

文章编号: 2095-2163(2024)01-0143-04

中图分类号: TP751; TP183

文献标志码: A

一种结合 Transformer 和 CNN 的散斑图复原网络

沈丽华, 戚伯特, 杨旭, 陈瑞品

(浙江理工大学 理学院, 杭州 310018)

摘要: 近年来,深度学习已成功应用于计算成像领域,并取得了显著的成果。目前,实现激光透过散射介质后的图像复原技术成为热门的研究课题,本文提出了一种包含双编码器的深度学习网络模型 Trans_CNN,从散斑图复原目标物图像,该网络融合了来自 Transformer 编码器和卷积神经网络编码器的信息,并将融合的编码信息传递给解码器获得复原结果,从散斑图中最大程度地学习全局特征和局部特征,从而更好地从散斑图复原出目标图像。实验结果证明 Trans_CNN 网络在复原散斑图像方面具有更好的性能和复原图像质量较好。

关键词: 计算成像; 图像复原; Transformer; 卷积神经网络

Trans_CNN: A network for speckle pattern restoration combining Transformer and CNN

SHEN Lihua, QI Bote, YANG Xu, CHEN Ruipin

(School of Science, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: In recent years, deep learning has been successfully applied to computational imaging, and has achieved remarkable results. At present, image restoration after passing through scattering media has become a hot research topic. In this paper, a deep learning network model Trans_CNN with dual encoders is proposed to reconstruct the target images from the speckle patterns. The network combines the information from the Transformer encoder and the CNN encoder, and transmits the encoded information to the decoder to obtain the restoration result. The purpose is to obtain the context information from the speckle pattern by combining the global representation with the local representation, improve the reconstruction of the target image from the speckle pattern. The experimental results confirm that Trans_CNN network has better performance for the image restoration from speckle.

Key words: computational imaging; image restoration; transformer; convolutional neural network

0 引言

光透过不均匀散射介质时,无法清晰成像,例如在浑浊水下、雨天、雾天无法看清目标物,因此透过散射介质实现清晰成像是需要解决的问题^[1]。近十几年来,国内外的研究者相继提出了实现透过散射介质进行图像复原的方法。传统的实现散射图像复原技术主要基于物理方法实现,包括波前整形技术、光学相位共轭、光学传输矩阵等技术,但是这些技术存在计算量大、移值性差、恢复性差等缺点。为了克服这些缺点,引入深度学习方法来解决散斑图像复原问题。

随着计算机技术的快速发展,深度学习依托其

强大的数据挖掘能力在计算机视觉、自动驾驶汽车和自然语言处理等领域有广泛的应用前景,特别是在图像分割领域取得了巨大进展^[2]。散斑图像复原问题与图像分割问题相似,神经网络通过对散斑图像每个像素点赋予类别从而复原散斑图。2015年,Ando等^[3]通过使用支持向量机(SVM)成功对收集到的包含人脸和非人脸的散斑图进行训练和分类;2018年,Li等^[4]提出 IDiffNet 网络,实现从散斑图复原出原始图像,并引入负皮尔逊相关系数(NPCC)作为损失函数,实验证明该损失函数对于稀疏图像复原效果更好;2021年,Lai等^[5]提出生成式对抗网络,通过强散射介质实现对位于相邻层中的两个不同对象的同时成像。

基金项目: 浙江省重点研发项目(2022C04007); 国家自然科学基金面上项目(11874323)。

作者简介: 沈丽华(1997-),女,硕士研究生,主要研究方向:深度学习、计算成像;戚伯特(1998-),男,硕士研究生,主要研究方向:计算成像;

杨旭(1988-),男,博士,副教授,主要研究方向:计算成像。

通讯作者: 陈瑞品(1970-),男,博士,教授,主要研究方向:计算成像、光场调控。Email:chenrp@zstu.edu.cn

收稿日期: 2022-12-29

以往利用深度学习散射成像的方法主要依靠卷积神经网络(CNN)进行散斑图像复原。CNN性能卓越,具有十分出众的特征提取能力,但也存在一些局限性,如具有局部感受野,因而很难感知全局信息。Transformer模型通过自注意力机制能捕获全局信息,能弥补卷积神经网络对全局信息感知不足的缺点。2020年,Dosovitskiy等^[6]提出了Vision Transformer(ViT)模型,在Transformer基础上进行改进,将图像分割成多个切片,并将这些切片作为线性嵌入序列输入到Transformer中训练,捕获输入图像中的长距离相关性,从而扩大图像感受野,获得更多的全局上下文信息。目前ViT已经在图像分类、语义分割等领域取得了较好的效果,许多工作提出了将Transformer和CNN相结合构成网络模型,使网络模型能保留全局和局部特征信息,一些领域取得了不错的成就^[7]。

本文基于Transformer和CNN的优点,将Transformer与CNN相结合用于实现散斑图复原,设计了一种深度学习网络模型Trans_CNN,该模型包含双编码器,即基于Transformer的编码器和基于CNN的编码器,将CNN提取到的局部特征和Transformer提取到的全局特征相结合,将成对的散斑图和目标图放入网络中训练学习,实现从散斑图复原出高质量图像。实验结果表明,Trans_CNN网

络模型能有效地复原散斑图像,复原效果较好。

1 Trans_CNN 网络模型

以往使用深度学习网络模型实现散射成像的方法主要依靠CNN实现,存在一定局限性,很难感知全局信息。Transformer在处理全局信息交互方面具有出色的性能。本文基于Transformer和CNN的特点,将Transformer和CNN相结合,构成新的网络模型Trans_CNN用于实现散斑图复原,搭建的Trans_CNN网络模型如图1所示,网络的输入为采集到的散斑图,输入图片大小为 256×256 。散斑图输入到网络后,分别进入到两个编码路径中提取信息,分别为CNN编码路径和Transformer编码路径,大小为 $x \in R^{H \times W \times C}$, H 、 W 和 C 分别表示图片长、宽和通道数。在CNN网络编码路径中,通过卷积和下采样操作编码信息,该路径由CNN网络的编码器组成,包含4个子模块,每个子模块包含2个 3×3 卷积、Relu激活函数和 2×2 的最大池化层。在Transformer编码路径中,将输入大小为 $x \in R^{H \times W \times C}$ 的散斑图像分成维度为 $P \times P$ 的 N 个patch,即 $x_p \in R^{(N \times (P^2 \cdot C))}$ 。为了保持位置信息,将这些patch投影到一个 K 维的嵌入空间中,并添加一个用于patch嵌入的位置嵌入 W_{pos} 以保留每个patch的空间信息,patch嵌入的位置,式(1):

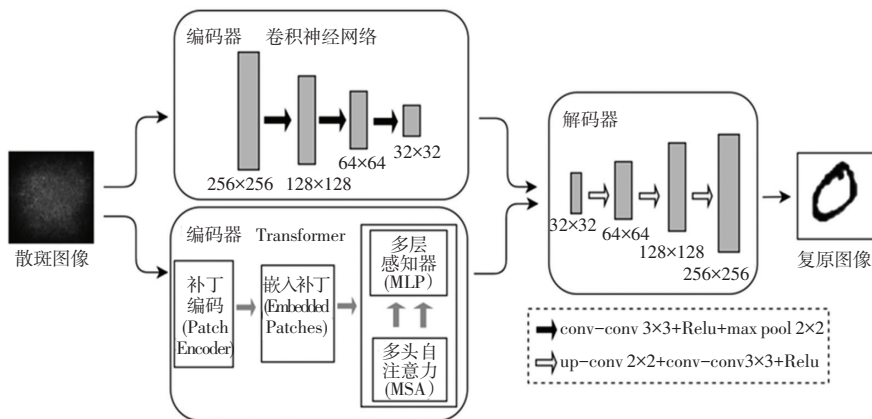


图1 Trans_CNN网络模型

Fig. 1 The model of Trans_CNN

$$t_0 = [x_p^1 W; x_p^2 W; \dots; x_p^N W] + W_{pos} \quad (1)$$

其中, $W \in R^{(N \times (P^2 \cdot C))}$ 表示patch的投影嵌入。

使用包含多头自注意(MSA)和多层感知器(MLP)的Transformer模块来学习全局上下文表示。MSA层由 M 个平行的自注意头部组成,以缩放嵌入的patch,MLP模块则学习长距离上下文依赖性。MSA和MLP的公式表示为式(2)和式(3), t'_i 和 t_i

为Transformer编码第 i 层的MSA和MLP的输出表示。

$$t'_i = \text{MSA}(\text{Norm}(t_{i-1})) + t_{i-1} \quad (2)$$

$$t_i = \text{MLP}(\text{Norm}(t'_i)) + t'_i \quad (3)$$

其中, $\text{Norm}()$ 表示层归一化。

两个编码路径后,CNN的编码信息与来自Transformer编码后的信息通过加法路径相加融合进

入解码器。解码器由 4 个子模块组成, 每个子模块包含 2×2 反卷积、2 个 3×3 卷积和 Relu 激活函数, 进行特征上采样, 将特征解码为高分辨率图像。

在网络训练过程中, 损失函数的变化情况可以反映目标图像的复原效果。本文选用负皮尔逊相关系数 (NPCC) 作为损失函数来观察散斑图像复原效果, 式(4):

$$Loss_{NPCC} = 1 - \frac{1 \times \sum_{i=1}^M \sum_{j=1}^N (M(i, j) - \bar{M})(N(i, j) - \bar{N})}{\sqrt{\sum_{i=1}^M \sum_{j=1}^N (M(i, j) - \bar{M})^2} \sqrt{\sum_{i=1}^M \sum_{j=1}^N (N(i, j) - \bar{N})^2}} \quad (4)$$

其中, $M(i, j)$ 表示网络复原的图像; $N(i, j)$ 表示原始目标; \bar{M} 和 \bar{N} 表示网络复原的图像和原始目标的均值。

2 实验结果及分析

2.1 数据集采集与预处理

本实验选取了 10 000 张各不相同的 MNIST 手写数字图用于采集散斑图。将 MNIST 手写数据图上采样到 $1\ 024 \times 1\ 024$ 后显示在空间光调制器 (SLM) 上, 由激光器先后通过两个透镜照射到 SLM 上, 携带了物体信息的 SLM 出射光通过散射介质并由 CMOS 相机采集散斑图像。实验中 SLM 和 CMOS 相机被连接到同一台电脑上, 并通过 Matlab 程序同时控制连续拍照收集散斑图像。

收集的 10 000 组数据集, 其中 8 000 组作为训练集, 1 000 组作为验证集, 1 000 组作为测试集。采集到的散斑图大小为 $2\ 048 \times 1\ 536$, 截取 $1\ 600 \times 1\ 000$ 大小中心方形区域以保留有用的散斑信息, 并更改大小为 256×256 作为网络模型的输入图像。原始的数字图像大小为 28×28 , 将其上采样到 256×256 作为数据集的标签图。

2.2 实验结果

在深度学习框架 PyTorch 基础上进行模型的搭建和训练, 在 RTX 3090Ti 上完成实验模型的训练。将训练集送到 Trans_CNN 中训练 20 次, 网络训练后测试集上的复原图如图 2 所示, 可以看出复原出的图像与原始图像基本类似, 即图像复原质量较高。

将 Trans_CNN 中的 Transformer 编码路径替换成 CNN 编码路径构成新的网络模型 CNN_CNN, 并将同样的数据集放在该模型中训练, 并与 Trans_

CNN 训练结果进行对比, 对比结果如图 3 所示, 可以看出两种网络模型均能实现散斑图的复原, 但 Trans_CNN 网络训练后复原的图像质量更高, 图像细节部分恢复得更好。

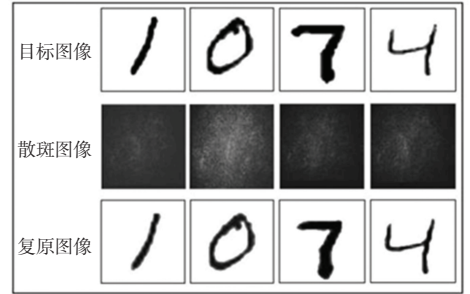


图 2 通过 Trans_CNN 训练后的复原图

Fig. 2 Recovered image after the training of Trans_CNN

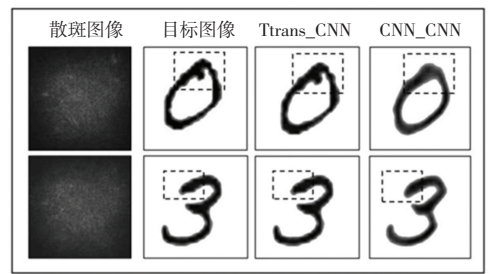


图 3 通过 Trans_CNN 和 CNN_CNN 网络复原的结果图对比

Fig. 3 Comparison of the restored images with the Trans_CNN and the CNN_CNN

2.3 评估分析

网络训练期间和验证期间的损失曲线图如图 4 所示, 可见 Trans_CNN 网络在前 3 轮训练过程中损失值急速下降并趋于稳定, 而 CNN_CNN 通过前 10 轮训练损失值才下降到一定程度并趋于稳定。相比之下, 整个训练过程中 Trans_CNN 网络的训练集损失值和验证集损失值都更低, 且收敛速度更快, 说明 Trans_CNN 在散斑图像复原方面具有更大优势。

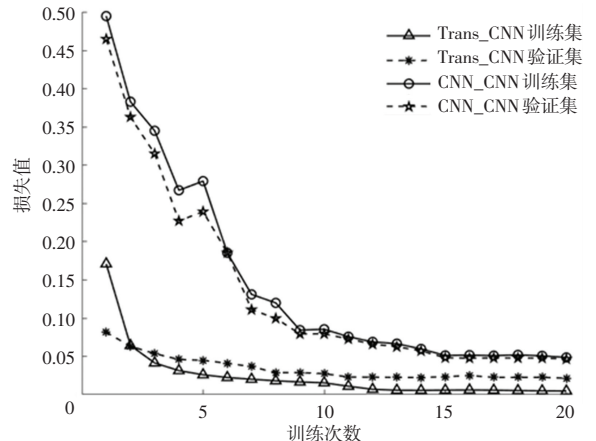


图 4 网络训练和验证期间的损失曲线图

Fig. 4 Loss curves during training and validation on the model

此外,为了评估比较两种网络对光散斑图像的复原效果,使用峰值信噪比(Peak Signal to Noise Ratio, *PSNR*)和结构相似性(Structural Similarity, *SSIM*)对网络复原效果进行量化。

*PSNR*的表达式(5):

$$PSNR = 10 \times \lg\left(\frac{255^2}{\frac{1}{WH} \sum_{i=1}^W \sum_{j=1}^H (M(i,j) - N(i,j))^2}\right) \quad (5)$$

其中, W 和 H 分别表示网络输出图像的宽度和高度; M 是原始目标图像; N 是网络重建的图像。

*SSIM*的表达式(6):

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (6)$$

其中, μ_x 、 μ_y 分别表示图像 x 和 y 的平均像素强度; σ_x 、 σ_y 分别表示图像 x 和 y 的标准差; σ_{xy} 是图像 x 和 y 的协方差; C_1 和 C_2 为常数。

分别计算出两种网络的测试集复原图像与目标图案平均*PSNR*和平均*SSIM*,结果见表1。可以看出,Trans_CNN网络复原的图像*PSNR*和*SSIM*均高于CNN_CNN,表明Trans_CNN对于散斑图的图像复原效果要优于CNN_CNN。

表1 两种网络复原图像的峰值信噪比(*PSNR*)和结构相似性(*SSIM*)的对比

Table 1 Comparison of *PSNR* and *SSIM* of restored images by two networks

网络	峰值信噪比(<i>PSNR</i>)	结构相似性(<i>SSIM</i>)
Trans_CNN	29.83	0.71
CNN_CNN	22.25	0.57

3 结束语

CNN存在局部感受野对全局信息感知不足,而

Transformer能弥补这种不足,本文提出了一种深度网络模型Trans_CNN来复原散斑图像。将Transformer和CNN两个编码器相结合构成双编码器;将两条编码路径提取到的信息通过加法路径相加融合,并将融合后的信息通过解码路径复原出目标图像。实验结果表明,Trans_CNN网络模型能结合Transformer和CNN的优点,最大程度地学习图像的全局特征和局部特征,恢复出的图像质量更高,对于图像细节信息的恢复能力更好。

参考文献

- [1] YOON S, KIM M, JANG M, et al. Deep optical imaging within complex scattering media[J]. Nature Reviews Physics, 2020, 2(3): 141-158.
- [2] MINAEE S, BOYKOV Y Y, PORIKLI F, et al. Image segmentation using deep learning: A survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(7): 3523-3542.
- [3] ANDO T, HORISAKI R, TANIDA J. Speckle-learning-based object recognition through scattering media[J]. Optics Express, 2015, 23(26): 33902-33910.
- [4] LI S, DENG M, LEE J. Imaging through glass diffusers using densely connected convolutional networks[J]. Optica, 2018, 5(7): 803-813.
- [5] LAI X, LI Q, CHEN Z, et al. Reconstructing images of two adjacent objects passing through scattering media via deep learning[J]. Optics Express, 2021, 29(26): 43280-43291.
- [6] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv: 2010.11929, 2020.
- [7] LIU W, LI C, XU N, et al. CVM-Cervix: A hybrid cervical pap-smear image classification framework using CNN, visual transformer and multilayer perceptron[J]. Pattern Recognition, 2022, 130: 108829.

(上接第142页)

情况的变动,对身份认证模型进行定期的调整,更好地完成多特征融合身份标识认证模型的构建。实验证明,该模型认证识别率可以达到90%以上,身份认证准确度较高。

参考文献

- [1] 杨丽红,尚泽昊. 基于区块链和模糊提取的多特征融合身份认

证模型[J]. 数字技术与应用, 2022, 40(8): 218-220.

- [2] 吕建国,王伟,黄笑冉. 基于区块链的走失者身份认证模型的研究[J]. 电脑知识与技术, 2021, 17(25): 149-150, 168.
- [3] 韦书令. 融合区块链技术的社区教育学习成果认证模型及实现机理研究[J]. 成人教育, 2021, 41(8): 71-77.
- [4] 李增,徐济成. 一种去中心化的身份认证模型的设计与实现[J]. 辽东学院学报(自然科学版), 2020, 27(1): 62-72.