文章编号：1005-6734(2025)10-0955-08 doi.10.13695/j.cnki.12-1222/o3.2025.10.001

深层次捕获时空特征的神经网络 IMU 定位模型

吴仕勋[1]，韩 金[1]，许登元[1]，侯忠伟[2]，聂 觅[3]
（1. 重庆交通大学 信息科学与工程学院，重庆 400074；2. 重庆交通大学 未来土木科技研究院，
重庆 400074；3. 重庆城投基础设施建设有限公司，重庆 400010）

摘要：针对现有神经网络模型在惯性导航中忽略惯性测量单元（IMU）序列时间特征、相互依赖性与周期性，导致定位精度下降的问题，提出一种深度融合 Xception 与 Transformer 结构的神经网络 IMU 定位模型。该模型通过构建适合学习速度向量的初步提取层、深层次提取层和速度回归层，以捕获 IMU 序列的复杂时空特性。在四种公开 IMU 数据集（RONIN、RIDI、IDOL 和 IMUNET）上验证模型的有效性。实验结果表明，与当前五种主流模型相比，所提模型在大多数已知与未知测试集上的定位性能都有所提升。其中，在规模最大的 RONIN 数据集上，与最差的模型相比绝对轨迹误差分别减少了 17.16%和 13.15%；在规模最小的 IDOL 数据集上，分别减少了 28.29%和 22.96%。这些结果表明模型能够提供更准确和鲁棒的速度预测，从而显著提升 IMU 定位精度。

关 键 词：惯性定位；神经网络；速度预测；惯性测量单元
中图分类号：TP183 **文献标志码：**A

Neural network IMU localization model for deep level capture of spatiotemporal features

WU Shixun[1], HAN Jin[1], XU Dengyuan[1], HOU Zhongwei[2], NIE Mi[3]
(1. School of Information Science and Engineering, Chongqing Jiaotong University, Chongqing 400074, China;
2. Institute of Future Civil Engineering Sciences and Technology, Chongqing Jiaotong University, Chongqing 400074, China; 3. Chongqing Urban Investment Infrastructure Construction Co., Ltd., Chongqing 400010, China)

Abstract: To address the issue of existing neural network models in inertial navigation overlooking the temporal characteristics, interdependencies, and periodicity of inertial measurement unit (IMU) sequences, which leads to degraded positioning accuracy, a IMU positioning neural network model is proposed that deeply integrates Xception and Transformer architectures. The proposed model employs an initial feature extraction layer, a deep feature extraction layer, and a velocity regression layer, which are tailored for learning velocity vectors, in order to capture the complex spatiotemporal characteristics of IMU sequences. To validate the effectiveness of the proposed model, experiments are conducted on four publicly available IMU datasets (RONIN, RIDI, IDOL and IMUNET). Experimental results demonstrate that, the proposed model achieves improved localization performance on most seen and unseen test sets compared with five state-of-the-art models. Specifically, on the largest RONIN dataset, the absolute trajectory error is reduced by 17.16% and 13.15% relative to the weakest baseline model. On the smallest IDOL dataset, the reductions reach 28.29% and 22.96%, respectively. These results indicate that the proposed model provides more accurate and robust velocity predictions, thereby significantly enhancing IMU-based localization accuracy.

Key words: inertial positioning; neural network; speed prediction; inertial measurement unit

收稿日期：2024-12-16；**修回日期：**2025-08-05
基金项目：国家重点研发计划项目子课题（2021YFB2600103-01）；重庆市教育委员会在渝高校与中科院所属院所合作项目（HZ2021009）；重庆市自然科学基金（CSTB2024NSCQ-MSX0275, CSTB2025NSCQ-GPX0839）；重庆市教委科学技术研究重点项目（KJZD-K202500702）

作者简介：吴仕勋（1983—），男，博士，副教授，硕士生导师，研究方向为无线通信，无线定位及人工智能。
通讯作者：侯忠伟（1986—），男，博士，副教授，硕士生导师，研究方向为土木工程智能化技术。

高效精确的自主定位技术对于提升行人与车辆的导航效能及复杂环境适应性具有关键作用。全球卫星导航系统虽是主流定位方案，但在室内或建筑物遮挡的情况下，其信号会遭受严重遮挡而丧失定位功能。在此背景下，惯性测量单元（Inertial Measurement Unit, IMU）凭借其低成本、高自主性以及强抗干扰能力，广泛应用于自动驾驶、航空航天、医疗检测和地质勘探等多个行业，展现出不可或缺的价值与潜力[1]。

常用的 IMU 由于制造精度和工艺限制，存在零偏误差、正交耦合误差、比例因子误差和随机噪声等误差源，导致定位结果随时间推移逐渐偏离实际值。为缓解定位过程中的误差漂移，行人航位推算通过检测步长与航向来更新位置，可减少惯性定位误差漂移，但步长和航向估计不准确仍会造成较大定位误差[2]。通过识别出行走时的静止状态，零速更新算法将零速度作为观测值补偿 IMU 误差，但 IMU 需固定在脚部，限制了应用场景且会受到运动干扰导致检测不佳[3]。此外，通过非线性滤波可将 IMU 测量数据与超宽带技术[4]、激光雷达[5]及里程计[6]等多传感器进行深度融合，从而有效抑制 IMU 误差漂移并提升定位精度，但需克服传感器间时间同步误差和空间杆臂误差[7]的挑战。

近年来，神经网络为 IMU 序列处理提供了新方法。端到端模型无需额外传感器辅助，通过学习连续 IMU 序列特征，可生成更准确的速度、航向角和轨迹信息，有效缓解误差漂移问题[8]。现有神经惯性定位方法按学习目标可分为学习速度和学习位移两类。在学习位移方法中，Yan 等人[9]首次采用神经网络处理 IMU 数据，先通过线性最小二乘法校正线性加速度，再进行二次积分获得精确位移。

尽管位移是 IMU 定位中最直观的输出，但速度与加速度之间存在更紧密的相互依存关系，这使得学习速度的神经网络模型成为该领域的研究热点。Herath 等人[10]提出了结构简洁的 RoNIN 模型，首次将 IMU 数据回归为二维速度向量，对此积分后获得位移。后续研究主要从三个方向进行改进：在注意力机制方面，Chen 等人[11]采用 Res2net 模块融合双卷积注意力模块进行速度回归，增强特征提取和细粒度表示能力以提升预测精度；在模型轻量化方面，Zeinali 等人[12]将深度可分离卷积引入 Resnet，显著减少训练参数并提升模型效率，使轻量化模型可部署于移动设备实现准确高效定位；在模型融合方面，Wang 等人[13]提出了一种引入时间注意机制的混合神经网络，在 CNN 提取空间特征的基础上加入 LSTM 来捕获全局时间信息，并利用时间注意机制对 LSTM 隐藏层输出进行加权求和，得到最终的速度及角度信息。

综上所述，学习速度的神经网络模型虽已取得了显著进步，但仍面临三重挑战：在注意力机制方面，过度关注序列长度压缩和特征维度提升，却忽略了 IMU 序列在行人行走中所体现出的时间特征、不同输入信号之间复杂的相互依赖性以及对空间特征捕获的不足；在轻量化模型方面，过度简化模块导致预测精度受限；在模块融合方面，模型之间无法优势互补，时间特征仅关注当前时刻及此前的序列，忽略了整个行走过程中 IMU 序列固有的周期性规律。为此，本文基于 Xception 模型[14]，提出了改进的深层次捕获时空特征网络模型，在有效缩减模型参数的同时，精准聚焦 IMU 传感器数据在空间上所展现出的独特特征并进行深层次提取。此外，进一步融合 Transformer 模型[15]，全面捕捉 IMU 序列的时间动态特征，通过多头注意力机制捕获行人在室内行走过程中 IMU 序列展现出的周期性特性。最后，通过深度可分离卷积与全连接层结合对数据特征进行降维得到二维速度向量，全面捕捉行人行走时 IMU 序列所蕴含的复杂时空特性，提升 IMU 定位精度。

1 系统模型

IMU 输出的角速度和加速度易受传感器内部偏差和噪声干扰，其数学表达式为：

$$\boldsymbol{w}_t = \boldsymbol{r}_t^w + \boldsymbol{b}_t^w + \boldsymbol{n}_t^w \tag{1}$$

$$\boldsymbol{a}_t = \boldsymbol{r}_t^a + \boldsymbol{b}_t^a + \boldsymbol{n}_t^a \tag{2}$$

其中，$\boldsymbol{w}_t$ 和 $\boldsymbol{a}_t$ 分别表示 t 时刻三轴角速度和加速度矢量，$\boldsymbol{r}_t^w$ 和 $\boldsymbol{r}_t^a$ 为真实的角速度和加速度，$\boldsymbol{b}_t^w$ 和 $\boldsymbol{b}_t^a$ 为角速度和加速度的时变零偏，$\boldsymbol{n}_t^w$ 和 $\boldsymbol{n}_t^a$ 为角速度和加速度的噪声值。

时变零偏和噪声导致 IMU 测量值与真实值存在显著偏差。传统惯性定位算法在通过误差标定模型补偿稳定误差源后，基于牛顿力学原理对角速度积分获得角度，该角度作为旋转向量将加速度从载体坐标系转换至全局坐标系，继而通过一次积分得到速度、二次积分获得位移，这就是捷联惯导更新算法[16]。然而，该算法未对 IMU 动态误差进行实时校正，仅通过复杂计算直接输出最终定位结果。由于动态误差会在积分过程中持续累积，最终导致位移误差显著增大。

为解决原始数据误差问题，可将加速度和角速度作为整体输入神经网络提取特征向量，并将其回归为速度[11-13]。随后，通过将计算的平均时间间隔与对应的速度值相乘并进行累加，从而得到位置信息。整个过程避免了连续积分的误差累积，通过神经网络学习数据的内在规律来识别随机漂移特性，并通过回归修正最小化其影响。处理流程可表述为：

$$\{\boldsymbol{a}_t, \boldsymbol{w}_t\} \to \vec{\boldsymbol{v}}_t \tag{3}$$

$$\Delta t = \frac{1}{n-1}\sum_{k=1}^{n-1}(t_{i_{k+1}} - t_{i_k}) \tag{4}$$

$$\boldsymbol{p}(K) = \boldsymbol{p}(0) + \sum_{j=1}^{K}\vec{\boldsymbol{v}}_j \Delta t \tag{5}$$

其中，$\vec{\boldsymbol{v}}_t$ 表示通过神经网络模型预测 t 时刻的速度向量；Δt 表示时间间隔的平均值；t_{i_k} 表示第 k 个时间点的时间戳；n 为时间点总数；$\boldsymbol{p}(0)$ 表示初始位置，即时间步 $K=0$ 时的位置；$\boldsymbol{p}(K)$ 为计算时间步 K 的位置；$\sum_{j=1}^{K}\vec{\boldsymbol{v}}_j \Delta t$ 为从时间步 1 到 K 的所有速度向量与时间间隔的乘积之和，即从初始位置到第 K 个时间步的总位移。

为使预测的轨迹点与真实的时间戳对齐，使轨迹在时间上更准确平滑，通过扩展时间戳并采用线性插值计算特定时间点的轨迹，其计算过程如下：

$$\boldsymbol{t}_{ext} = [t_0 - \varepsilon, t_1, t_2, L, t_n, t_n + \varepsilon] \tag{6}$$

$$\boldsymbol{p}(t) = \boldsymbol{p}_i + \frac{t - t_i}{t_{i+1} - t_i}(\boldsymbol{p}_{i+1} - \boldsymbol{p}_i) \quad t \in [t_i, t_{i+1}] \tag{7}$$

其中，$\boldsymbol{t}_{ext}$ 表示扩展的时间序列；$\boldsymbol{p}_i$ 为每个时间点 t_i 的预测位置，用来作为插值的起点；时间扩展量 ε 取值为 1×10^{-6}，这一微秒级的扩展量远低于 100 Hz 和 200 Hz 的采样频率，因此对数据序列的时间精度影响可忽略。同时，该值又足够大，能避免数值计算中的舍入误差，确保插值过程不会因时间戳边界问题导致插值失败或轨迹不平衡。

2　深层次捕获时空特征的神经网络定位模型

鉴于 Xception 模型在图像分类、目标检测等领域展现出的优秀性能[14]，本文以此模型为基础设计了如图 1 所示的系统框架图，整个框架核心分为初步提取层、深层次提取层和速度回归层。通过该系统框架对 IMU 序列的时空特征进行深层次提取并最终回归为二维速度向量。

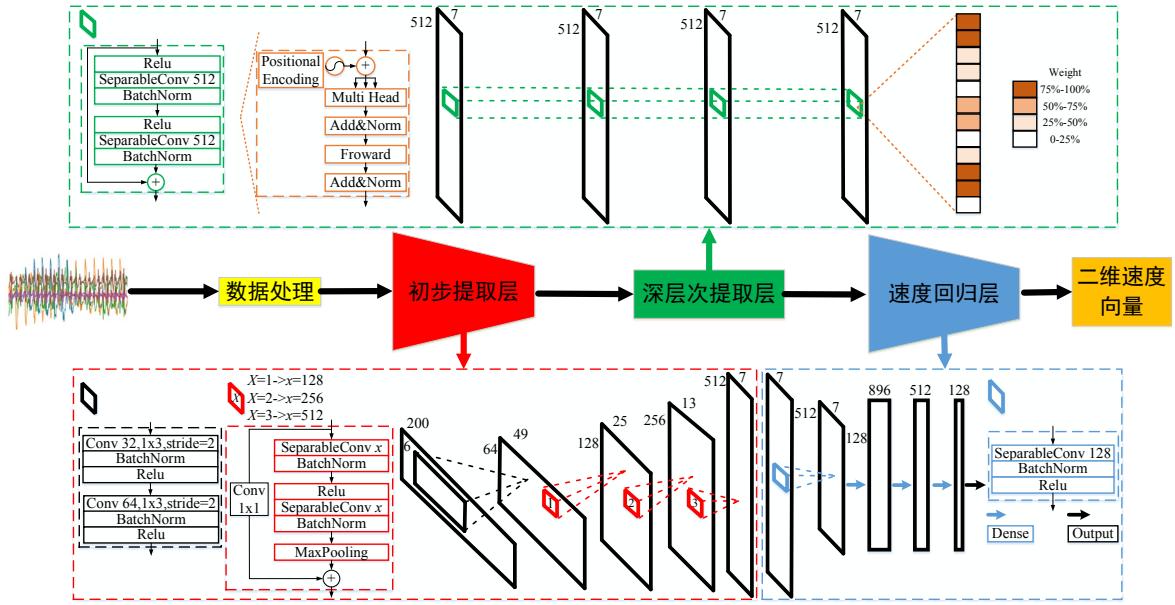

图 1　系统框架

Fig.1　System framework

2.1　数据处理

由于 IMU 设备与真实值采集设备的坐标系独立，因此在将数据输入初步提取层之前，需先将 IMU 数据转换至真实值坐标系，以确保预测速度计算的轨迹与真实轨迹对齐。转换公式如下：

$$\boldsymbol{R}_{W_I}^{W_T} = \boldsymbol{R}_{L_T}^{W_T} \cdot \boldsymbol{R}_{L_I}^{L_T} \cdot \left(\boldsymbol{R}_{L_I}^{W_I}\right)^{-1} \tag{8}$$

$$[\boldsymbol{a}_t, \boldsymbol{w}_t]_{W_T} = \boldsymbol{R}_{W_I}^{W_T} \cdot [\boldsymbol{a}_t, \boldsymbol{w}_t]_{W_I} \tag{9}$$

其中，W_T 为真实值设备的导航坐标系，W_I 为 IMU 设备的导航坐标系，$\boldsymbol{R}_{W_I}^{W_T}$ 表示从 IMU 设备导航坐标系旋转到真实值设备导航坐标系的旋转矩阵；L_T 为真实值设备的载体坐标系，L_I 为 IMU 设备的载体坐标系，$\boldsymbol{R}_{L_I}^{L_T}$ 表示从 IMU 设备载体坐标系旋转到真实值设备载体坐标系的旋转矩阵，通过将两个设备对齐的方式获得，如果 IMU 数据采集设备与真实值采集设备为同一个则该矩阵可以忽略；$\boldsymbol{R}_{L_T}^{W_T}$ 表示从真实设备载体坐标系转换到其导航坐标系的旋转矩阵，可从真实设备中得到，真实设备包括谷歌开发的 Tango 技术[10]、光学动作捕捉系统 Vicon 等[17]；$\boldsymbol{R}_{L_I}^{W_I}$ 表示从 IMU 设备的载体坐标系转换到其导航坐标系的旋转矩阵，从采集 IMU 数据的设备中获得。

在完成坐标系转换后，对 IMU 序列采用窗口大小

为 200、步长为 10 的滑动窗口进行切割（步长指窗口每次移动 10 个数据点）。通过引入随机偏移的数据增强技术，窗口起始位置可在预设范围内随机偏移 0 至 10 个数据点，从而提升数据多样性以增强模型训练时的泛化能力。最后，数据以 128 为批次大小进行训练，使神经网络输入为批次大小为 128、特征维度为 6、序列长度为 200 的 IMU 序列。

2.2 初步提取层

以 Xception 模型的 Entry flow 层[14]为基础构建初步提取层，该层由两个卷积层（Conv）和三个深度可分离卷积层（DS-Conv）组成，用于从输入的 IMU 序列中提取基础特征，为后续深层网络的精细处理提供特征基础。具体参数配置见表 1。

表 1 初步提取层的结构

Tab.1 Structure of preliminary extraction layer

层	输入尺寸	输出尺寸	具体细节
Conv1	6×200	32×99	1×3, 32, stride=2
Conv2	32×99	64×49	1×3, 64, stride=2
DS-Conv1	64×49	128×25	[1×3,128]×2 1×3 MaxPool, stride=2
DS-Conv2	128×25	256×13	[1×3,256]×2 1×3 MaxPool, stride=2
DS-Conv3	256×13	512 x 7	[1×3,512]×2 1×3 MaxPool, stride=2

由表 1 可以看出，与 Xception 模型结构[15]不同，初步提取层将卷积核尺寸由 3×3 改为适用 IMU 数据处理的 1×3，并将第二个卷积核步长设为 2，使卷积层感受野增加到 7 个时间点，具体参数对比如图 2 所示。

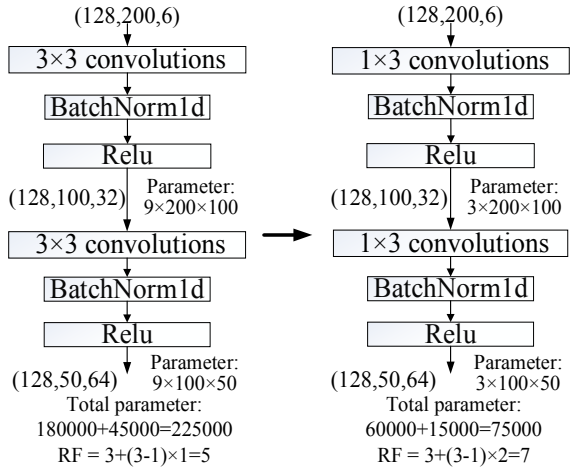

图 2 卷积核的替换

Fig.2 Replacement of convolutional kernel

图 2 对比显示：相较于原始结构，替换后的结构在减少计算参数的同时扩大了卷积核感受野。关键改进在于：1）采用大感受野卷积核能捕捉更远距离的依赖关系，帮助模型更好理解加速度和角速度随时间的变化，这对行人运动状态识别至关重要。该设计的初衷在于为后续层提取 IMU 数据的特征输入更合适的序列长度和特征维度，最大限度减少信息在传递过程中的丢失。2）采用深度可分离卷积对压缩数据中的每个序列进行独立空间卷积，不断扩大数据特征维度以丰富其表征能力。3）引入池化层进行空间下采样，有效减少了数据序列长度。同时，采用残差连接的跳跃结构直接将输入信息传递到后续层，有效避免了梯度消失的问题。

初步提取层的设计旨在逐步缩短 IMU 数据的序列长度，并在此过程中不断提升特征维度。这样，在早期阶段就能有效提取到 IMU 数据的基础特征，为后续更深层次的特征提取层提供更加有效和更具代表性的输入数据。

2.3 深层次提取层

初步提取层主要聚焦于数值大小的基础属性，诸如峰值、谷值等直观的数据特征。相比之下，深层次提取层则借鉴了 Xception 模型中的 Middle flow 层设计[14]，致力于挖掘更为复杂且精细的特征，例如峰值的相对位置、数值间的动态变化趋势等，这些深层次特征恰好对应行人行走时的转向、加速或减速等关键行为模式。与初步提取层不同，深层次提取层摒弃了对 IMU 数据的下采样处理，转而聚焦于现有特征的增强，通过多次应用深度可分离卷积技术，提取到更丰富、更细致的特征信息，并引入残差连接和非线性激活函数来保证网络的高效训练。

鉴于 IMU 序列在行走过程中表现出的周期性特征，卷积结构虽擅长提取局部特征，却难以应对这类序列中存在的长距离依赖问题。因此，在深层次提取层的最后阶段融入 Transformer 架构，通过其独特的注意力机制建模输入序列中不同位置的全局依赖性，从而为每个时间步的 IMU 数据赋予更高权重，对受噪声干扰较大的区域赋予较低权重，以此来弱化某个时间段内受噪声信号干扰较大的 IMU 数据。具体参数配置见表 2。

表 2 深层次提取层的结构

Tab.2 Structure of deep extraction layer

层	输入尺寸	输出尺寸	具体细节
DS-Conv1	512×7	512×7	[1×3, 512]×3
DS-Conv2	512×7	512×7	[1×3, 512]×3
DS-Conv3	512×7	512×7	[1×3, 512]×3
DS-Conv4	512×7	512×7	[1×3, 512]×3
Transformer	512×7	512×7	FFN=1024, Heads=4 Layer=1, dropout=0.1

通过优化 Xception 模型的 Middle flow 层实现了对 IMU 数据特征的精细提取，并融合 Transformer 的全局建模优势，显著增强了网络表达能力。即便面对

庞大的数据集，该网络也能迅速且高效捕捉其中关键的数据特征。此外，卷积层利用局部感受野特性有效平滑 IMU 数据的局部噪声，而 Transformer 则通过注意力机制抑制全局噪声，并利用长距离依赖关系滤除短期噪声。深层次提取层将卷积与 Transformer 相结合，不仅提升了 IMU 数据处理的鲁棒性和准确性，还显著降低了 IMU 噪声的影响。

2.4　速度回归层

速度回归层由一个深度可分离卷积和全连接层组成，结构比 Xception 模型中的 Exit flow 层[14]更简洁高效，用于整合深层次提取层提取到的特征向量。首先通过深度可分离卷积处理特征向量，并采用对 IMU 数据进行降维的策略。与原始 Xception 模型中 Exit Flow 层通过升维以增强图像空间表征能力不同，速度回归层的降维策略侧重于凝练深层次提取层进一步提取的有效特征，同时抑制噪声与冗余信息，从而为后续的回归任务提供高质量的输入。

经过深度可分离卷积处理后，特征向量依次通过三层全连接结构进行转换：第一层（Linear 1）将三维数据初步压缩至二维空间；第二层（Linear 2）在保持维度不变的前提下对特征进行平滑；第三层（Linear 3）进一步压缩并输出最终的二维速度向量。这一设计基于行人室内行走的特性：z 轴方向的速度和位移基本保持恒定，因此仅需回归二维速度分量，具体参数如表 3 所示。

表 3　速度回归层的结构

Tab.3　Structure of velocity regression layer

层	输入尺寸	输出尺寸	具体细节
DS-Conv1	128×512×7	128×128×7	[1×3, 128]×1
Linear1	128×128×7	128×512	128×7->512
Linear2	128×512	128×512	512->512
Linear3	128×512	128×2	512->2

通过全连接层将压缩后的特征向量映射为二维输出，生成预测速度向量，通过真实位移值计算的速度向量构成均方误差损失函数。该损失函数记作 L_{MSE}，其表达式为：

$$L_{\mathrm{MSE}}=\frac{1}{l}\sum_{i=1}^{l}\| \boldsymbol{Y}_i-\hat{\boldsymbol{Y}}_i \|^2 \tag{10}$$

其中，$\boldsymbol{Y}_i$ 为真实位移值计算的速度向量，$\hat{\boldsymbol{Y}}_i$ 为模型预测的速度向量，l 为数据点的总数。

为增强网络的非线性表达能力，每个线性层后均加入相应的激活函数，显著提升了整个模型的拟合能力。但全连接层参数过多易导致过拟合，因此在激活函数后引入正则化技术以抑制过拟合，从而提高模型回归性能。

3　实验结果与分析

为全面评估模型性能，将本文模型与五种现有 IMU 神经网络定位模型在四个公开数据集上进行对比分析。所有实验基于 Pytorch 2.4.0 实现，采用 Adam 优化器更新网络参数（设置学习率为 0.0001），并通过学习率的调度器在训练模型时动态调整学习率：当损失函数经过 10 个 epoch 后仍未改善时，则按 0.1 的比例降低学习率，下限设为 1×10^{-12} 以免学习率过于接近零。

3.1　数据集

实验选用 RONIN[10]、RIDI[9]、IDOL[17] 和 IMUNET[12]四个公开数据集进行验证分析。如表 4 所示，除 RIDI 采用单设备采集数据外，其余数据集均使用两台设备分别记录 IMU 数据与真实轨迹。对于 RONIN 与 IMUNET 数据集，基准设备固定于胸前而 IMU 设备随机放置，导致两者相对位置并不固定，需通过相应预处理措施实现坐标统一；而 IDOL 数据集的两台设备同装于手持平台，相对位置保持恒定，可直接完成坐标对齐。这些数据集覆盖手持、口袋和腿部等多种运动场景，并已预先划分为训练集和测试集。测试集既包含训练集中已出现的场景，也涵盖未见场景（IMUNET 除外），确保验证全面性。

表 4　数据集描述

Tab.4　Dataset description

数据集	测试设备/基准设备	频率	携带方式	采集时长
RONIN	Galaxy S9 Pixel 2XL / Asus Zenfone AR	200 Hz	腿部，背包，手持，胸前固定	42.5 h
RIDI	Lenovo Phab2 Pro / Lenovo Phab2 Pro	200 Hz	自然附着	25 h
IDOL	IPhone 8 / Kaarta Stencil	100 Hz	自然附着	20 h
IMUNET	Lenovo Phab2 Pro / Samsung S10	200 Hz	手持	28 h

3.2　基线模型

选取 ResNet[11]、MobNet[18]、MnasNet[19]、EffNet[20] 和 IMUNet[12]作为基线模型，与本文模型作对比分析。为确保各模型能够适配 IMU 序列处理，对每个基线模型进行了相应的修改与优化。

3.3　评估指标

经由模型输出的速度向量，通过时间戳扩展和线性插值处理后形成最终轨迹。为科学评估不同模型的

定位性能，需引入相应评价指标，首要考虑绝对轨迹误差（Absolute Trajectory Error, ATE），记作 E_{ATE}。该指标通过整体比较预测轨迹与真实轨迹的均方根误差来评估全局轨迹准确性，计算公式如下：

$$E_{\mathrm{ATE}}=\sqrt{\frac{1}{m}\sum_{t=1}^{m}\left\|\boldsymbol{x}_t-\hat{\boldsymbol{x}}_t\right\|^2} \tag{11}$$

其中，$\boldsymbol{x}_t$ 表示时间戳 t 对应的真实位置，$\hat{\boldsymbol{x}}_t$ 表示预测位置，m 为轨迹中位置点总数。

其次，相对轨迹误差（Relative Trajectory Error, RTE）表示真实位置与局部预测位置之间的位置一致性，记作 E_{RTE}。该指标通过计算对应时间间隔内真实值与预测值的均方根误差得到，其数学表达式为：

$$E_{\mathrm{RTE}}=\sqrt{\frac{1}{m}\sum_{t=1}^{m}\left\|\left(\boldsymbol{x}_{t+td}-\boldsymbol{x}_t\right)-\left(\hat{\boldsymbol{x}}_{t+td}-\hat{\boldsymbol{x}}_t\right)\right\|^2} \tag{12}$$

其中，td 表示对应的时间间隔。

3.4 模型性能分析

相较于 Xception 模型，本文提出的模型更适合处理 IMU 数据的初步提取层和深层次提取层。为验证这些改进的合理性与有效性，采用 RONIN 数据集进行评估与验证。从表 5 可以看到，感受野为 7 个时间点（RF: 7）的卷积能更有效提取 IMU 数据特征，而 5 个时间点（RF: 5）和 9 个时间点（RF: 9）的感受野因对 IMU 数据特征提取过于简单或冗余，对系统性能提升有限。

表 5 不同卷积核感受野的轨迹误差（单位：米）

Tab.5 Trajectory error of convolutional kernels with different receptive fields (Unit: m)

测试对象	评估指标	RF:5	RF:7	RF:9
已知	ATE	4.16	3.38	4.63
	RTE	2.83	2.66	2.74
未知	ATE	5.85	5.35	5.39
	RTE	4.63	4.48	4.49

表 6 不同卷积层数的轨迹误差（单位：米）

Tab.6 Trajectory error with different convolutional layers (Unit: m)

测试对象	评估指标	卷积层数		
		2 层	4 层	8 层
已知	ATE	3.75	3.38	3.56
	RTE	2.77	2.66	2.67
未知	ATE	5.42	5.35	5.91
	RTE	4.52	4.48	4.48

对深层次提取层的优化聚焦于深度可分离卷积的层数配置。本文模型采用四层深度可分离卷积，而 Xception 模型使用八层。实验结果表明（表 6），四层结构效果最佳。层数不足会导致特征提取不充分，层数过多则易造成信息分散难以整合，因此合适的层数有助于捕捉有效特征。

在深层次提取层中，Transformer 的参数配置同样关键，直接影响模型最终性能。其参数包括多头注意力机制的头数、前馈神经网络维度、层数以及 Dropout 率，其中头数、前馈神经网络维度和 Dropout 率均相对固定[16]。然而，层数的设定需依据数据量的规模灵活调整，因为增加层数必然伴随着训练参数的增多。表 7 的实验结果表明，层数并非越多越好。实际上，过多的层数往往导致提取到的特征信息过于离散，不仅无法提升模型性能，还会增加不必要的训练参数。相反，仅使用一层 Transformer 在达到最佳效果和最小轨迹误差的同时，还能保持最少的训练参数量，实现效率与性能的双重优化。

表 7 不同 Transformer 层数的轨迹误差（单位：米）

Tab.7 Trajectory error of different Transformer layers (Unit: m)

测试对象	评估指标	Transformer 层数		
		1 层	2 层	3 层
已知	ATE	3.38	3.59	3.72
	RTE	2.66	2.71	2.72
未知	ATE	5.35	5.62	5.51
	RTE	4.48	4.55	4.59

3.5 不同模型的性能对比分析

为全面评估各模型在不同数据集上的性能表现，表 8 给出不同模型在已知与未知测试集上的 ATE 和 RTE。在规模最大的 RONIN 数据集上，本文模型（下文图表简记为 proposed）在已知测试集上的 ATE 为 3.38 m，相较于 ResNet、MobNet、MnasNet、EffNet、IMUNet 分别减小 6.89%、17.16%、10.58%、7.65%和 10.11%；RTE 相较于基线模型也有所减小。在未知测试集上，除 MnasNet 外，本文模型的 ATE 分别减小 5.31%、13.15%、5.81%和 12.44%。在规模最小的 IDOL 数据集上，本文模型在已知测试集的 ATE 相较于上述五种主流模型分别减少 13.95%、18.05%、15.94%、28.29%和 4.66%，同时在 RTE 上也保持最低误差。在未知测试集上，其 ATE 亦分别降低 1.77%、22.96%、2.85%、14.15%和 7.52%。尽管 MnasNet 模型在 RONIN 未知测试集上 ATE 最小，但在其它数据集上的误差却相对较大。原因在于该模型的自动搜索策略可以从未知测试集中学习新的特征；同时，RONIN 数据集提供了更丰富的训练样本，使其能够挖潜更多潜在信息从而进行更广泛的训练。然而，对于其他特征稀疏的数据集，搜索特征较少导致其性能下降。相比之下，本文模型在绝大多数数据集上均实现了最小的轨迹误差，这主要归功于 Transformer 结构的设计：该结构能够赋予整个序列的时间周期性特征足够的权重，并通过深

层次特征提取进一步挖掘 IMU 序列的空间特征，使其在未知测试集上的 ATE 和 RTE 也有所减小。

表 8　总体轨迹预测精度（单位：米）

Tab.8　Overall trajectory prediction accuracy (Unit: m)

数据集	测试对象	评估指标	ResNet	MobNet	MnasNet	EffNet	IMUNet	Proposed
RONIN	已知	ATE	3.63	4.08	3.78	3.66	3.76	3.38
		RTE	2.76	2.83	2.75	2.79	2.73	2.66
	未知	ATE	5.65	6.16	5.19	5.68	6.11	5.35
		RTE	4.57	4.75	4.54	4.60	4.72	4.48
RIDI	已知	ATE	1.56	1.60	1.56	1.49	1.36	1.30
		RTE	1.92	1.87	1.87	1.81	1.58	1.56
	未知	ATE	1.71	1.89	1.75	1.92	1.58	1.38
		RTE	1.81	1.86	1.77	1.85	1.53	1.52
IDOL	已知	ATE	3.80	3.99	3.89	4.56	3.43	3.27
		RTE	2.40	2.43	2.44	2.63	2.31	2.17
	未知	ATE	4.51	5.75	4.56	5.16	4.79	4.43
		RTE	2.76	3.08	2.80	2.83	2.81	2.73
IMUNET	已知	ATE	4.65	5.02	6.54	4.33	4.32	4.25
		RTE	3.72	4.25	4.99	3.67	3.61	3.46

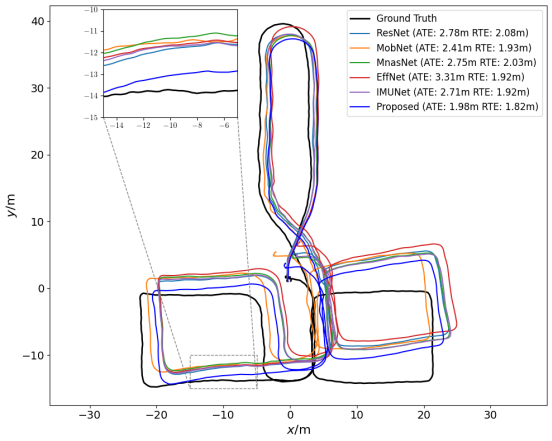

图 3　六种模型在 RONIN 已知测试集上的轨迹对比（轨迹长度为 310 m）

Fig.3　Trajectory comparison of six models from the seen RONIN test set (trajectory length 310 m)

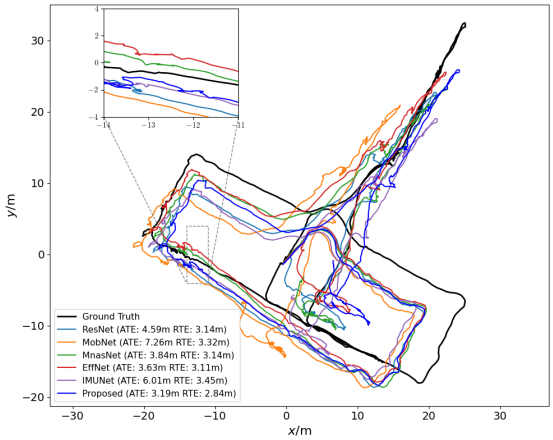

图 4　六种模型在 RONIN 未知测试集上的轨迹对比（轨迹长度为 705 m）

Fig.4　Trajectory comparison of six models from the unseen RONIN test set (trajectory length 705 m)

为直观评估不同定位模型的性能，本文在每个数据集中均选取一个具有代表性的序列，并通过图形化方法将各模型的轨迹进行对比展示。对于规模最大的 RONIN 数据集，图 3 和图 4 分别展示了各模型在已知和未知测试集上的轨迹对比结果：在已知测试集中，本文模型生成的轨迹与真实轨迹最为接近，并在一定时间内能有效抑制 IMU 误差累积；在未知测试集中，该模型仍保持最小轨迹误差，充分验证了其出色的泛化能力。

在 RIDI 数据集上，图 5 和图 6 分别展示了本文模型在 134 m 已知路径和 173 m 未知路径中的轨迹生成效果，均与真实轨迹保持高度吻合。IDOL 数据集的图 7 和图 8 进一步验证了该模型在 467 m 已知路径和 736 m 未知路径中的优异表现。此外，IMUNET 数据集的图 9 对比结果表明，本文模型生成的轨迹仍最接近真实轨迹。

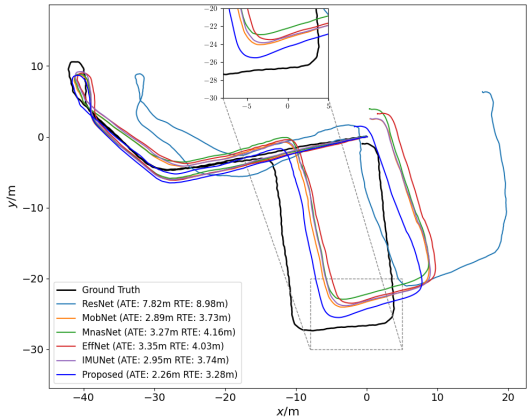

图 5　六种模型在 RIDI 已知测试集上的轨迹对比（轨迹长度为 134 m）

Fig.5　Trajectory comparison of six models from the seen RIDI test set (trajectory length 134 m)

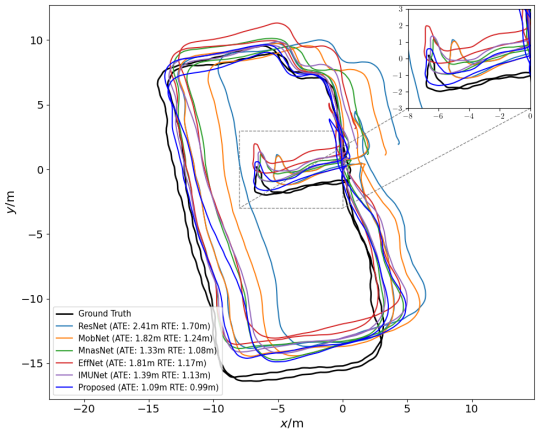

图 6 六种模型在 RIDI 未知测试集上的轨迹对比（轨迹长度为 173 m）

Fig.6 Trajectory comparison of six models from the unseen RIDI test set (trajectory length 173 m)

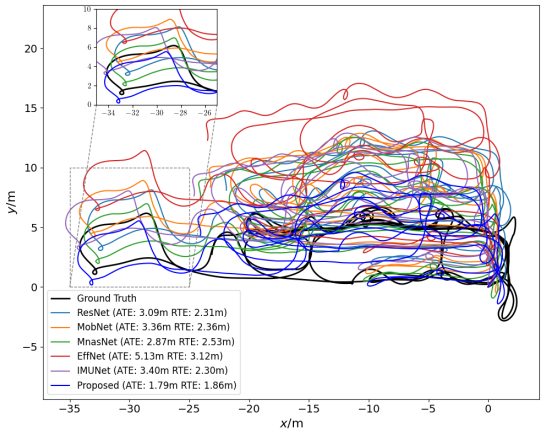

图 7 六种模型在 IDOL 已知测试集上的轨迹对比（轨迹长度为 467 m）

Fig.7 Trajectory comparison of six models from the seen IDOL test set (trajectory length 467 m)

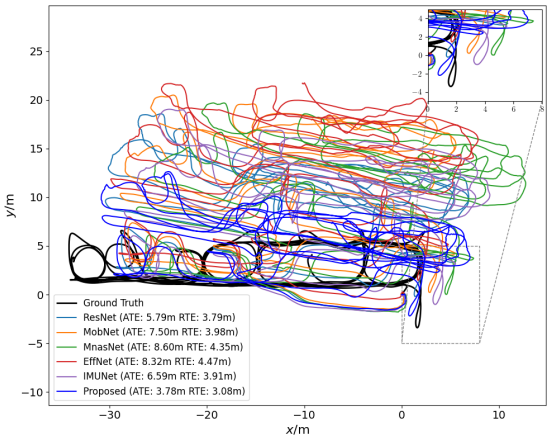

图 8 六种模型在 IDOL 未知测试集上的轨迹对比（轨迹长度为 736 m）

Fig.8 Trajectory comparison of six models from the unseen IDOL test set (trajectory length 736 m)

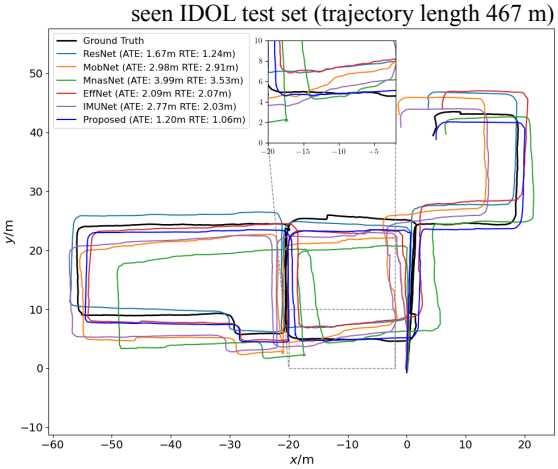

图 9 六种模型在 IMUNET 数据集上的轨迹对比（轨迹长度为 314 m）

Fig.9 Trajectory comparison of six models from the IMUNet dataset (trajectory length 314 m)

4 结 论

本文分析了行人行走过程中 IMU 序列所展现的独特特征，采用 Xception 模型设计了一种深层次捕获时空特征的模型。模型架构包含三个核心模块：1）初步提取层对 IMU 数据进行压缩和降维以提取基础特征；2）深层次提取层对提取到的特征进行更精炼的加工和提取，同时结合 Transformer 弥补模型对全局时间周期性特征捕获的缺失；3）速度回归层将数据压缩为速度向量并重构为轨迹以此来提高速度的预测精度。实验结果表明，该模型在四个公开 IMU 数据集上均能有效抑制误差累积，保持优异的预测精度和泛化性能。虽然本文模型训练的参数量有所增加，但实现了轨迹误差的最小化。未来将会对模型不断进行优化，在保证相应轨迹误差最小的同时不断减少训练参数以实现更广泛的应用。

参考文献（References）：

[1] Hoang Q H, Kim G. IMU augment tightly coupled LiDAR-visual-inertial odometry for agricultural environments[J]. IEEE Robotics and Automation Letters, 2024, 9(10): 8483-8490.

[2] Niu Z, Cong L, Qin H, et al. Pedestrian dead reckoning based on complex motion mode recognition using hierarchical classification[J]. IEEE Sensors Journal, 2024, 24(4): 4935-4947.

[3] Wdavid G, Helen B. Application of ZUPT algorithm in mobile robotics for precise localization[J]. Robotics and Autonomous Systems, 2024, 123(4): 567-582.

[4] Wang X, Gao F, Huang J, et al. UWB/LiDAR tightly coupled positioning algorithm based on ISSA optimized particle filter[J]. IEEE Sensors Journal, 2024, 24(7): 11217-11228.

[5] 双丰，马翰林，杨杰，等. 基于改进 EKF_LOAM 的电缆沟巡检机器人精准定位策略[J]. 中国惯性技术学报，2024, 32(4): 326-335.
Shuang F, Ma H, Yang J, et al. The precise positioning strategy of cable trench inspection robot based on improved EKF_LOAM[J]. Journal of Chinese Inertial Technology, 2024, 32(4): 326-335.

（下转第 971 页）

scale-invariant keypoints[J]. International Journal of Computer Vision, 2004,60(2):91-110.
[4] Tian Y, Chen C, Shah M. Cross-view image matching for geo-localization in urban environments[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 1998-2006.
[5] 王小攀，李建胜，王安成，等. 面向无人机绝对定位的遥感影像快速检索方法[J]. 中国惯性技术学报，2024, 32(04): 363-370+378.
Wang X, Li J, Wang A, et al. Fast retrieval method of remote sensing image for UAV absolute location[J]. Journal of Chinese Inertial Technology, 2024, 32(04): 363-370+378.
[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] Dai M, Hu J H, Zhuang J D, et al. A transformer-based feature segmentation and region alignment method for UAV-view geo-localization[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(7): 4376-4389.
[8] Yang H, Lu X, Zhu Y. Cross-view geo-localization with evolving transformer[J]. arxiv preprint arxiv:2107.00842, 2021.
[9] Dai M, Zheng E, Feng Z, et al. Vision-based UAV self-positioning in low-altitude urban environments[J]. IEEE Transactions on Image Processing, 2023, 33: 493-508.
[10] Zhang K, Qi S, Cai J, et al. Content-based image retrieval with a convolutional siamese neural network: Distinguishing lung cancer and tuberculosis in CT images[J]. Computers in biology and medicine, 2022, 140: 105096.
[11] Yuan Z, Zhang H, Lu P, et al. Ditfastattn: Attention compression for diffusion transformer models[J]. arXiv preprint arXiv:2406.08552, 2024.
[12] Wang P, Wang X, Wang F, et al. Kvt: k-nn attention for boosting vision transformers[C]//European conference on computer vision. Cham: Springer Nature Switzerland, 2022: 285-302.
[13] Fang S, Li K, Li Z. Salient positions based attention network for image classification[J]. arxiv preprint arxiv:2106.04996, 2021.
[14] Gao T, Li Z, Wen Y, et al. Attention-free global multiscale fusion network for remote sensing object detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 62: 5603214.
[15] Huo J. A study of spatial attention and squeeze excitation block fusion improved resnet for identifying bank notes[J]. Security and Communication Networks, 2021: 1-8.
[16] Touvron H, Cord M, Douze M, et al. Training data-efficient image transformers & distillation through attention[C]//International conference on machine learning. PMLR, 2021: 10347-10357.
[17] Howard J, Ruder S. Fine-tuned language models for text classification[J]. arxiv preprint arxiv:1801.06146, 2018.
[18] Zhuang J, Dai M, Chen X, et al. A faster and more effective cross-view matching method of UAV and satellite images for UAV geolocalization[J]. Remote Sensing, 2021, 13(19): 3979.
[19] Wang T, Zheng Z, Yan C, et al. Each part matters: Local patterns facilitate cross-view geo-localization[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 32(2): 867-879.
[20] Chen Q, Wang T, Yang Z, et al. SDPL: Shifting-dense partition learning for UAV-view geo-localization[J]. arXiv preprint arXiv: 2403.04172, 2024.

（上接第 962 页）

[6] 杨秀建，皇甫尚昆，颜绍祥. 基于改进 UKF 的 UWB/IMU/里程计融合定位方法[J]. 中国惯性技术学报，2023, 31(5): 462-471.
Yang X, Huangfu S, Yan S. Fusion positioning method with UWB/IMU/odometer based on the improved UKF[J].
Journal of Chinese Inertial Technology, 2023, 31(5): 462-471.
[7] 韩勇强，于潇颖，纪泽源，等. 面向城市复杂环境的 GNSS/INS 高精度图优化算法[J]. 中国惯性技术学报，2022, 30(5): 582-588.
Han Y, Yu X, Ji Z, et al. The high-precision factor graph optimization algorithm of GNSS/INS for urban complex environment[J]. Journal of Chinese Inertial Technology, 2022,30(5): 582-588.
[8] Mi J, Wang Q, Liu P, et al. A performance enhancement method for redundant IMU based on neural network and geometric constraint[J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 1-11.
[9] Yan H, Shan Q, Furukawa Y. RIDI: Robust IMU double integration[C]//15th European Conference on Computer Vision (ECCV). 2018: 621-636.
[10] Herath S, Yan H, Furukawa Y. RONIN: Robust neural inertial navigation in the wild: Benchmark, evaluations & new methods[C]//2020 IEEE International Conference on Robotics and Automation (ICRA). France, 2020: 3146-3152.
[11] Chen B, Zhang R, Wang S, et al. Deep-learning-based inertial odometry for pedestrian tracking using attention mechanism and Res2NET module[J]. IEEE Sensors Letters, 2022, 6(11): 1-4.
[12] Zeinali B, Zanddizari H, Chang M J. IMUNet: Efficient regression architecture for inertial IMU navigation and positioning[J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 1-13.
[13] Wang Y, Cheng H, Meng MQH. Inertial odometry using hybrid neural network with temporal attention for pedestrian localization[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1-10.
[14] Chollet F. Xception: Deep learning with depthwise separable convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 1251-1258.
[15] Ashish V, Noam S, Niki P, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems 30 (NIPS 2017), 2017: 5999-6009.
[16] Li J, Yuan G, Duan H. Adaptive Kalman filter for SINS/GPS integration system with measurement noise uncertainty[J]. IEEE Transactions on Industrial Electronics, 2022, 69(12): 13925-13935.
[17] Sun S, Melamed D, Kitani K. IDOL: Inertial deep orientation-estimation and localization[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(7): 6128-6137.
[18] Zhu Q, Zhuang H, Zhao M, et al. A study on expression recognition based on improved mobilenetV2 network[J]. Scientific Reports, 2024, 14(1): 8121.
[19] Li Y, Yu A, Meng T, et al. Deepfusion: LiDAR-camera deep fusion for multi-modal 3D object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 17182-17191.
[20] Arora L, Singh S K, Kumar S, et al. Ensemble deep learning and EfficientNet for accurate diagnosis of diabetic retinopathy[J]. Scientific Reports, 2024, 14(1): 30554.