

文章编号 1004-924X(2025)15-2455-13

基于 Swin Transformer 的位移与应变测量

但西佐, 杨 浩, 李桂华*

(安徽大学 电气科学与自动化学院, 安徽 合肥 230601)

摘要:为解决传统数字图像相关(DIC)方法依赖固定子集而导致的高频位移捕捉能力不足及对复杂应变适应性差的问题,提出了一种基于 Swin Transformer 的位移与应变同步预测网络模型。该模型由 ST-DIC-d(位移预测)和 ST-DIC-s(应变预测)两个子网络组成。首先,利用 Swin Transformer 的分层编码器-解码器结构进行多尺度特征提取;然后,通过注意力机制与跳跃连接有效捕捉图像的局部和全局特征,从而得到位移场;最后,通过应变计算层从位移场精确计算应变分量。在 DIC Challenge 数据集上的实验结果表明,ST-DIC 相比传统 DIC 方法和基于 CNN 的模型,平均绝对误差在垂直位移预测上减少 21%。在实际拉伸实验中,ST-DIC 在高梯度区域最大应变精度提高 8%。本文所提出的方法在测量复杂变形方面具有更高的可靠性与精度。

关键词:数字图像相关;Swin Transformer;高频位移;多尺度特征提取;应变测量

中图分类号:TP394.1;TH691.9 **文献标识码:**A

doi:10.37188/OPE.20253315.2455 **CSTR:**32169.14.OPE.20253315.2455

Displacement and strain measurement based on Swin Transformer

DAN Xizuo, YANG Hao, LI Guihua*

(School of Electrical Engineering and Automation, Anhui University, Hefei 230601, China)

* Corresponding author, E-mail: guihuali1@163.com

Abstract: To address the limited high-frequency displacement capture capability and poor adaptability of traditional digital image correlation (DIC) methods that rely on fixed subsets, a strain-displacement synchronous prediction network based on the Swin Transformer is proposed. The model comprises two sub-networks, ST-DIC-d (displacement prediction) and ST-DIC-s (strain prediction). A hierarchical encoder-decoder Swin Transformer is employed for multi-scale feature extraction, enabling effective capture of local and global image features via attention mechanisms and skip connections to recover the displacement field. A trainable strain computation layer then derives strain components directly from the predicted displacement field. Evaluation on the DIC Challenge dataset shows that ST-DIC reduces the average absolute error in vertical displacement prediction by 21% relative to traditional DIC and CNN-based approaches. In tensile experiments, ST-DIC attains an 8% improvement in maximum strain accuracy within high-gradient regions and yields more finely resolved measurements. The proposed method demonstrates enhanced reliability and accuracy for measuring complex deformations.

收稿日期:2025-05-16;修订日期:2025-06-18.

基金项目:安徽省高校自然科学基金资助项目(No. 2022AH050098)

Key words: digital image correlation; swin transformer; high-frequency displacement; multi-scale feature extraction; strain measurement

1 引言

数字图像相关 (Digital Image Correlation, DIC) 是一种在 20 世纪 80 年代发展起来的非接触式全场光学测量技术, 可精确测量目标表面的位移和应变。随着计算机视觉与图像处理技术的不断进步, DIC 已成为不可或缺的测量手段, 被广泛应用于材料科学、土木工程和生物力学等领域^[1-2]。其核心原理是对机械载荷前后的物体表面散斑图像进行拍摄, 并利用图像相关算法计算位移场与应变场^[3]。基于其全场测量且非接触的特点, DIC 在传统方法难以胜任的场合表现出显著优势。由于具备更强的计算能力和更优的图像处理算法, DIC 的准确度和可靠性大幅提高, 已成为工程和科学研究中的标准测量方法^[4-8]。

DIC 最初主要应用于测量平面内的位移^[9]。但随着技术的成熟, 它逐渐扩展到三维变形测量, 可在复杂环境或特殊场景下全面监测结构变形^[1,10-11]。传统的 DIC 方法通过将图像划分为子集来进行相关计算, 而子集大小直接影响到位移和应变测量的准确性。较小的子集虽然对细微的局部变形更敏感, 从而提高了空间分辨率, 但在噪声较大或散斑稀疏时易出现匹配不稳定的情况。相反, 较大子集可降低噪声影响, 提高匹配稳定性, 但会降低空间分辨率, 可能忽略局部变形的关键细节, 且难以准确捕捉高空间频率的位移场或应变场^[5,12-13]。当结构变形量较大或散斑图案破碎时, 传统 DIC 的预测常变得不稳定且易产生误差, 其根本原因在于 DIC 依赖子集进行相关计算, 而在大变形下的匹配难度会显著增大, 导致预测精度下降^[14]。

为了克服这些局限性, 研究人员开始关注深度学习方法在 DIC 中的应用, 尤其是卷积神经网络 (Convolutional Neural Network, CNN)。深度学习模型具有强大的特征提取和模式识别能力, 可直接基于端到端框架从图像中提取位移和应变信息, 减少了传统方法中图像预处理和特征匹

配的需求。Min 等提出了一种基于 3D CNN 的应变测量模型, 通过卷积、池化和全连接层来预测拉伸测试过程中的 x 与 y 方向位移。然而, 该方法不仅依赖于传统 DIC 生成训练数据, 其精度还受限于所使用的 DIC 方法^[15]。Boukhtache 等提出了 StrainNet 模型, 可直接从变形散斑图像中提取位移场。该网络基于 FlowNetS 架构, 由特征提取模块和位移预测模块组成, 先利用卷积层提取图像特征, 再采取上采样过程生成全分辨率光流, 通过加深网络结构与优化架构, 该模型提升了光流预测的准确度^[16-17]。

此类模型的输出通常为像素级位移, 需进一步基于位移计算应变。Yang 等提出了一个由两个 CNN 网络构成的可同时计算位移和应变场的端到端模型。这两个网络分别采用下采样和上采样结构处理一对散斑图像, 并输出相应的位移和应变。由于其训练数据主要使用均匀散斑和低空间频率位移, 该模型的泛化能力受到限制, 对于高空间频率位移和应变的预测并未超越传统 DIC^[14]。Wang 等提出 DIC-Net 模型, 结合改进的 U-Net 结构、残差模块和相对位置提取器 (Relative Position Extractor, RPE) 来精确预测复杂变形, 提高了位移场与应变场的测量精度并减少了“棋盘效应”, 所使用的 Hermite 数据集包含不同空间频率和散斑模式下的大量位移场, 但位移范围仅限于 2 像素以内, 模型在大位移场景下的测量能力尚未得到验证^[18]。此外, Wang 等还提出了基于无监督 CNN 的 DIC 方法, 可准确测量二维位移。其模型依旧采用包含多层特征提取、双通道相关块和注意力机制的 U-Net 架构, 从具有不同特征的散斑图像中提取信息。在训练阶段, 通过变形模型预测参考散斑图并与真实参考图像进行对比, 实现无监督训练。尽管该方法在包含大、小位移的数据上进行了训练和验证, 但其只能输出位移值, 且整体精度并未显示出显著提升^[19]。

尽管基于深度学习的 DIC 方法取得了一些不错的成果, 但这些方法大多数仍以 CNN 为核

心。然而,由于卷积运算的局部性,CNN难以捕捉远距离交互与全局语义信息^[20]。并且,CNN的结构设计本身强调特征的平移不变性,而不是对位置敏感,这在图像分类中是优势,但在DIC中,尤其是测量高频位移时,要求网络必须对微小的空间位移保持高度敏感。这些局限性影响了高频位移的测量精度。近年来出现的Swin Transformer^[21]作为一种新颖的架构,其特有的层次化结构与窗口注意力机制能同时捕捉局部与全局信息,克服了CNN的局限并在图像处理等任务中展现出优异性能,广泛应用于计算机视觉领域。Zhou等将类似GMFlow的Transformer模型引入深度学习DIC框架,虽然只使用相对较小的训练数据,但在位移测量上也取得了较好的效果^[22]。Cao等的研究^[23-25]也表明,基于Swin Transformer的U-Net模型在医学图像分割、遥感和图像修复等领域具有出色表现。

因此,本文提出了一种改进的U-Net结构的Swin Transformer模型ST-DIC。与CNN模型相比,Swin Transformer中的自注意力机制可有效捕捉图像全局相关性,使模型在识别与跟踪高频和微小位移方面更具优势。而添加了跳跃连接的U-Net结构可进行多尺度特征提取,既能处理大尺度变形,又能保留细节信息,从而精准测量高频变形。与之前提出的模型^[26]相比,ST-DIC应变网络在保留高可靠性的同时,参数大幅减少,提高了训练和推理效率。实验证明,在DIC Challenge^[27]数据集上,与CNN模型和商业DIC软件相比,ST-DIC尤其在捕捉高频位移和应变方面表现更优,准确度和稳定性均显著提升。此外,该模型在实际拉伸实验中也获得了可靠的结果。

2 方 法

如图1所示,本文提出的模型包括编码器、解码器以及跳跃连接。模型由两个网络组成:一个用于计算位移场(ST-DIC-d),另一个用于计算应变场(ST-DIC-s)。其基本单元为Swin Transformer Block。

在编码器部分,输入的参考图像与变形图

像首先进行叠加(stack)并被分割为不重叠的 4×4 patch。首先,每个patch包含两幅 4×4 的单通道图像(共2通道),因此其初始特征维度为 $4 \times 4 \times 2 = 32$ 。随后,通过线性嵌入层将这些patch特征映射到给定的高维空间,记为 C 。接着,对嵌入后的patch token依次进行多层Swin Transformer模块与patch merging模块,逐渐获得分层的特征表示。patch merging模块用于下采样并增大通道维度,而Swin Transformer模块主要用于捕捉图像间的位移关联。模型的bottleneck层位于编码器与解码器之间,用于进一步压缩特征表示,并提取高层次抽象特征,在降低计算量的同时保留核心信息。

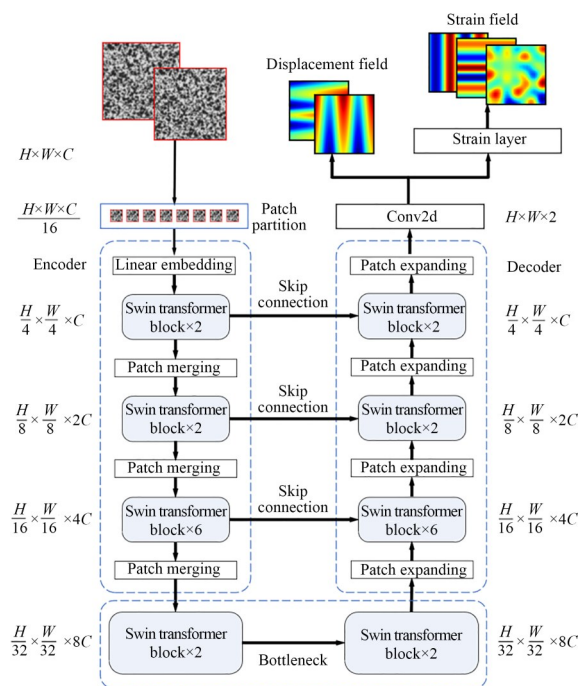


图1 ST-DIC网络结构

Fig. 1 Model structure of ST-DIC net

在解码器中,通过Swin Transformer模块与patch expanding层对特征图进行上采样,并利用跳跃连接将不同尺度的编码器特征融合到解码器对应层,以弥补编码过程中空间信息的损失。最后,通过patch expanding层逐步恢复分辨率,输出高精度位移场。

应变网络相比位移网络增加了一个应变层,通过使用Savitzky-Golay滤波器实现了位移场对

空间坐标的数值微分,从而得到应变分量 ϵ_x, ϵ_y 以及剪切应变 ϵ_{xy} 。各模块功能如下所示。

2.1 输入层、分割与线性嵌入

输入层接收两张 128×128 的单通道灰度图,即参考图像与变形图像,并在通道维度进行堆叠,形成一个双通道的输入张量。接着, Patch Partition层将输入张量分割为 4×4 固定尺寸的patch,并使用一个步幅和卷积核大小均为4的卷积操作,将每个patch转化到高维嵌入向量空间,使得二维图像数据被转化为一维序列表示。

2.2 Transformer 模块

Swin Transformer模块旨在保持较低计算量的同时高效捕捉图像间的位移关系。该模块首先在局部窗口内使用多头自注意力机制,通过限定注意力计算范围来降低计算开销。多头自注意力机制将输入特征映射到 (Q, K, V) 向量上,并通过以下公式计算注意力权重并得到加权输出:

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}} + B\right)V, \quad (1)$$

其中: d_k 表示键向量的维度, B 为每个注意力头的相对位置偏置项。

如图2所示,左侧模块中的W-MSA表示窗口多头自注意力(Shifted Window Multi-head Self-Attention, W-MSA)机制。输入首先经过层

归一化进行标准化处理,随后进入W-MSA模块,在局部窗口内进行独立的注意力计算。W-MSA输出再次经过层归一化后,被送入多层感知器(Multi-Layer Perceptron, MLP)以实现非线性变换和通道维度的信息建模。在W-MSA和MLP模块之后均引入残差连接,以保留原始特征并增强网络的稳定性。右侧模块中的SW-MSA与W-MSA类似,但它使用了窗口平移策略,可通过跨窗口的信息交互提升局部特征关联。

2.3 Patch Merging与Patch Expanding

网络的下采样阶段采用Patch Merging层,将相邻patch合并以降低特征图的分辨率并增加通道维度。这样不仅能减少空间大小和降低计算量,还能提升特征表示的抽象层次。Patch Merging通过线性变换扩大合并后的patch特征维度,从而丰富特征表达。而在上采样阶段,Patch Expanding层执行与之相反的操作,以逐步恢复特征图的分辨率。

2.4 跳跃连接

为了在图像重建过程中充分利用多尺度信息,模型在不同下采样层间添加了跳跃连接,将编码器中的多尺度特征与解码器对应层相融合,以弥补下采样过程中丢失的空间信息,使得在融合全局特征的同时保留丰富的局部细节,最终通过一个卷积层输出高分辨率的位移场。

2.5 基于Savitzky-Golay滤波器的应变计算层

应变网络(ST-DIC-s)在位移网络(ST-DIC-d)的基础上增加了一个非可训练的应变计算层,用于从位移场精确计算应变分量。为提高偏导数计算的精度并抑制测量噪声的影响,选用Savitzky-Golay滤波器实现位移场对空间坐标的数值微分,从而获得应变分量 ϵ_x, ϵ_y 和剪切应变 ϵ_{xy} 。

Savitzky-Golay滤波是一种基于局部最小二乘多项式拟合的数据平滑与微分方法,可以有效降低数据中的随机噪声并提高导数计算的准确性。具体而言,首先,在给定尺寸的滑动窗口内采用最小二乘法对数据点进行多项式拟合,再通过对拟合多项式进行解析求导,得到对应的数值导数。使用该方法计算导数时需设定两个主要参数:滑动窗口长度(通常为奇数)与拟合多项式

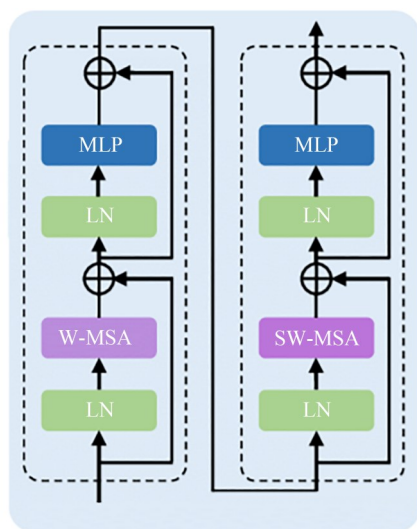


图2 Swin Transformer块

Fig. 2 Swin Transformer block

阶数,这两个参数的选择决定了滤波的平滑程度与导数估计的局部准确性。在充分参考 Pan^[28]等关于随机噪声抑制与空间分辨率折衷的讨论基础上,固定窗口长度为 5 像素(对应一维核长 5),多项式阶数为 2。选择依据:阶数 2 即可准确近似一阶导数且不会过拟合;长度 5 在本工作典型散斑尺寸(3~5 pixel)和噪声水平下兼顾平滑与细节保持。再对 3~11 像素窗口的敏感性分析,发现窗口 5~7 的应变均方误差差异不足 2%,因此采用计算量更低的窗口 5。

以位移网络 ST-DIC-d 输出的位移分量 $u(x, y), v(x, y)$ 为例,采用一维 Savitzky-Golay 滤波器分别计算各偏导数,即:

$$\begin{cases} \frac{\partial u}{\partial x} = u(x, y) * K_{SG, x} \\ \frac{\partial v}{\partial y} = v(x, y) * K_{SG, y} \\ \frac{\partial u}{\partial y} = u(x, y) * K_{SG, y} \\ \frac{\partial v}{\partial x} = v(x, y) * K_{SG, x} \end{cases}, \quad (2)$$

其中: $K_{SG, x}$ 与 $K_{SG, y}$ 分别表示沿 x 轴和 y 轴计算导数的 Savitzky-Golay 滤波器核,实际上二者为相同的滤波核沿不同方向的应用。利用上述获得的位移偏导数,可进一步计算各应变分量,法向应变分别为:

$$\epsilon_{xx} = \frac{\partial u}{\partial x}, \quad \epsilon_{yy} = \frac{\partial v}{\partial y}. \quad (3)$$

剪切应变为:

$$\epsilon_{xy} = \frac{1}{2} \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right). \quad (4)$$

$$L_{\text{strain}} = \frac{1}{N} \sum_{i=1}^N \left(\left\| \epsilon_{x, \text{pred}, i} - \epsilon_{x, \text{true}, i} \right\|_2^2 + \left\| \epsilon_{y, \text{pred}, i} - \epsilon_{y, \text{true}, i} \right\|_2^2 + \left\| \epsilon_{xy, \text{pred}, i} - \epsilon_{xy, \text{true}, i} \right\|_2^2 \right), \quad (6)$$

其中: $\epsilon_{x, \text{pred}}, \epsilon_{y, \text{pred}}, \epsilon_{xy, \text{pred}}$ 分别代表预测的 x 方向、 y 方向和剪切方向的应变值; $\epsilon_{x, \text{true}}, \epsilon_{y, \text{true}}$ 和 $\epsilon_{xy, \text{true}}$ 分别为对应的真实应变值; N 为图像中的像素总数。

如表 1 所示, ST-DIC 每个 patch 的推理时间约为 0.013~0.014 s, DICnet 大约为 0.010~0.011 s, StrainNet-f 为 0.008 251 s; 然而, ST-DIC 的单轮训练时间大幅缩短, 仅约 200~230 s (DICnet 需 495~520 s), 且显存占用减少一半 (ST-DIC: 7.3 GB, DICnet: 15.6 GB)。由此可

值得指出的是, 尽管该应变计算层中使用的 Savitzky-Golay 滤波器参数预先确定且本身不包含可训练参数, 但实验表明, 将 ST-DIC-s 与 ST-DIC-d 联合进行端到端训练, 即通过优化位移网络参数以提高应变预测精度, 能够显著提升最终应变计算的准确性。

3 实 验

实验与测试在配备 NVIDIA GeForce RTX 4080 GPU 和 Intel Core i9-13900K CPU 的计算机上进行, 使用 PyTorch 2.12 与 Python 3.9 编程, CUDA 版本为 11.8。本文使用的训练和验证数据集在文献[26]中有详细的阐述。

在超参数设置上, 初始学习率设为 10^{-4} , 采用 AdamW 优化器, 权重衰减值 10^{-5} , 以控制 L2 正则化以防止过拟合。批大小为 16, 训练总共进行 300 轮。为进一步降低过拟合风险, 采用学习率固定的策略, 使参数更新更稳定。损失函数选用均方误差 (Mean Square Error, MSE), 以加强模型在高变形梯度情况下的预测能力。

位移网络和应变网络的训练采用基于 L_2 范数的 MSE 损失函数, 以最小化预测位移场与真实位移场之间的差异。位移损失函数定义如下:

$$L_{\text{disp}} = \frac{1}{N} \sum_{i=1}^N \left(\left\| u_{\text{pred}, i} - u_{\text{true}, i} \right\|_2^2 + \left\| v_{\text{pred}, i} - v_{\text{true}, i} \right\|_2^2 \right), \quad (5)$$

其中: u_{pred} 和 u_{true} 分别为预测和真实的 x 方向位移, v_{pred} 和 v_{true} 分别为预测和真实的 y 方向位移, N 表示图像中的像素总数。应变损失函数表示为:

见, 在效率与资源占用方面, ST-DIC 优势明显。以 1024×1024 (约 1.05 MP) 分辨率的 DIC 图像为例, 需划分 64 个 128×128 patch。按照 0.014 s/patch 的平均推理时间, 整幅图推理耗时约 0.90 s (含 patch 组装及两次 I/O), 基本满足工业场景的应用需求。这里需要说明的是, 对比验证时优化了 D-ST 与 S-ST 的数据加载代码, 使之与本文的模型处于相同条件, 因此推理时间与文献[26]不同。

表 1 不同深度学习 DIC 模型的训练和推理效率比较

Tab. 1 Comparison of training and inference efficiency of different deep learning DIC models

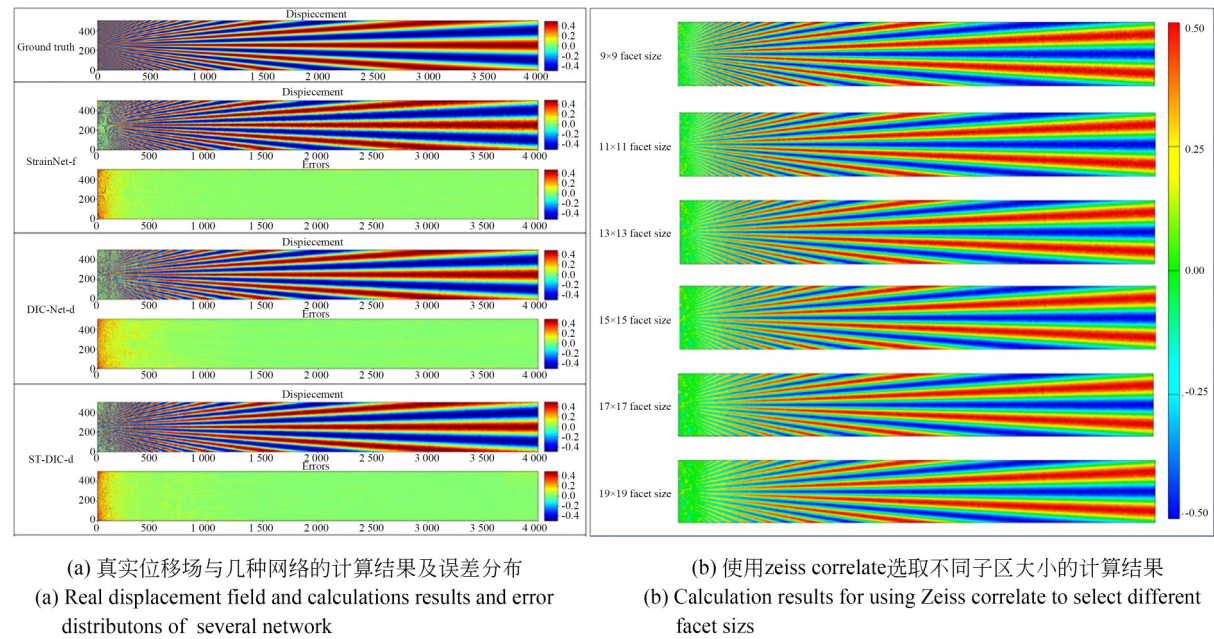
Model	Inference time (s/patch)	Training time (s/epoch)	GPU memory (GB)
ST-DIC-d	0.014 681	232.47	7.30
ST-DIC-s	0.013 702	202.52	7.30
D-ST	0.015 382	225.17	4.09
S-ST	0.016 169	244.09	13.80
DICnet-d	0.011 283	519.94	15.60
DICnet-s	0.010 387	494.87	15.60
StrainNet-f	0.008 251	—	—

3.1 位移测试

选择 StrainNet-f^[16]与 DIC-Net-d^[18]作为对比模型,在 DIC Challenge 2.0 的 Star5 测试图上进行位移测量。Star5 图像的分辨率为 4 000×501 像素,在 y 方向上呈现出正弦周期位移特征,其周期范围从左到右为 10 像素至 300 像素。图像左侧高频区域的测量结果能够直观体现模型在高频位移测量方面的性能。由于 ST-DIC-d 和 DIC-Net-d 无法直接处理该大尺寸图像,因此图像被分割为边长 128 像素、步长 112 像素(即重叠

宽度 16 像素)的 patch。因为步长 112 使每个内部像素至少处于距 patch 边缘 8 像素的位置,可完全覆盖自注意力感受野而不明显增加 patch 数量。检验 8,16,24 像素 3 种重叠宽度,当宽度<16 像素时,边缘出现可见拼接伪影,而 24 像素对最终误差改善不足 0.3%,故采用 16 像素。重叠区采用简单均值融合,其差异与高斯加权($\sigma=8$ pixel)在 MAE 上的平均差仅为 0.02 pixel。而 StrainNet-f 则可直接对整张图像进行处理。为了检验泛化性,未将 DIC Challenge 的变形模式纳入训练集中,而将 DIC-Net 重新训练了 300 轮。此外,还使用 Zeiss Correlate 软件(窗口尺寸设为 9~19)对 Star5 图像进行位移计算,以便进行对比。

图 3(a)展示了多种基于深度学习方法计算的位移场及误差分布;图 3(b)则给出了传统子集相关法(通过商业软件 Zeiss Correlation)计算的位移场,但受软件限制,无法获得误差分布。可以看出,随着子区尺寸的增大,传统方法对高频位移的测量能力明显减弱;即便采用最小的 9×9 子区(在实际应用中很少使用小于 9×9 的子区,因为很可能会导致误匹配),其测量效果依然不理想。位移场展示了 u,v 的合成结果,误差图则使用均方根(Root Mean Square, RMS)形式可视



(a) 真实位移场与几种网络的计算结果及误差分布
(a) Real displacement field and calculations results and error distributions of several network

(b) 使用zeiss correlate选取不同子区大小的计算结果
(b) Calculation results for using Zeiss correlate to select different facet sizes

图 3 DIC Challenge 2.0 的 Star5 位移计算结果及误差分布

Fig. 3 Calculation results and error levels for DIC Challenge 2.0 star-5 displacement challenge

化。可见深度学习模型在图像边缘及高梯度区域的误差波动较大,而传统 DIC 因子集尺寸固定限制了其对高频位移的分辨能力。从计算结果看,DIC-Net-d 和 ST-DIC-d 生成的位移场相比 StrainNet-f 更加平滑。

图 4 为 3 个模型在 y 方向预测误差的分布直方图。由图可以看出,DIC-Net-d 和 ST-DIC-d 的误差集中在较小区间,且 DIC-Net-d 在较小误差 (<0.5 像素) 部分数据略优。这与 DIC-Net-d 采

用插值而非反卷积上采样有关,其抑制了部分噪声,但也带来细节损失。如图 5 所示,在 $x=0\sim 500$ 时,DIC-Net-d 的预测较为光滑但在局部存在失真。同时,ST-DIC-d 在绝对误差 $0.05\sim 0.1$ 内比 DIC-Net-d 的分布点更少,这与其在全局 RMSE 上的优势一致。

图 6 则展示了 3 种模型在 x 方向(理想值为 0)的预测与误差分布。可以看出,DIC-Net-d 与 ST-DIC-d 在整体水平上都优于 StrainNet-f。

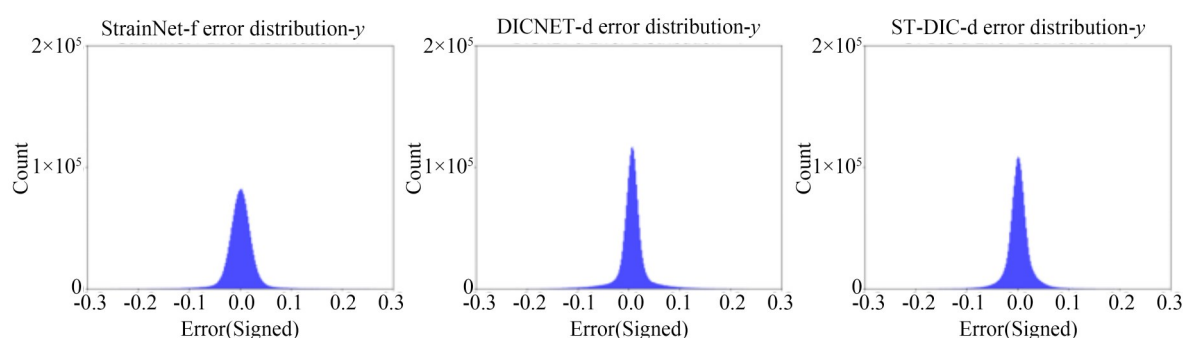


图 4 三种模型对 Star5 测试中 y 方向预测误差的分布(Zeiss 结果因未包含边缘像素故未统计在内)

Fig. 4 Distribution of the y -direction prediction error values for three models(The Zeiss Correlate results were not included in the statistics because its calculations exclude the pixels at the image edges)

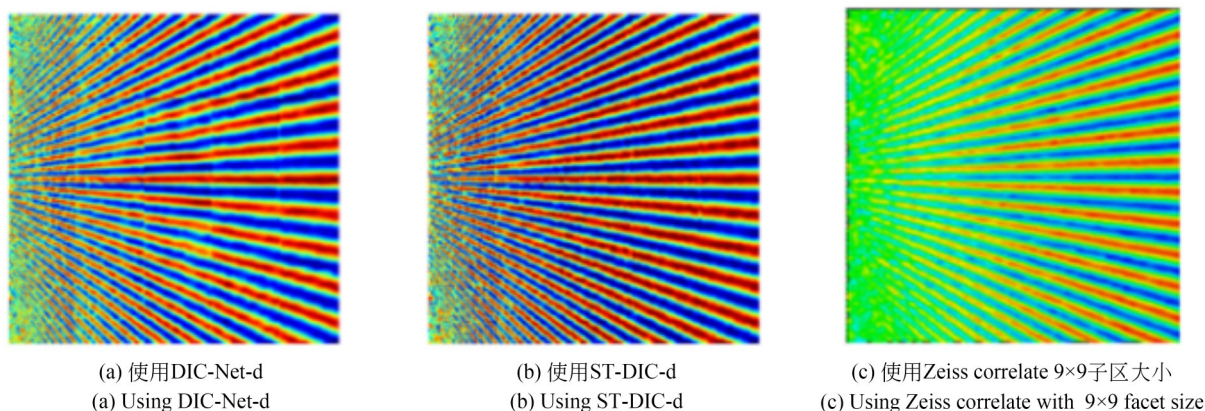


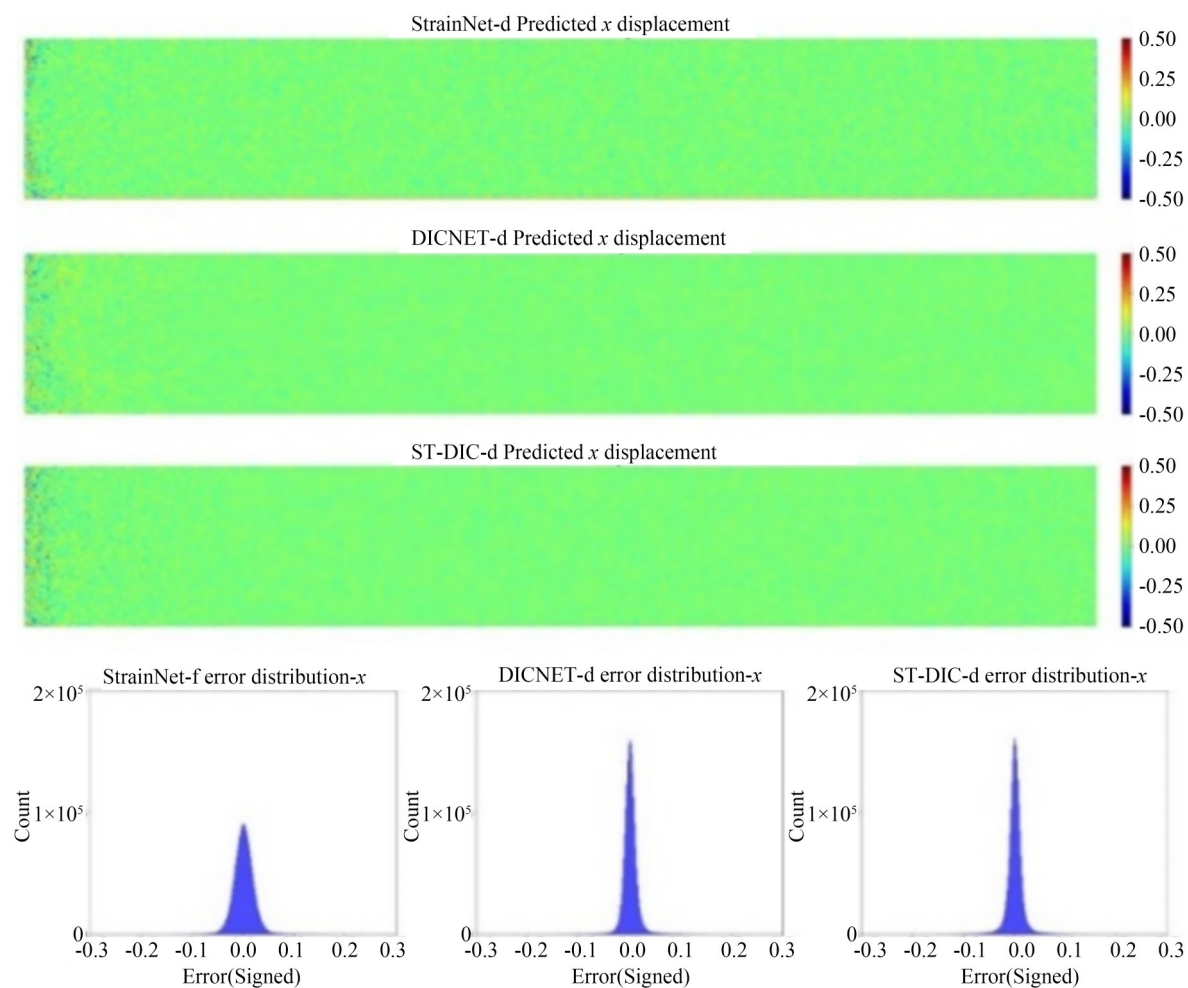
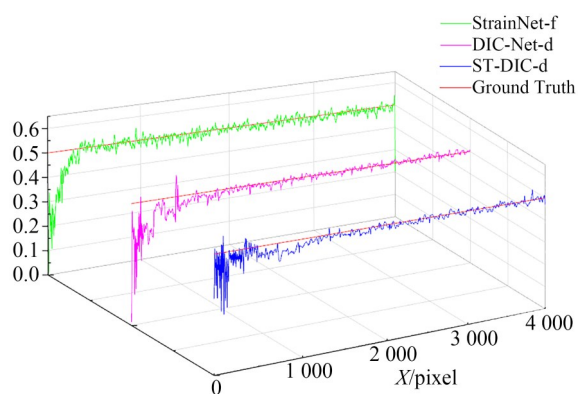
图 5 对 Star5 中 $x=0$ 到 $x=500$ 区域 y 方向位移的预测

Fig. 5 Prediction of y -direction displacement in region from $x=0$ to $x=500$ in Star5

为了更直观地对比空间频率对预测精度的影响,我们取 Star5 图像垂直中线($y=250$)上的位移进行分析,结果如图 7 所示。其中理想值设定为 0.5(实际变形曲线),StrainNet-f 整体偏差最大;在高频区段($x=0\sim 1\,000$),DIC-Net-d 误差显著大于 ST-DIC-d;但在中频区段($x=1\,000\sim 2\,000$)时,ST-DIC-d 相对误差又略微增加。在低

频区域,两者表现相当。

本文还比较了各模型对 Star5 位移场的空间分辨率。空间分辨率用于衡量测量方法区分位移场细节特征的能力,通常定义为位移衰减 10% 时对应的位移周期。选取 Star5 位移场,提取各方法在 $y=250$ 处的测量结果(理想值为 0.5),通过多项式拟合降低波动和误差的影响,计算测量

图 6 三种模型对 Star5 中 x 方向位移的预测与误差分布Fig. 6 Calculation results of x displacement and distribution of error values for three models图 7 理想值为 0.5 时三种模型在 $y=250$ 处的垂直位移计算结果Fig. 7 Vertical displacement calculations of three models at $y=250$ with ground-truth set at 0.5

值衰减至 0.45 像素时对应的周期,作为空间分辨率指标。计算结果如图 8 所示, DIC-Net-d 与 StrainNet-f 的空间分辨率分别为 23.44 pixel 和 19.40 pixel;而本文所提 ST-DIC-d 的空间分辨率为 16.56 pixel,由此可知,本文方法可以更好地区分细节特征。

表 2 总结了各模型在 Star5 数据集上的平均绝对误差 (Mean Absolute Error, MAE) 与 RMSE 结果。可见 ST-DIC-d 在多个关键指标上均优于 StrainNet-f 与 DIC-Net-d,尤其在 MAE_y 与 RMSE_y 上有明显优势。MAE_y 为 0.022 219,比 StrainNet-f 和 DIC-Net-d 分别减少 21% 和 16%。RMSE_y 亦是如此。该结果表明,ST-DIC-d 在复杂位移预测中具备更高的精度与稳定性。

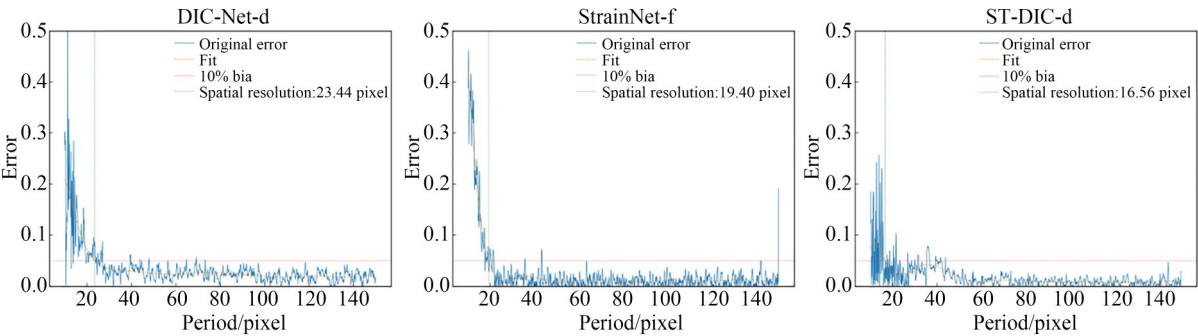


图 8 三种模型的空间分辨率
Fig. 8 Spatial resolution of three models

表 2 DIC Challenge(Star5)上各模型的预测误差

Tab. 2 Prediction errors of each model on DIC-Challenge(Star5) (pixel)

Model	MAE			RMSE		
	x	y	Total	x	y	Total
StrainNet-f	0.017 956	0.028 093	0.046 049	0.033 498	0.057 655	0.066 680
DIC-Net-d	0.014 258	0.026 586	0.040 844	0.024 875	0.050 632	0.056 413
ST-DIC-d	0.012 462	0.022 219	0.034 681	0.026 623	0.044 531	0.051 883

3.2 应变测试

在 DIC Challenge 中,ST-DIC-d 对位移的预测已显示出一定优势,但其应变网络(ST-DIC-s)在多个方面有更显著的提升。由于目前端到端预测应变的研究相对较少,为此选用 CNN 应变网络(CNN-Strain Net)进行对比。采用了该网络作者提供的数据集及其推荐的超参数,对 CNN-Strain Net 和 ST-DIC-s 各训练 300 轮,并随机选取 10% 数据做验证集。图 9 展示了两个模型在训练过程中的损失函数变化趋势。可见 CNN-Strain Net 在训练集上持续收敛,但验证集损失

波动较大,出现过拟合;而 ST-DIC-s 在训练集与验证集上具有更一致的收敛趋势。

选取训练中验证集 loss 最优的模型权重,对 DIC Challenge 2.0 中的 Star6 图进行应变预测,并与 Zeiss Correlate 软件计算结果对比,如图 10 所示。高频区域内,CNN-Strain Net 的预测出现明显失真,且在低频区域也存在可见误差。而 CNN-Strain Net 在训练早期的一些模型参数其实对高频区域更敏感,但整体性能仍不足。相比之下,采用 Sobel 卷积核的 ST-DIC-s 在低频区域块边界略显不连续,但通过缩小滑动步长可改善这一问题。对于高频应变区域,ST-DIC-s 预测效果明显优于 CNN-Strain Net,也优于传统 DIC。可以看出,Swin Transformer 对高空间频

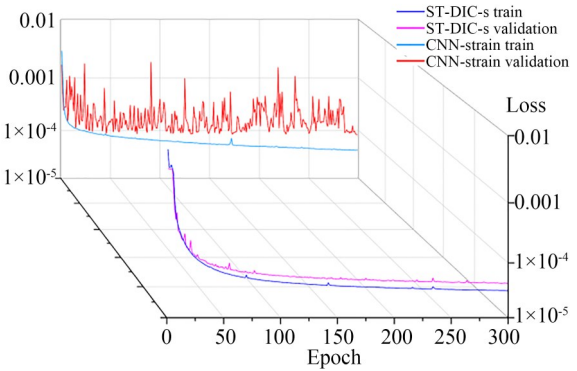


图 9 ST-DIC-s 与 CNN-Strain Net 在训练与验证集上的损失函数收敛对比
Fig. 9 Variation of loss during training process of ST-DIC-s compared to CNN-based strain network

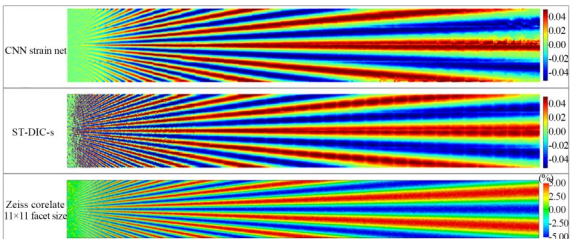


图 10 CNN-Strain Net, ST-DIC-s 与 Zeiss Correlate 的 Star6 图像应变分布
Fig. 10 Strain field of Star6 image predicted by CNN-strain net, ST-DIC-s and Zeiss correlate

率应变的捕捉能力较强;而 Zeiss Correlate 在低频区域总体效果不错,但在高频区域表现一般,也会带来更多噪声。

3.3 真实实验验证

为了验证 ST-DIC 在实际位移与应变测量中的性能,对铝合金试件进行了标准拉伸实验,如图 11 所示。实验设定拉伸速率为 5 mm/min,摄像频率为 5 Hz。使用 Dantec Dynamics 公司的 Q400 DIC 系统进行图像采集与标定,并使用 Istra4D 软件进行 DIC 计算(窗口尺寸为 17×17 ,步长 7 像素)。在对比中,将图像裁剪并分割为若干 128×128 的区域来输入 ST-DIC-d。

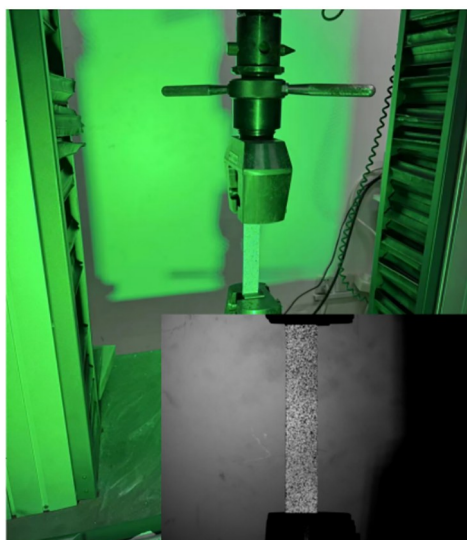


图 11 拉伸实验现场

Fig. 11 Tensile test

如图 12 所示,ST-DIC-d 与 Istra4D 在时域和空间分布上整体相似。ST-DIC-d 具有端到端高频测量的优势,对细节特征更敏感。例如,ST-DIC-d 更能凸显上方夹具区域与下方散斑区域的差异及裂纹附近的局部应变梯度变化。但在小位移区域,ST-DIC-d 也出现噪声,需要进一步针对低位移场景进行模型优化或结合工程降噪手段。而传统基于子集相关的 Istra4D 结果则较为平滑,但其内在插值必然会损失细节。

为验证 ST-DIC-d 对大位移的测量能力,在相同实验中测试了最大约 16 像素的位移,如图 13 所示。ST-DIC-d 和 Istra4D 在大位移测量上的结果高度吻合。这说明只要训练数据覆盖足够大的变形范围,SwiN Transformer 的窗口注意

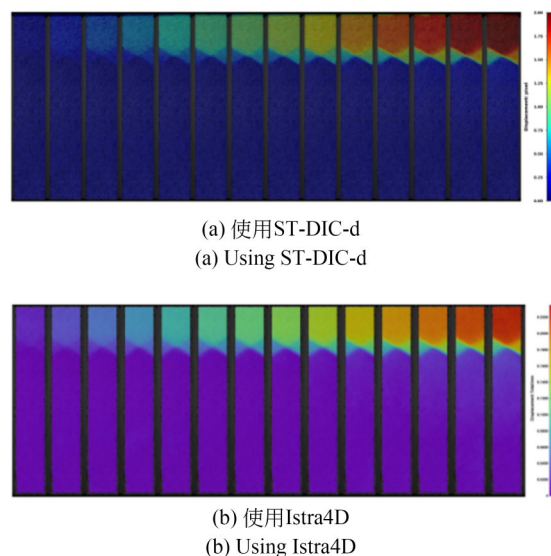


图 12 拉伸试验过程中不同模型 x 方向和 y 方向的复合位移测量结果对比(包含颈缩阶段连续捕获的 15 幅图像)

Fig. 12 Comparison of composite displacement measurement results in x and y directions using ST-DIC-d and Dantec Dynamics Istra4D during tensile testing(including 15 consecutive images captured during the necking phase)

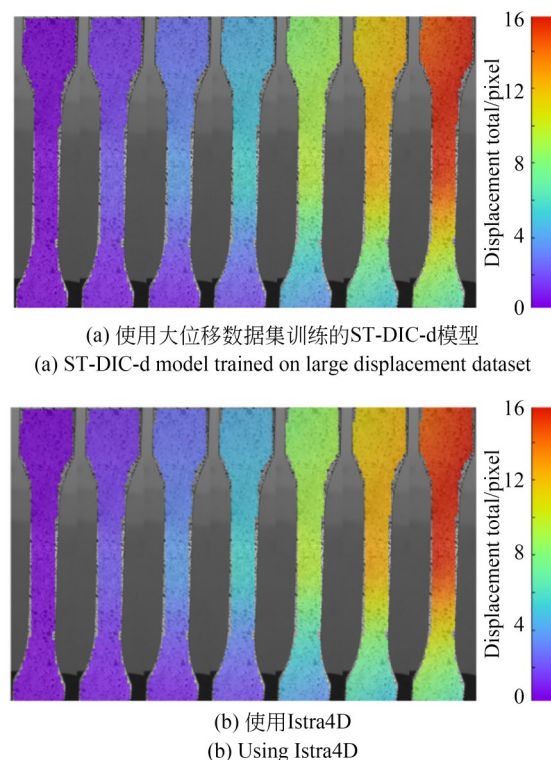


图 13 拉伸实验中的位移测量结果

Fig. 13 Displacement measurement results in tensile test

力便可应对大尺度位移。在可视化时,对云图加了与商业软件类似的高斯滤波进行平滑。

图 14 则对比了试件在颈缩阶段的应变分布。ST-DIC-s 与 Istra4D 的测量结果显示,二者在最大应变(约 8%)上相差不大,但 ST-DIC-s 的应变分布在高应变梯度区域更细腻,在应变值较小的下部区域则容易受到噪声影响。因此,仍需针对小应变区域做进一步优化。

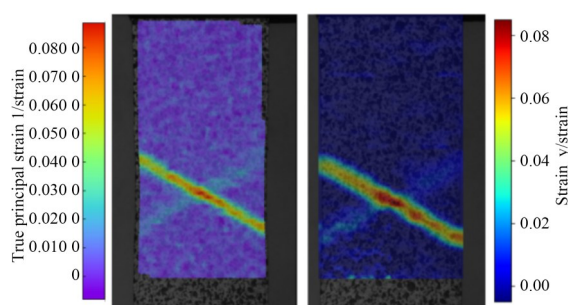


图 14 ST-DIC-s 与 Istra4D 在颈缩阶段应变云图对比

Fig. 14 Comparison of strain measurements between ST-DIC-s and Istra4D in necking phase

4 结 论

本文提出了基于 Swin Transformer 的数字图像相关法(ST-DIC)。在 DIC Challenge 和真实拉伸实验中,相比传统 DIC 方法及基于 CNN 的模型,ST-DIC 在应对高频位移和复杂变形模式时更具优势。与 StrainNet-f 和 DIC-Net-d 相比,该方法在 MAE_y 指标上减少了 21% 和 16%,突

显了其在垂直位移预测上的高精度。进一步分析显示,ST-DIC-d 在高频位移区域的性能尤其优异,证明了 Swin Transformer 通过全局自注意力来捕捉远距离依赖的独特优势。相比 CNN-Strain Net,ST-DIC-s 的训练收敛更平稳,验证集波动更小。在 DIC Challenge 的 Star6 测试中,ST-DIC-s 在高频应变区域的预测明显优于 CNN-Strain Net 及传统 DIC 方法,证实了其在复杂应变场测量中的有效性。与商业软件 Istra4D 相比,ST-DIC-d 测量的位移在时域与空间分布上高度一致,且对局部细节更敏感。ST-DIC-s 在高梯度区域的应变预测更准确,最大应变量与 Istra4D 基本相当;但在小应变区域需进一步抑制噪声。

总体而言,ST-DIC 在大变形、高频位移及复杂应变测量方面展现出优势,验证了 Swin Transformer 在 DIC 任务中的适用性与潜力。其层次化结构及窗口注意力机制能高效融合局部与全局特征,大幅提高位移与应变预测精度。然而,真实应用中图像尺寸和分辨率通常较训练集更大,需要将图像拆分为多个 patch 来处理,如何在拼接时平滑边缘区域、避免边界效应是下一步的研究方向。

作者贡献声明:

但西佐:模型搭建,论文构思和撰写;
杨浩:实验设计及数据分析;
李桂华:论文审核与编辑。

参考文献:

- [1] JANELIUKSTIS R, CHEN X. Review of digital image correlation application to large-scale composite structure testing [J]. *Composite Structures*, 2021, 271: 114143.
- [2] 张贵阳, 霍炬, 杨明, 等. 多相机网络联合约束优化的高精度三维变形全场测量[J]. *光学精密工程*, 2021, 29(7): 1653-1666.
ZHANG G Y, HUO J, YANG M, *et al.* High-precision and full-field measurement of 3D deformation based on multi-camera network joint constraint optimization[J]. *Opt. Precision Eng.*, 2021, 29(7): 1653-1666. (in Chinese)
- [3] HILD F, ROUX S. Comparison of local and global approaches to digital image correlation[J]. *Experimental Mechanics*, 2012, 52(9): 1503-1519.
- [4] PAN B. Recent progress in digital image correlation[J]. *Experimental Mechanics*, 2011, 51(7): 1223-1235.
- [5] LI J R, YAO Y F, ZHAO Q H, *et al.* Simultaneous measurement of six displacement gradients using Digital Shearography [J]. *Measurement*, 2025, 240: 115520.
- [6] 王立忠, 张振, 赵建博, 等. 小幅面高温焊缝变形的数字图像相关测量[J]. *光学精密工程*, 2021, 29(4): 802-812.
WANG L ZH, ZHANG ZH, ZHAO J B, *et al.* Digital image correlation measurement of high tem-

- perature weld seam deformation with small format [J]. *Opt. Precision Eng.*, 2021, 29(4): 802-812. (in Chinese)
- [7] SUTTON M A, HILD F. Recent advances and perspectives in digital image correlation[J]. *Experimental Mechanics*, 2015, 55(1): 1-8.
- [8] BORNERT M, BRÉMAND F, DOUMALIN P, *et al.* Assessment of digital image correlation measurement errors: methodology and results[J]. *Experimental Mechanics*, 2009, 49(3): 353-370.
- [9] KHOO S W, KARUPPANAN S, TAN C S. A review of surface deformation and strain measurement using two-dimensional digital image correlation[J]. *Metrology and Measurement Systems*, 2016, 23(3): 461-480.
- [10] MA Q J, REJAB M R M, HALIM Q, *et al.* Experimental investigation of the tensile test using digital image correlation (DIC) method[J]. *Materials Today: Proceedings*, 2020, 27: 757-763.
- [11] YONEYAMA S. Basic principle of digital image correlation for in-plane displacement and strain measurement [J]. *Advanced Composite Materials*, 2016, 25(2): 105-123.
- [12] XU X H, SU Y, CAI Y L, *et al.* Effects of various shape functions and subset size in local deformation measurements using DIC [J]. *Experimental Mechanics*, 2015, 55(8): 1575-1590.
- [13] SUN Y F, PANG J H L. Study of optimal subset size in digital image correlation of speckle pattern images [J]. *Optics and Lasers in Engineering*, 2007, 45(9): 967-974.
- [14] YANG R, LI Y, ZENG D, *et al.* Deep DIC: Deep learning-based digital image correlation for end-to-end displacement and strain measurement [J]. *Journal of Materials Processing Technology*, 2022, 302: 117474.
- [15] MIN H G, ON H I, KANG D J, *et al.* Strain measurement during tensile testing using deep learning-based digital image correlation[J]. *Measurement Science and Technology*, 2020, 31(1): 015014.
- [16] BOUKHTACHE S, ABDELOUAHAB K, BERRY F, *et al.* When deep learning meets digital image correlation[J]. *Optics and Lasers in Engineering*, 2021, 136: 106308.
- [17] ILG E, MAYER N, SAIKIA T, *et al.* FlowNet 2.0: evolution of optical flow estimation with deep networks[C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. July 21-26, 2017, Honolulu, HI, USA. IEEE, 2017: 1647-1655.
- [18] WANG Y, ZHAO J Q. DIC-Net: Upgrade the performance of traditional DIC with Hermite dataset and convolution neural network[J]. *Optics and Lasers in Engineering*, 2023, 160: 107278.
- [19] WANG Y X, ZHOU C L. Unsupervised CNN-based DIC method for 2D displacement measurement[J]. *Optics and Lasers in Engineering*, 2024, 174: 107981.
- [20] RONNEBERGER O, FISCHER P, BROX T. *U-Net: Convolutional Networks for Biomedical Image Segmentation*[M]. Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Cham: Springer International Publishing, 2015: 234-241.
- [21] LIU Z, LIN Y T, CAO Y, *et al.* Swin transformer: hierarchical vision transformer using shifted windows[C]. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*. October 10-17, 2021, Montreal, QC, Canada. IEEE, 2021: 9992-10002.
- [22] ZHOU Y F, ZUO Q J, CHEN N, *et al.* Transformer based deep learning for digital image correlation[J]. *Optics and Lasers in Engineering*, 2025, 184: 108568.
- [23] CAO H, WANG Y Y, CHEN J, *et al.* Swin-unet: Unet-like Pure transformer for medical image segmentation [M]. Computer Vision-ECCV 2022 Workshops. Cham: Springer Nature Switzerland, 2023: 205-218.
- [24] HE X, ZHOU Y, ZHAO J Q, *et al.* Swin transformer embedding UNet for remote sensing image semantic segmentation [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 4408715.
- [25] FAN C M, LIU T J, LIU K H. SUNet: swin transformer UNet for image denoising [C]. 2022 *IEEE International Symposium on Circuits and Systems (ISCAS)*. May 27-June 1, 2022, Austin, TX, USA. IEEE, 2022: 2333-2337.
- [26] DAN X Z, GUO H D, HU Y W, *et al.* Transformer-enhanced end-to-end models for accurate displacement and strain fields in digital image correlation[J]. *Optics Express*, 2025, 33(3): 5191.

- [27] REU P L, BLAYSAT B, ANDÓ E, *et al.* DIC challenge 2.0: developing images and guidelines for evaluating accuracy and resolution of 2D analyses [J]. *Experimental Mechanics*, 2022, 62 (4): 639-654.
- [28] PAN B, XIE H M, GUO Z Q, *et al.* Full-field strain measurement using a two-dimensional Savitzky-Golay digital differentiator in digital image correlation [J]. *Optical Engineering*, 2007, 46 (3): 033601.

作者简介:



但西佐(1991—),男,安徽安庆人,博士,讲师,2019年于合肥工业大学获得博士学位,主要从事计算机视觉及深度学习方面的研究。E-mail: danxizuo@ahu.edu.cn

通讯作者:



李桂华(1971—),女,安徽合肥人,博士,教授,硕士生导师,1994年、1999年、2006年于合肥工业大学分别获得学士、硕士和博士学位,主要从事热误差理论及仪器精度及生物材料的热物理性能方面的研究。E-mail: guihuali1@163.com