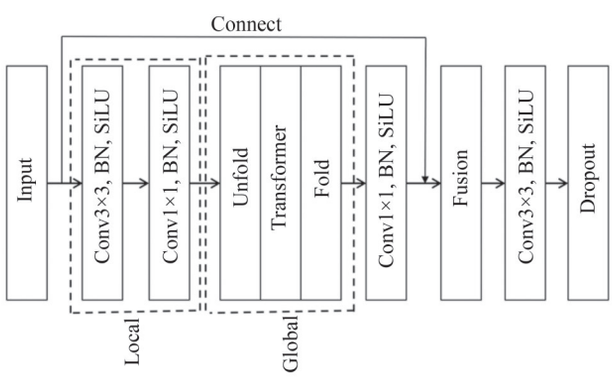


图3　MobileViT block网络结构

Fig. 3　MobileViT block model structure

MobileViT Block是该模型的核心架构，每个卷积层都包含BN和SiLU激活函数，该模块先由Conv3×3局部提取图像的特征信息，然后利用Conv1×1进行通道压缩，并将特征图划分为多个patch输入Transformer模块。通过自注意力机制捕捉它们的全局依赖关系，通过Conv1×1将特征重新整合为空间特征图，并与原始局部特征融合（Feature Fusion），再通过Conv3×3输出，并加入Dropout层提升泛化能力。MobileViT结构如图4所示，图片通过MV2Block结合MobileViT Block处理后，最后经过全连接层（Fully Connected，FC）输出像差系数。

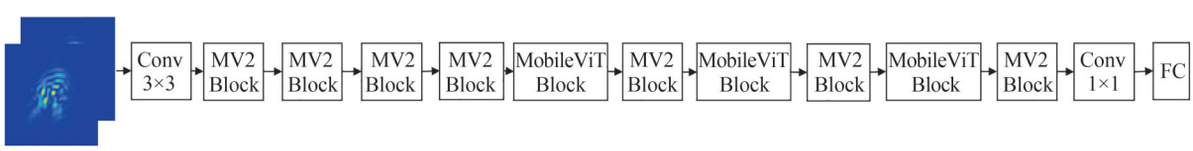


图4　MobileViT模型结构

Fig. 4　MobileViT model structure

在波前复原任务中，MobileViT的混合架构展现了独特优势。MV2Block通过深度可分离卷积高效提取PSF图像的局部细节，而Transformer模块通过自注意力机制捕捉湍流条件下波前畸变的全局特性。这种局部与全局特征的协同作用可显著提升Zernike系数的预测精度，适用于复杂波前建模需求。

1.2　其他常用模型

为了评估本文提出的MobileViT模型在波前重构任务中的性能，另外选择了三种典型深度学习模型作

为对比对象：传统卷积神经网络CNN、ResNet-50和EfficientNet-B0。CNN结构简单、计算量小，可作为基础卷积网络的代表；ResNet-50通过残差连接捕捉更深层特征，有助于复杂波前的恢复；EfficientNet-B0则是一种轻量化网络，具有较高的参数效率，能够在保证精度的同时降低计算开销。拟从重构精度、模型复杂度和计算效率等多维度全面评估MobileViT的性能优势。

图5中，采用的CNN[24]模型共有8层网络结构，包含4个卷积层，1个展平层和3个全连接层。模型输入为尺寸为128×128的双通道图像。卷积层均采用3×3的卷积核，步长为1，填充为1，输出通道数依次为32、64、128和256。每层卷积后均接BN和ReLU激活函数，并使用2×2的最大池化层进行下采样。经过四层卷积与池化后，特征图被展平并输入至全连接层，通道维度依次为512、128，激活函数为ReLU，并在中间加入Dropout防止过拟合。最后的输出层节点数为33，对应于需要预测的Zernike系数个数。

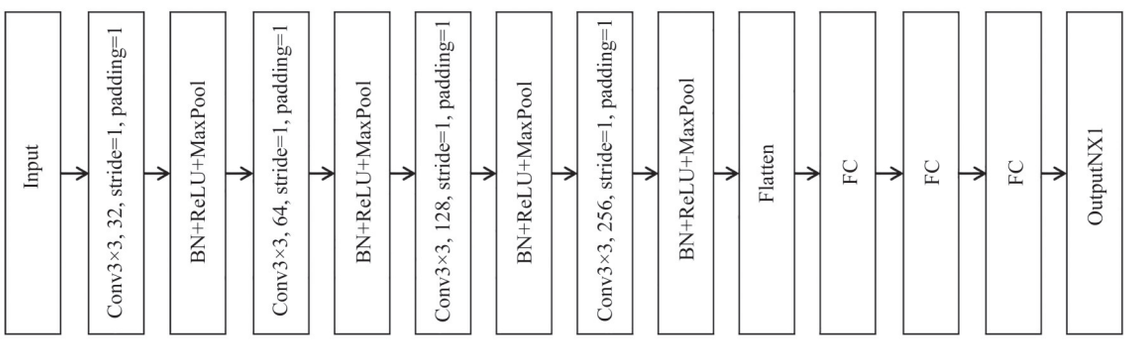


图5 CNN模型结构

Fig. 5 CNN model structure

ResNet模型由何凯明[25]等于2015年提出，解决了深度神经网络随着层数的增加，性能反而退化的问题。核心思想是引入残差学习模块，学习输入与目标之间的残差，这种结构能使梯度更有效的反向传播，解决了深层网络的优化难题。

残差模块如图6所示，残差定义为真实值与预测值之间的差值。在深度神经网络中，残差学习通过引入残差映射来解决深层网络的退化问题。假设目标映射为$H(x)$，表示期望的输出特征映射，其中x为前一层或前一残差块的输出。传统方法直接学习$H(x)$，而残差学习将目标转化为学习残差映射$F(x)=H(x)-x$。因此，原始映射可表示为$F(x)=H(x)+x$。当网络深度增加时，可能会出现退化问题，即深层网络的性能劣于较浅层网络。残差学习的优势在于，当残差映射$F(x)=0$时，$H(x)=x$，即新层等价于恒等映射。此时，深层网络的性能至少与浅层网络相当，从而避免退化问题。在实际训练中，残差$F(x)$通常不为零，但通过优化$F(x)$，网络能够逐步逼近最优解。相比直接学习$H(x)$，学习残差$F(x)$通常需要较小的权重更新，这使得深层网络更容易优化，并能有效学习新的特征。

文献[12]中，ResNet-50模型在ResNet系列网络中损失最小，表现最好，本文采用ResNet-50模型作为对比。ResNet-50模型的具体网络结构如图7所示。模型以128×128×2的双通道PSF图像作为输入，首先通过

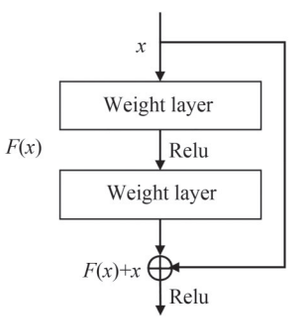


图6 残差模块

Fig. 6 Residual module

初始卷积层和最大池化层，将特征图尺寸降至32×32。随后，网络进入4个残差阶段，分别包含3、4、6和3个残差模块，通道数依次为64、128、256和512。特征图尺寸在每个阶段逐步减小，最终为2 048×4×4。之后，通过全局平均池化将特征图降维至2 048×1×1，展平为2 048维向量，最后通过全连接层输出Zernike系数。由于本文预测3～35阶Zernike系数，因此将最后一层全连接层的输出维度调整为33，即输出3～35阶Zernike系数。

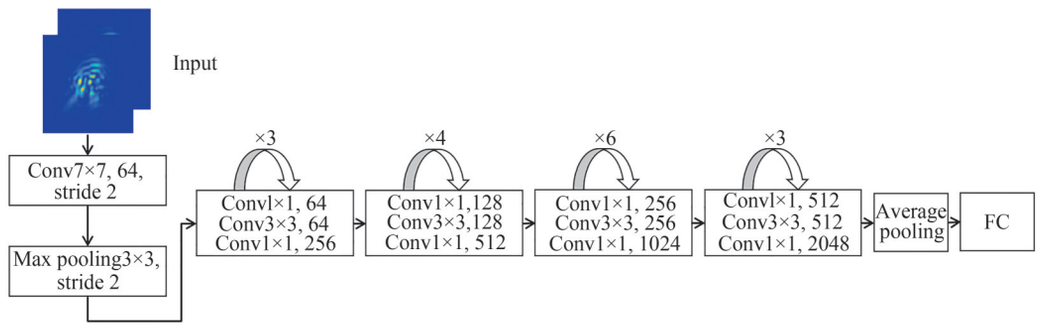


图7　ResNet-50模型结构

Fig. 7　ResNet-50 model structure

EfficientNet系列是由Google于2019年提出的一类卷积神经网络，旨在保持较低计算量和参数量的前提下，获得更高的图像分类精度。其核心思想是复合缩放：同时对网络的深度、宽度和输入分辨率进行统一缩放，从而在不同计算预算下实现网络性能最优化。相比传统CNN网络，EfficientNet系列在相同精度下显著减少了参数量和FLOPs。

其网络结构主要由MBConvBlock构成，如图8所示。MBConvBlock与ResNet的残差结构类似，同样采用shortcut连接来加深网络层数。其主要由MBConv和DropConnect两部分组成。其中，MBConv内部也包含shortcut结构，但不同之处在于其中引入了Squeeze-and-Excitation（SE）模块。SE模块通过对每个通道计算权重，并与原始特征图相乘，从而增强模型对重要通道特征的敏感性，仅增加少量计算量即可提升表示能力。同时，DropConnect替代了传统的Dropout方法，不再随机丢弃隐藏层节点的输出，而是随机丢弃残差路径的输入，以有效缓解过拟合。

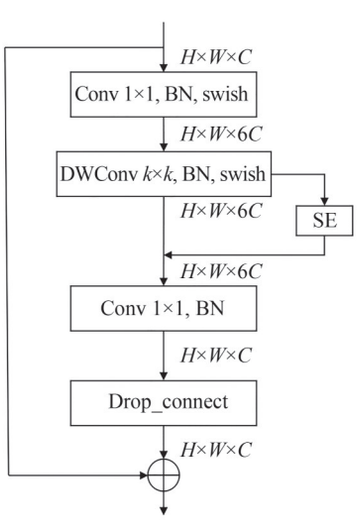


图8　MBConvBlock网络架构

Fig. 8　MBConvBlock network structure

EfficientNet系列提出了B0～B7不同规模的网络，本文采用基础模型B0，其结构如表1所示。表1详细列出了EfficientNet-B0网络架构，包括网络阶段、模块类型、分辨率、卷积核大小、输入张量大小、输出通道数及层数等信息。EfficientNet-B0网络通常输入为224 × 224 × 3的图像，本文输入128 × 128 × 2的图像，输出为33阶的Zernike系数。

表1 EfficientNet-B0的网络架构

Table 1 Network architecture of EfficientNet-B0

Stage	Module type (including kernel size and expansion factor)	Input resolution	Output channels	Layers
1	Conv 3×3	128×128	32	1
2	MBConv1, k3×3	64×64	16	1
3	MBConv6, k3×3	64×64	24	2
4	MBConv6, k5×5	32×32	40	2
5	MBConv6, k3×3	16×16	80	3
7	MBConv6, k5×5	8×8	112	3
7	MBConv6, k5×5	8×8	192	4
8	MBConv6, k3×3	4×4	320	1
9	Conv1×1 & Pooling & FC	4×4	33	1

2 模拟环境

2.1 像差表示

波前畸变可由圆域的Zernike多项式进行描述，残余相位$\phi(r)$可以看作是Zernike多项式的线性组合

$$\phi(r)=\sum_{i=1}^{M}a_i Z_i(r) \tag{1}$$

式中，M为Zernike阶数；a_i为第i阶Zernike系数；$Z_i(r)$为第i项Zernike多项式。

Zernike多项式通常按照Noll序数排列[26]，前几阶分别对应倾斜(Z_1,Z_2)、离焦(Z_3)等光学像差。考虑到实际大气湍流导致的波前像差中，低频成分占绝大数，本研究选取3～35阶Zernike系数作为研究对象，探究深度学习模型在这些像差上的重构能力。一般来讲，仅从单幅强度图像难以准确重构出Zernike系数，这是由于波前相位和强度图像之间存在着多对一的映射关系。例如，某些Zernike多项式Z_i会在正负符号相反的Zernike系数下生成相同的PSF图像，从而引入符号模糊性。这种特性使得基于一副图像强度信息直接预测Zernike系数时面临困难，通常引入相位差法来解决符号模糊性问题[27]。核心思想在于同时采集两幅PSF图像：一幅为焦平面PSF，另一幅为引入已知相位的PSF，已知相位通常为一定大小的离焦像差。这两幅图像的差异带来的信息冗余使得精确预测带符号的Zernike系数成为可能。

在焦平面，已知残余相位$\phi(r)$，对应的PSF计算公式为

$$I_1(u,v)=\left|F\left\{p(x,y)\mathrm{e}^{\mathrm{j}\phi(x,y)}\right\}\right|^2 \tag{2}$$

式中，F为傅里叶变换；$p(x,y)$为孔径函数；(x,y)为相位平面坐标；(u,v)为焦平面坐标。引入一个已知的离焦相位$\phi_d(x,y)=aZ_3$，其中a为常数。本文根据大量的模拟实验发现a为2π时，既能看出焦平面和离焦平面PSF的差异，模型又能表现出良好的性能，此时$\phi_d(x,y)$的RMS值为1λ。总的波前相位变为

$$\phi_d'(x,y)=\phi(x,y)+\phi_d(x,y) \tag{3}$$

离焦平面PSF计算为

$$I_2(u,v)=\left|F\left\{p(x,y)\mathrm{e}^{\mathrm{j}\phi_d'(x,y)}\right\}\right|^2 \tag{4}$$

相位差法的优势在于在成像系统中引入已知像差然后采集图像，利用两幅图像来消除相位恢复的不确定性，通过将两幅图像输入到深度学习模型中，可以直接预测带符号的Zernike系数值。

2.2 数据集的生成

根据Kolmogorov湍流理论，可以得Zernike多项式系数向量$\boldsymbol{A}=\{a_1,a_2,\cdots,a_N\}$的协方差矩阵$\boldsymbol{C}$[28]，则有

$$\left\langle a_i a_j\right\rangle=c_{ij}\left(\frac{D}{r_0}\right)^{5/3} \tag{5}$$

式中，c_{ij}为协方差系数，a_i和a_j为第i和第j阶Zernike系数。D/r_0表征光波受湍流影响的大小，其中D为望远镜口径，r_0为大气相干长度。系数矩阵A可由Karhumen-Loeve多项式求得

$$\boldsymbol{C}=\mathrm{V}\mathrm{S}\,\boldsymbol{V}^{\mathrm{T}} \tag{6}$$

$$\boldsymbol{A}=\mathrm{VB} \tag{7}$$

式中，$\boldsymbol{S}$是对角阵；$\boldsymbol{V}$是Karhumen-Loeve多项式系数矩阵；$\boldsymbol{B}$是相位$\phi(r)$的系数。生成特定大小的Zernike系数，再由$\phi(r)=\sum_{i=1}^{M}a_i Z_i(r)$计算得到相应畸变波前，焦平面和离焦平面的PSF图像由式(2)和(4)可得。本测试了D/r_0分别为5、10、15和20时的4种不同湍流条件下的数据集，每种湍流随机生成15 000组作为训练集，1 000组作为测试集。

3 仿真结果与分析

3.1 模型训练结果

模拟实验基于Windows服务器构建的PyTorch深度学习环境中进行，所使用的GPU为Nvidia GeForce 4060。构建了CNN、ResNet-50、EfficientNet-B0和MobileViT模型，本文所用的ResNet-50与EfficientNet-B0模型结构均基于PyTorch官方库torchvision.models提供的实现，并在输入层和输出层做了相应修改以适应本研究任务，所有模型均使用预先建立的数据集从头开始训练。每个网络均训练400个周期，批量大小设置为64。

采用平均绝对误差(Mean Absolute Error, MAE)损失函数作为目标函数

$$\mathrm{MAE}=\frac{1}{M}\sum_{i=1}^{M}|\hat{a}_i-a_i| \tag{8}$$

式中，M为Zernike阶数；$\hat{a}_i$为预测的第i阶Zernike系数；a_i为真实的Zernike系数。优化器采用Adam优化器，其能够自主调节学习率，在不同阶段能保持较好的性能。与其他的优化器相比，Adam具有高效性和较好的鲁棒性。学习率衰减策略采用余弦退火衰减，CNN，ResNet-50，EfficientNet-B0和MobileViT的初始学习率设为5×10^{-4}。

图9给出了不同湍流条件下、不同深度学习模型对应的测试集损失曲线，其中红色实线为CNN，青色点线为EfficientNet-B0，蓝色点划线为ResNet-50，黑色虚线为MobileViT。从图9中可以看出，不同湍流条件

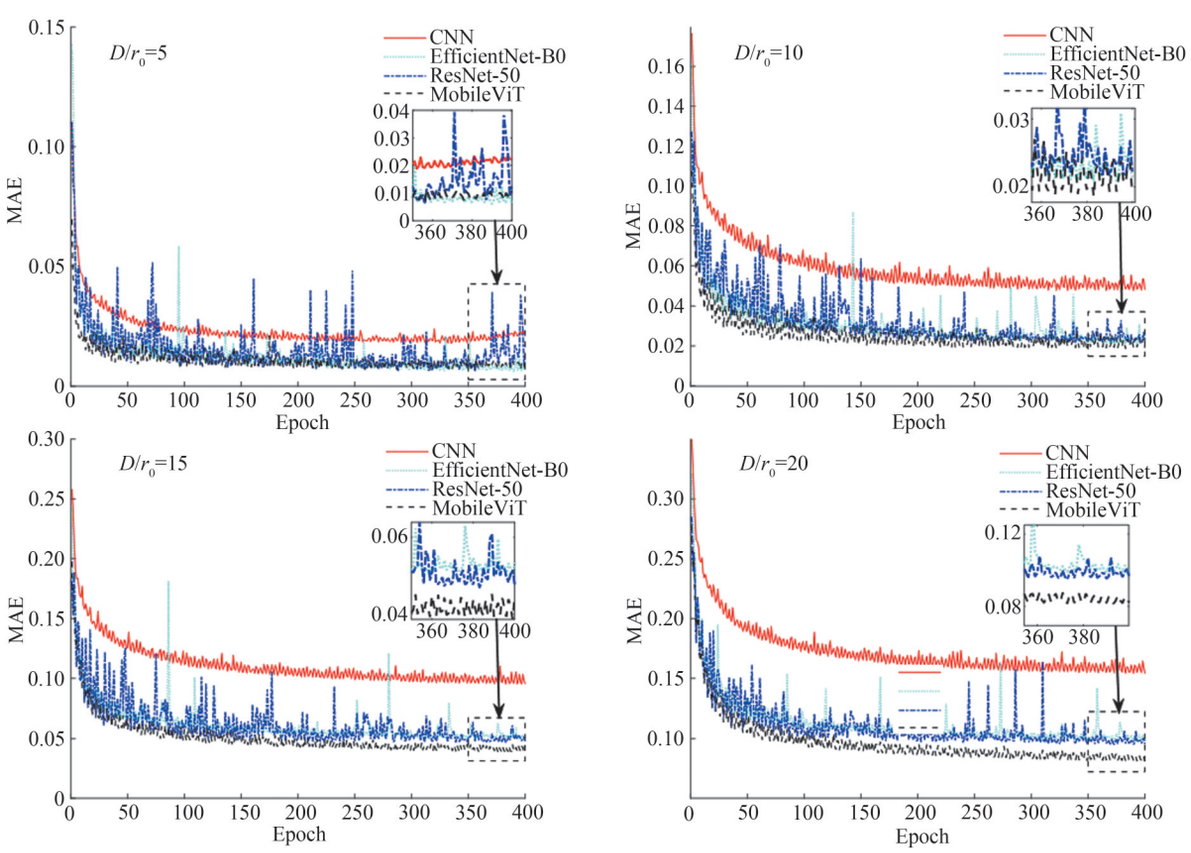


图9 不同湍流条件下训练集的损失曲线

Fig. 9 Loss curves of the training set under different turbulence levels

下，三种模型约经过400次训练过后，测试集损失都趋于稳定。从测试结果上看，CNN模型在不同湍流条件下的损失都高于其他三种模型。$D/r_0=5$时，其它三种模型的损失曲线交织在一起，结果差异不大。但是，从$D/r_0=10$中等强度的湍流条件开始，随着湍流强度的增加，三种模型之间的差异越来越大。从不同湍流条件下的损失值可以看出，MobileViT模型损失最小，ResNet-50模型与EfficientNet-B0模型损失曲线基本交织在一起，损失最大的为CNN模型。

3.2 波前复原结果分析

为全面评估不同模型在不同大气湍流条件下的波前重建与图像复原性能，对比了CNN、ResNet-50、EfficientNet-B0与MobileViT在不同湍流强度下的重构结果，结果如图10所示，其中RMS（Root Mean Square）为复原后的残余波前均方根，单位为波长λ；SR(Strehl Ratio)为残余波前对应的PSF图像的斯特列尔比。

从图10可以看出，与初始RMS和SR值相比，三种模型重构后的残余波前RMS值都大大降低、残余波前对应的PSF的SR都大大提升，证明了三种模型用于波前复原的有效性。但随着湍流强度D/r_0的增大，所有模型的RMS误差均有所上升、SR值有所下降，表明三种模型复原强湍流条件下的波前畸变时，重构精度都有所下降。进一步对比不同模型、不同湍流条件下的重构结果可以发现，CNN模型的复原精度整体最低；而EfficientNet-B0模型与ResNet-50模型的复原结果相近，二者在大多数湍流条件下表现差别不大；相比之下，MobileViT模型在不同湍流条件下均保持最低的RMS值和最高的SR值，表现出最优的复原能力。在弱湍流时（$D/r_0=5$），所有模型均有很好的复原结果，CNN相比于其它模型复原结果较差，其余三种模型复原精度基本一致。随着湍流强度增加，从$D/r_0=10$开始，CNN模型复原后的残余波前的RMS最大（对应的SR最小），表明复原精度最低；ResNet-50模型与EfficientNet-B0模型的复原精度处于中等水平，复原精度最高的是MobileViT模型，残余波前RMS最小（对应的SR最大），这表明MobileViT模型对强湍流条件下的波前畸变鲁棒性强。尤其在$D/r_0=20$时，MobileViT模型对应的残余波前RMS为0.073λ，相比于ResNet-50模型的0.109λ和EfficientNet-B0模型的0.112λ降低了33.03%和34.82%、相较CNN的0.176λ降低了58.52%；此时MobileViT对应的SR值为0.792，相较于ResNet-50的0.670和EfficientNet-B0模型的0.650，提升了18.21%和21.85%，相较于CNN的0.378则提升了109%。由以上分析可知，对于强湍流条件下的波前像差，基于MobileViT模型的波前重构方法有很强的适应性、复原精度高。

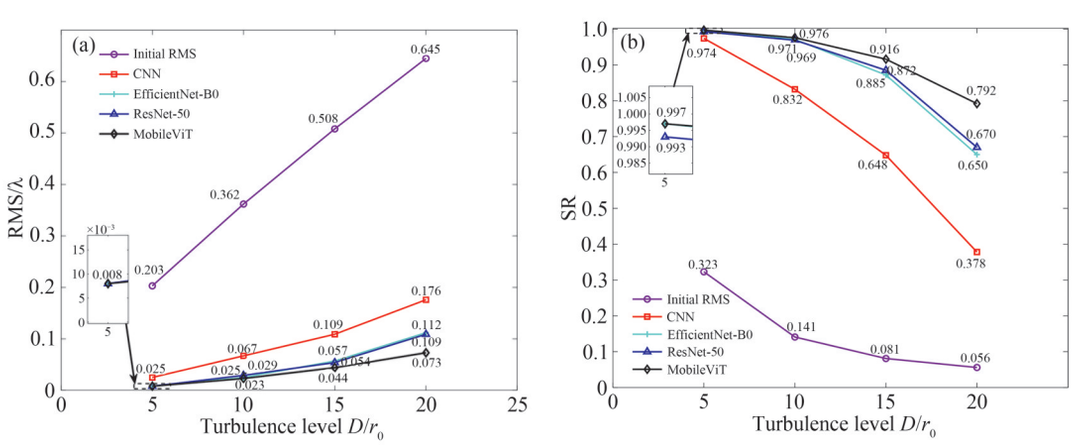


图10 不同湍流条件下，1 000组测试集重构前后的平均RMS和平均SR对比。(a) RMS对比；(b) SR对比

Fig.10 Comparison of averaged RMS and SR before and after reconstruction under different turbulence conditions based on 1 000 test samples. (a) Comparison of RMS; (b) Comparison of SR

图11分析了Zernike预测值与真实值的MAE，计算如式(8)所示。图11中可以看出湍流越强，各个模型的MAE呈上升趋势。$D/r_0=5$时，CNN模型的精度最低，其余三种模型的MAE一致，可以作为三者的起点，随着湍流增强，CNN的曲线陡峭程度最大，上升的最快，从开始的0.018 rad上升到0.154 rad；EfficientNet-B0模型从0.006 rad上升到0.098 rad；ResNet-50模型从0.006 rad上升到0.095 rad；最稳定的是MobileViT，从0.006 rad上升到0.070 rad，综上可以看出MobileViT在不同湍流下受影响最小，具有良好的精确性和稳定性。

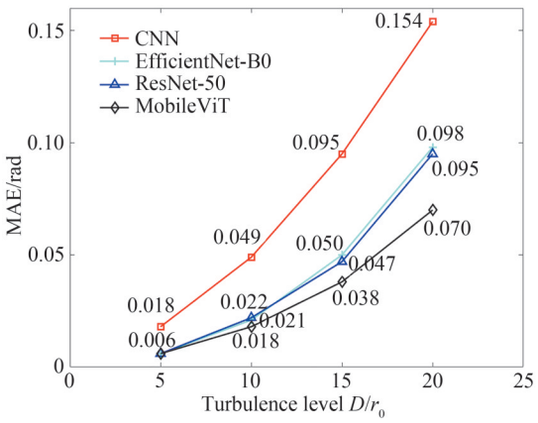


图 11　不同湍流强度下Zernike系数真实值与预测值的MAE对比

Fig. 11　Comparison of MAE between true and predicted Zernike coefficients under different turbulence intensities

表2是各模型的参数量、训练时间及推理时间统计。从表中可以看出，CNN模型虽然在参数量(0.98 M)、训练时间(62.66 min)和推理时间(0.63 ms)上占据明显优势，但复原精度明显较低，难以满足复杂湍流条件下的应用需求。ResNet-50在推理时间上具有一定优势(5.4 ms)，然而其参数量高达23.04 M、训练时间最长(255.56 min)，且在精度上仅与EfficientNet-B0相近，其规模优势并未能转化为性能优势。EfficientNet-B0在参数量(4.05 M)、训练时间(105.86 min)和推理时间(6.3 ms)之间取得一定平衡，但精度仍明显落后于MobileViT。相比之下，MobileViT在保持适中参数量(5.01 M)和训练时间(155.11 min)的情况下，取得了最优测试损失。虽其推理时间(8.4 ms)略高于其他模型，但相较于精度的显著提升，这一延迟完全可接受。由此可见，MobileViT在轻量化与高精度之间实现了最优平衡，体现出在不同湍流条件下对波前重构的强大适应性和复原能力。

图12和13是$D/r_0=15$和$D/r_0=20$随机抽取的一帧波前畸变及其复原结果，其中(a)原始波前；(a1)-(a4)

表2　各模型参数量、训练与推理时间统计

Table 2　Statistics of parameters, training time, and inference time for different models

Model	Parameters(M)	Training time/min	Inference time/ms
CNN	0.98	62.66	0.63
EfficientNet-B0	4.05	105.86	6.3
ResNet50	23.04	255.56	5.4
MobileViT	5.01	155.11	8.4

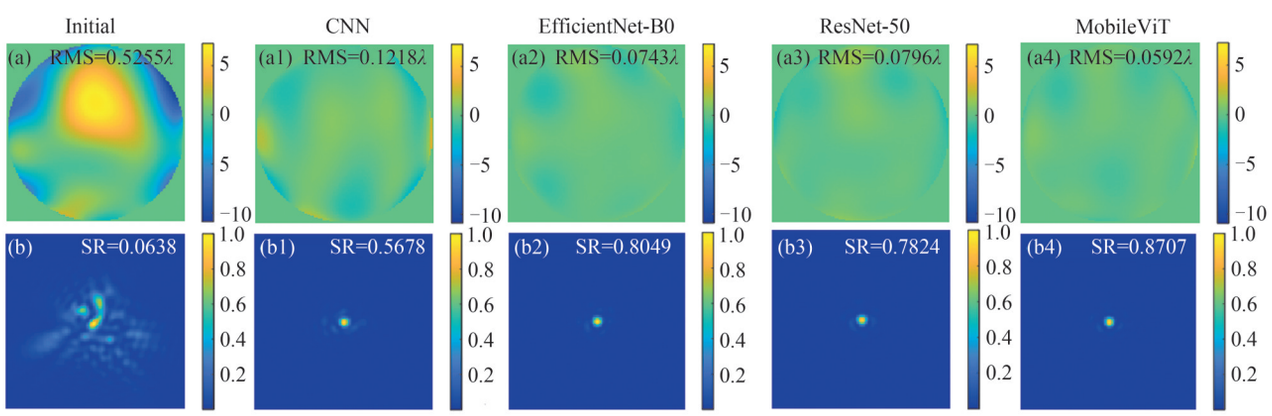


图12 $D/r_0=15$时随机抽取的一帧像差以及复原结果对比。(a)为原始波前；(a1)～(a4)分别为CNN，EfficientNet-B0，ResNet-50和MobileViT的复原后的残余波前；(b)为初始的PSF；(b1)～(b4)分别为各自模型对应的复原后的PSF

Fig. 12 Comparison of reconstructed results for a random wavefront aberration under $D/r_0=15$. (a) The original wavefront; (a1) ～ (a4) correspond to the residual wavefronts reconstructed by CNN, EfficientNet-B0, ResNet-50, and MobileViT, respectively. (b) Initial PSF; (b1)～(b4) represent the reconstructed PSFs corresponding to each model

分别为CNN,EfficientNet-B0,ResNet-50和MobileViT复原后的残余波前,(b)为初始的PSF,(b1)-(b4)分别为CNN,EfficientNet-B0,ResNet-50和MobileViT复原后的PSF。从图中可以看出,$D/r_0=15$和$D/r_0=20$时,MobileViT在强湍流情况下均保持最优的性能。在$D/r_0=15$时,其复原波前最接近真实波前,残余波前对应的PSF质量也最高,SR的值为0.870 7,高于ResNet50的0.782 4、EfficientNet-B0的0.804 9,和CNN的0.567 8。$D/r_0=20$时,优势更加明显,其SR的值0.802 7,显著优于ResNet-50的0.767 0,EfficientNet-B0的0.617 2和CNN的0.379 4。综上表明,在湍流越强的情况下,MobileViT相比于其他两种模型稳定性更好,复原精度最高。

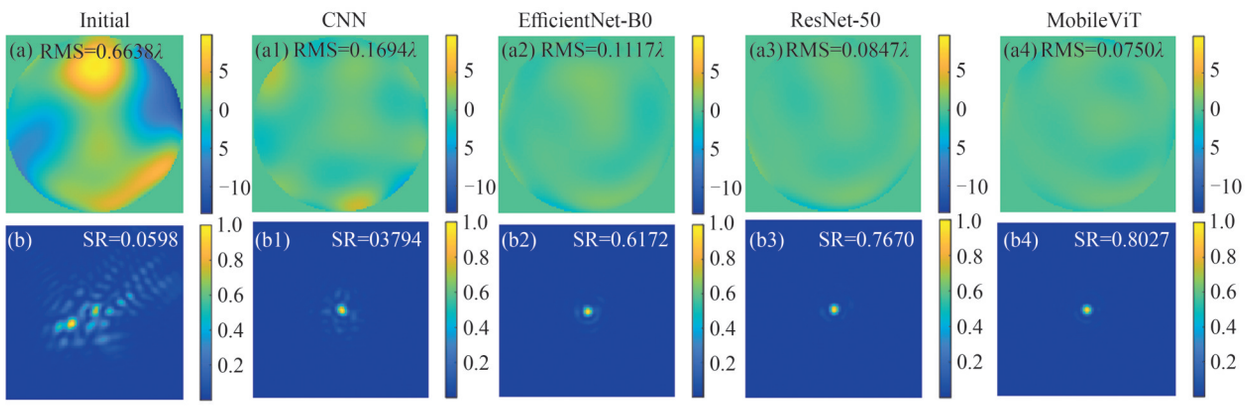


图13 $D/r_0=20$时随机抽取的一帧像差以及复原结果对比。(a)为原始波前;(a1)~(a4)分别为CNN,EfficientNet-B0, ResNet-50和MobileViT的复原后的残余波前;(b)为初始的PSF;(b1)~(b4)分别为各自模型对应的复原后的PSF

Fig. 13 Comparison of reconstructed results for a random wavefront aberration under $D/r_0=20$. (a) The original wavefront; (a1)~(a4) correspond to the residual wavefronts reconstructed by CNN, EfficientNet-B0, ResNet-50, and MobileViT, respectively. (b) The initial PSF; (b1)~(b4) represent the reconstructed PSFs corresponding to each model

4 结论

本文提出了一种基于MobileViT模型的相位差法点目标Zernike系数预测与波前复原方法,通过深度学习技术实现了高效、非迭代的波前校正。相较于传统的迭代优化算法,该方法利用焦平面和离焦平面两帧PSF图像直接预测第3~35阶Zernike系数。以不同湍流条件下的波前像差作为复原对象,分析基于MobileViT的波前重构性能。为便于对比,同时给出了CNN、ResNet-50和EfficientNet-B0模型的复原结果。

与CNN、ResNet-50和EfficientNet-B0对比,MobileViT在不同湍流强度下均表现出优越性能。结果表明,CNN模型虽然参数量小、推理时间快,但复原效果明显最差。弱湍流条件下,CNN、ResNet-50、EfficientNet-B0和MobileViT四种模型均能较好地复原波前。但随着湍流强度增加,CNN的复原精度低于其它三种模型,EfficientNet-B0的复原精度略低于ResNet-50,而MobileViT模型始终保持最高的复原精度,其残余波前RMS最小,复原的SR最大,显示出在强湍流条件下对波前畸变具有更强的鲁棒性。同时,通过对比Zernike系数的预测值与真实值的MAE,可以看出MobileViT在不同湍流下受影响最小,精确性和稳定性均优于其它模型。结合参数量、训练时间和推理速度的对比,MobileViT在轻量化与高精度之间实现了最佳平衡。

该方法验证了深度学习结合相位差法在无波前探测自适应光学系统中的应用潜力,不仅提升了波前校正的精度和速度,还降低了硬件依赖,为天文观测、激光通信等领域的实际应用提供了新思路。下一步拟搭建实验平台进一步验证本文所提方法的先进性。

参考文献

[1] JIANG Wenhan. Overview of adaptive optics development[J]. Opto-Electronic Engineering, 2018, 45(3): 170489.
姜文汉. 自适应光学发展综述[J]. 光电工程, 2018, 45(3): 170489.

[2] ANGEL J R P, WIZINOWICH P, LLOYD-HART M, et al. Adaptive optics for array telescopes using neural-network techniques[J]. Nature, 1990, 348(6298): 221-224.

[3] NEMOTO K, FUJII T, GOTO N, et al. Transformation of a laser beam intensity profile by a deformable mirror[J]. Optics Letters, 1996, 21(3): 168-170.

[4] JIANG Jisong, JIANG Aimin. Numerical model and verification of anisoplanatic imaging based on time-varying phase

screens[J]. Acta Photonica Sinica, 2020, 49(3): 0301001.
蒋佶松，姜爱民. 基于时变相位屏的非等晕成像数值模型及验证[J]. 光子学报，2020，49(3)：0301001.
[5] VORONTSOV M A, CARHART G W, COHEN M, et al. Adaptive optics based on analog parallel stochastic optimization: analysis and experimental demonstration[J]. Journal of the Optical Society of America A, 2000, 17(8): 1440-1453.
[6] GENG C, LUO W, TAN Y, et al. Experimental demonstration of using divergence cost-function in SPGD algorithm for coherent beam combining with tip/tilt control[J]. Optics Express, 2013, 21(21): 25045-25055.
[7] ZOMMER S, RIBAK E N, LIPSON S G, et al. Simulated annealing in ocular adaptive optics[J]. Optics Letters, 2006, 31(7): 939-941.
[8] YANG P, AO M W, LIU Y, et al. Intracavity transverse modes controlled by a genetic algorithm based on Zernike mode coefficients[J]. Optics Express, 2007, 15(25): 17051-17062.
[9] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
[10] CAO Yang, ZHANG Zupeng, PENG Xiaofeng. Wavefront distortion restoration method based on residual attention network[J]. Acta Photonica Sinica, 2022, 51(12): 1206002.
曹阳，张祖鹏，彭小峰. 基于残差注意力网络的波前畸变复原方法[J]. 光子学报，2022，51(12)：1206002.
[11] PAINE S W, FIENUP J R. Machine learning for improved image-based wavefront sensing[J]. Optics Letters, 2018, 43(6): 1235-1238.
[12] LI Y, YUE D, HE Y. Prediction of wavefront distortion for wavefront sensorless adaptive optics based on deep learning [J]. Applied Optics, 2022, 61(14): 4168-4176.
[13] SIDDIK A B, SANDOVAL S, VOELZ D, et al. Deep learning estimation of modified zernike coefficients and recovery of point spread functions in turbulence[J]. Optics Express, 2023, 31(14): 22903-22913.
[14] GONSALVES R A, CHIDLAW R. Wavefront sensing by phase retrieval[C]. Applications of Digital Image Processing III, SPIE, 1979, 207: 32-39.
[15] MA Huimin, JIAO Jun, QIAO Yan, et al. Wavefront restoration method based on light intensity image deep learning[J]. Laser & Optoelectronics Progress, 2020, 57(8): 081103.
马慧敏，焦俊，乔焰，等. 一种基于光强图像深度学习的波前复原方法[J]. 激光与光电子学进展，2020，57(8)：081103.
[16] MA H, ZHANG W, NING X, et al. Turbulence aberration restoration based on light intensity image using GoogLeNet [C]. Photonics. MDPI, 2023, 10(3): 265.
[17] WU Yu, GUO Youming, BAO Hua, et al. Sub-millisecond phase retrieval for phase-diversity wavefront sensor[J]. Sensors, 2020, 20(17): 4877.
[18] KOK Y E, BENTLEY A, PARKES A, et al. Direct Zernike coefficient prediction from point spread functions and extended images using deep learning[J/OL]. arXiv: 2404.15231, 2024. (2024-04-23) [2024-12-21]. https://doi.org/10.48550/arXiv.2404.15231.
[19] XU Yuanzhai, TANG Qiuyan, WANG Xiaojun, et al. Double spots based deep learning wavefront reconstruction and correction[J]. Chinese Journal of Lasers, 2024, 51(13): 1301005.
许元斋，唐秋艳，王小军，等. 基于双光斑的深度学习波前复原及校正[J]. 中国激光，2024，51(13)：1301005.
[20] DING Y, YI Z, LI M, et al. HI-MViT: A lightweight model for explainable skin disease classification based on modified MobileViT[J]. Digital Health, 2023, 9: 20552076231207197.
[21] ZHANG M, LIN Z, TANG S, et al. Dual-attention-enhanced MobileViT network: a lightweight model for rice disease identification in field-captured images[J]. Agriculture, 2025, 15: 571.
[22] MEHTA S, RASTEGARI M. MobileViT: Light-weight, general-purpose, and mobile-friendly vision transformer[J/OL]. arXiv: 2110.02178, 2021. (2021-10-05) [2025-9-20]. https://doi.org/10.48550/arXiv.2110.02178.
[23] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4510-4520.
[24] CHEN Y. Convolutional neural network for sentence classification[D]. University of Waterloo, 2015.
[25] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
[26] NOLL R J. Zernike polynomials and atmospheric turbulence[J]. Journal of the Optical Society of America, 1976, 66(3): 207-211.
[27] ZHANG Xingyun, LUO Fanglin, LI Nan, et al. Phase diversity wavefront sensing and image reconstruction[J]. High Power Laser and Particle Beams, 2021, 33(8): 081010.
张杏云，罗芳琳，李楠，等. 相位差波前探测与图像重建[J]. 强激光与粒子束，2021，33(8)：081010.
[28] RODDIER N A. Atmospheric wavefront simulation using Zernike polynomials[J]. Optical Engineering, 1990, 29(10): 1174-1180.

Wavefront Reconstruction Using Phase Diversity Based on the MobileViT Model

WU Jiawei[1], MIAO Yongqiang[1], ZHANG Zhiguang[1], YANG Huizhen[2]
(1 *Jiangsu Ocean University, School of Electronic Engineering, Lianyungang* 222005, *China*)
(2 *Engineering School of Networks & Telecommunications, Jinling Institute of Technology, Nanjing* 211169, *China*)

Abstract: In adaptive Optics (AO) systems, wavefront aberrations are commonly represented by Zernike polynomials, whose coefficients describe distortions such as tilt, defocus, coma, and spherical aberration. Accurate estimation of Zernike coefficients is critical for high-quality imaging and effective wavefront correction. Due to the symmetry of the Fourier transform, Zernike coefficients with opposite signs can produce identical Point Spread Function (PSF) images, causing ambiguity in coefficient prediction. The Phase Diversity (PD) method mitigates this issue by simultaneously capturing focal and defocused intensity images, introducing additional phase information to improve reconstruction accuracy.

Based on PD theory, this study proposes a MobileViT-based wavefront reconstruction method. The approach uses a pair of focal and defocused intensity images as input and automatically extracts image features to achieve non-iterative phase retrieval. The reconstruction performance of MobileViT is systematically evaluated under varying turbulence intensities and compared with CNN, ResNet-50, and EfficientNet-B0 models, providing a reference for practical applications.

Experimental results demonstrate that MobileViT consistently outperforms the other models across all turbulence conditions. Although CNN has the fewest parameters and fastest inference, its reconstruction accuracy is the lowest. Under weak turbulence, all four models achieve reasonable wavefront reconstruction; as turbulence intensity increases, CNN′s accuracy deteriorates, EfficientNet-B0 performs slightly worse than ResNet-50, and MobileViT maintains the highest accuracy, with the smallest residual RMS and largest Strehl Ratio (SR), indicating stronger robustness under strong turbulence. Mean Absolute Error (MAE) comparisons of predicted versus true Zernike coefficients further show that MobileViT is least affected by turbulence, achieving superior precision and stability.

Considering reconstruction accuracy, model size, training time, and inference speed, MobileViT achieves an optimal balance between lightweight design and high precision. This method demonstrates strong robustness under varying turbulence conditions, improving wavefront correction efficiency while reducing hardware dependency. These findings highlight the potential of combining deep learning with phase diversity for practical sensorless AO systems and suggest promising applications in fields such as astronomical imaging and laser communications. Future work will involve building an experimental platform to further validate the proposed approach.

Key words: Adaptive optics; MobileViT model; Phase diversity; Zernike coefficients; Wavefront reconstruction

OCIS Codes: 110.1080; 010.1330; 080.1010

CSTR: 32255.14.gzxb20265501.0101001

Foundation item: National Natural Science Foundation of China (12473081, U2141255), Fund of the State Key Laboratory of Adaptive Optics (FNLAO-24-MS-001)