一种大视角变换场景下的图像匹配方法

向梦丽，黄志勇，佘雅丽，丁妥君

(三峡大学计算机与信息学院，湖北 宜昌 443000)

摘　　要：针对现有图像匹配方法在大视角变换场景中匹配精度和匹配数量显著下降的问题，提出了一种改进的 E-LoFTR 图像匹配方法。首先，采用先视角调整后精细匹配策略，提出了一种新颖的双阶段 SIFT 视角矫正模块，该模块结合了尺度不变特征变换(SIFT)算法的视角不变性与单应性变换(homography)的几何对齐能力，提高了模型对大视角变换的适应能力。然后，设计了方向感知门控注意力机制，使用多方向卷积和动态门控的级联结构提取查询(Q)、键(K)、值(V)，注入的几何先验显著提升了模型的鲁棒性。最后，为了避免特征融合过程中的信息损失问题，使用 Fusion-DySample 上采样模块提升匹配性能。在公开数据集 MegaDepth 上的实验结果表明，所提出的方法在旋转误差阈值为 5°，10°和 20°下的相对位姿估计累计曲线下面积分别为 57.1%，72.7%和 83.9%，较 E-LoFTR 分别提升 0.7%，0.5%和 0.4%；在基于 MegaDepth 构建的全新数据集 NewMega 和私有工业数据集上，匹配点对数量和匹配正确率均显著提升。

关　键　词：图像匹配；E-LoFTR；大视角变换；SIFT；注意力机制

中图分类号：TP 391.41　　**DOI**：10.11996/JGj.2095-302X.2026010090

文献标识码：A　　**文 章 编 号**：2095-302X(2026)01-0090-09

An image matching method for large viewpoint variation scenarios

XIANG Mengli, HUANG Zhiyong, SHE Yali, DING Tuojun

(College of Computer and Information Technology, China Three Gorges University, Yichang Hubei 443000, China)

Abstract: To address the significant decline in matching accuracy and the number of correspondences exhibited by existing image-matching methods under large viewpoint variations, an improved image-matching approach based on E-LoFTR was proposed. Firstly, based on a strategy of viewpoint rectification followed by fine-grained matching, a novel two-stage SIFT-based viewpoint-rectification module was proposed, which leveraged the viewpoint invariance of the Scale-Invariant Feature Transform (SIFT) algorithm and the geometric alignment capability of homography to enhance matching accuracy under large viewpoint variations. Then, a directional-gated attention mechanism was designed that employed a cascaded structure of multi-directional convolutions and dynamic gating to extract queries (Q), keys (K), and values (V). The injected geometric priors significantly enhanced the model’s robustness. Lastly, to mitigate information loss during the upsampling of fused features, the Fusion-DySample module was incorporated to further improve performance. Experimental results on the public MegaDepth dataset showed that our method achieved relative pose estimation AUCs of 57.1%, 72.7%, and 83.9% under rotation error thresholds of 5°, 10°, and 20°, respectively, outperforming E-LoFTR by 0.7%, 0.5%, and 0.4%. On the newly constructed NewMega dataset based on MegaDepth and on a private industrial dataset, our method also demonstrated substantial improvements in both the number of matches and matching accuracy.

Keywords: image matching; E-LoFTR; large perspective variation; SIFT; attention mechanism

---

收稿日期：2025-06-24；**定稿日期**：2025-08-27；**通信作者**：黄志勇，E-mail：hzy@hzy.org.cn

Received：24 June，2025；**Finalized**：27 August，2025；**Corresponding author**：HUANG Zhiyong，E-mail：hzy@hzy.org.cn

基金项目：国家自然科学基金(62371271)

Foundation items：National Natural Science Foundation of China (62371271)

图像匹配是计算机视觉中的一个核心问题，旨在识别并定位 2 幅或多幅图像之间的相似区域或对象。该技术广泛应用于三维重建、即时定位与地图构建(pping, SLAM)[1]和图像拼接[2]等计算机视觉任务，同时在医学影像分析、遥感图像处理、增强现实、机器人导航[3-5]及工业自动化等多个应用领域展现出重要价值。图像匹配算法主要包括特征检测、描述符提取和特征匹配 3 个阶段。基于手工特征的传统方法以尺度不变特征变换 SIFT(Scale-Invariant Feature Transform)[6]，SURF(SURFp Robust Features)[7]，ORB(Oriented FAST and Rotated BRIEF)[8]，FAST(Features from Accelerated Segment Test)[9] 和 BRIEF(Binary Robust Independent Elementary Features)[10]为代表。其中，SIFT 算法通过高斯差分金字塔实现尺度不变性，利用主方向分配赋予旋转鲁棒性，即使是在大视角变化场景下仍能保持稳定的匹配性能。但由于传统方法采用的手工特征极度依赖梯度极值检测，在弱纹理以及光照变化强烈区域适应能力差，易导致漏检、错检。

随着卷积神经网络(Convolutional Neural Networks, CNN)不断发展，基于深度学习的图像匹配方法取得了显著进展。现有的方法按特征提取方式可分为基于检测器和无检测器 2 类匹配框架。其中，基于检测器的匹配方法以 SuperGlue[11]为代表，利用 SuperPoint[12]特征提取器提取关键点和描述子。受 SuperGlue 的启发，后续研究者们相继提出了 LightGlue[13]，ClusterGNN[14]，OmniGlue[15]和 Ada-Matcher[16]等。SUN 等[17]在 SuperGlue 的基础上提出了无检测器的图像匹配架构 LoFTR(Detector-Free Local Feature Matching with Transformers)，实现了直接在 2 幅图像之间建立半密集的特征对应关系。并通过引入 Transformer[18]，分别在粗粒度特征和细粒度特征上捕捉全局关系，从而增强特征表示。该方法在弱纹理、重复纹理和视角变化等挑战性场景下的表现显著优于基于检测器的方法。随后涌现出一系列基于 LoFTR 的改进模型，如 MatchFormer[19]，AspanFormer[20]，ASTR(Adaptive Spot-guided Transformer for Robust Local Feature Matching)[21]，Ada-LoFTR[22]以及郭印宏等[23]提出的基于重复性和特异性约束的图像特征匹配方法等。由于在全图范围内进行注意力计算效率较低，WANG 等[24]基于 LoFTR 提出了 E-LoFTR，通过聚合相邻 token 的方法减少冗余计算。此外，E-LoFTR 还提出了一种两阶段相关性细化模块，首先通过像素级匹配定位，再在小窗口内进行亚像素级细化，有效提升匹配精度。

尽管基于深度学习的图像匹配方法相比传统方法性能更佳，但在大视角变换场景中的表现仍存在明显不足。其局限性主要源于以下 3 方面因素：①大部分方法在特定数据集上进行训练，如分别在 ScanNet[25]和 MegaDepth[26]上训练室内和室外模型，而此类数据集包含的大视角变换场景较少，导致模型难以学习极端视角下的鲁棒特征表示；②由于 CNN 架构的局部感受野难以捕捉大视角变换下的远程依赖，而基于 Transformer 的方法虽支持全局建模，但注意力机制对旋转敏感；③大视角变换还会引起特征空间的非线性畸变，部分描述符匹配失效。

为解决上述问题，本文提出了一种结合传统方法与深度学习方法的新颖匹配框架，与 E-LoFTR 等现有方法相比，其核心创新如下：

1）提出了双阶段 SIFT 视角矫正模块，采用先矫正后匹配的思想，通过引入 SIFT、随机抽样一致(Random Sample Consensus, RANSAC)算法以及单应性变换，实现大视角差异图像的几何对齐，增强了匹配的鲁棒性。

2）引入了方向感知门控注意力机制，采用不同的卷积核提取水平、垂直和对角线方向的特征，并通过可学习的门控抑制无关区域噪声使模型能够关注重要区域，以提升匹配的精度。

3）设计了一个结合局部感知与全局感知的上采样模块 Fusion-DySample，通过动态权重分配机制自适应融合局部感知和全局感知，有效结合 2 种采样模式的优点，保留了局部信息和全局信息，为精匹配阶段提供高保真特征。

1　方法

给定一对图像 I^A 和 I^B，图像匹配的目标是建立一组可靠的对应关系。

采用无检测器的半密集匹配方法，总体框架如图 1 所示。首先，对输入图像进行双阶段 SIFT 视角矫正得到几何对齐的图像对 I^A 和 $\tilde{I}^B$；再将矫正后的图像输入 RepVGG[27]骨干网络提取多尺度特征，生成 3 个分辨率分别为 1/8，1/4 和 1/2 的特征图；将 1/8 粗特征图传入方向感知门控注意力机制进行特征重构，再经过粗匹配层的动态阈值筛选和 Dual-Softmax 计算，得到粗匹配点对集合 M_c；然

后，利用 Fusion-DySample 上采样策略对 1/8，1/4 和 1/2 维度的特征进行融合上采样，得到的融合特征作为精匹配阶段的输入进行细化匹配，以得到精匹配点对集合 M_f；在最终的精细匹配预测结果上应用单应性逆变换将矫正后的匹配点坐标通过映射，得到原始图像的匹配点对。

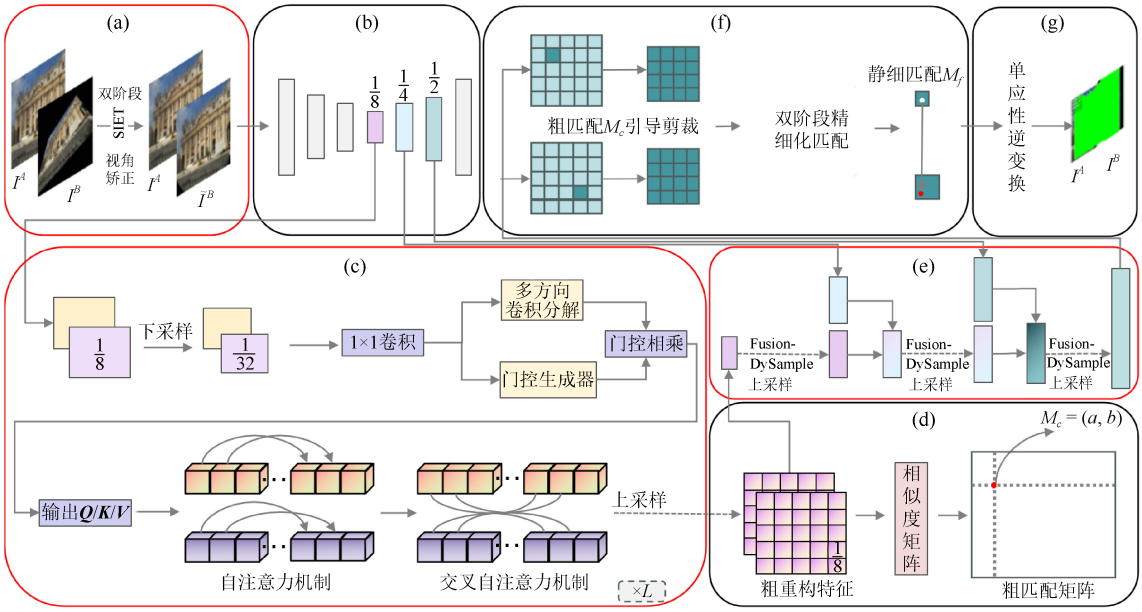

图 1 方法总体架构((a) 预处理；(b) 特征提取；(c) 特征重构；(d) 粗匹配；(e) 特征融合；(f) 精匹配；(g) 逆变换)

Fig. 1 Overall framework of the method ((a) Preprocessing; (b) Feature extraction; (c) Feature reconstruction; (d) Coarse matching; (e) Feature fusion; (f) Fine matching; (g) Inverse transformation)

1.1 双阶段 SIFT 视角矫正

为提升模型在大视角变换条件下的适应能力，本文提出了一种基于传统特征匹配方法的图像几何对齐模块：双阶段 SIFT 视角矫正。该模块作为整体模型的预处理步骤，通过 2 个阶段的 SIFT 匹配，得到可靠的对应点，应用单应性变换，实现图像之间的粗几何对齐，为后续匹配网络减轻负担。双阶段 SIFT 视角矫正过程如图 2 所示。

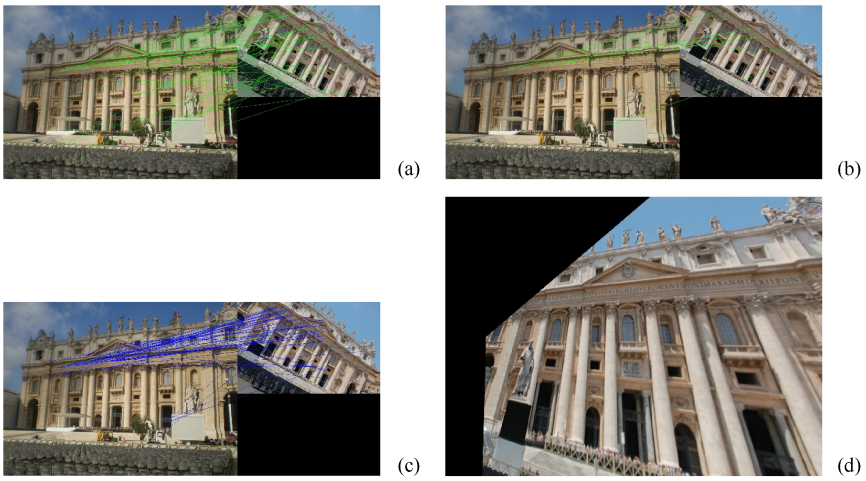

图 2 双阶段 SIFT 视角矫正过程((a) 第一阶段；(b) RANSAC 筛选；(c) 第二阶段；(d) 单应性变换)

Fig. 2 Two-stage SIFT-based viewpoint rectification Process ((a) First stage; (b) RANSAC filtering; (c) Second stage; (d) Homography warping)

1) 第一阶段 SIFT 匹配。首先全图匹配，提取待匹配图像对 I^A 和 I^B 中的特征点，并通过快速近似最近邻库(Fast Library for Approximate Nearest Neighbors, FLANN)匹配器对特征点进行匹配。然后经过 RANSAC 筛选出一组可靠匹配点对作为第二阶段 SIFT 匹配的输入。

2) 第二阶段局部区域增强 SIFT 匹配。将筛选后获得的可靠匹配点对作为中心，在一定范围内构造特征提取掩码，再次提取更精准的局部特征点后进行增量匹配。将新增的匹配点对加入输入点对作为总匹配点对集合。

3) 利用总匹配点对集合估计单应性变换矩阵 $\boldsymbol{H}$，将图 I^B 变换至图 I^A 视角，得到几何对齐的图像 $\tilde{I}^B$，即

$$\tilde{I}^B = warp\left(I^B, \boldsymbol{H}\right) \tag{1}$$

1.2 特征提取

在特征提取阶段，采用 RepVGG 作为轻量级骨干网络。给定几何对齐后的图像对 I^A 和 $\tilde{I}^B$，将其输入骨干网络进行多尺度特征提取，输出 1/8 下采样特征图作为粗特征 $feature_c^0$，1/4 和 1/2 下采样特征图作为精特征 $feature_f^0$ 和 $feature_f^1$。

1.3 特征重构

特征重构阶段主要采用方向感知门控自注意力和交叉注意力机制对粗特征 $feature_c^0$ 进行上下文信息交互，得到重构粗特征 $feature_c$。

由于直接在整个粗特征 $feature_c^0$ 上进行特征重构计算成本较大，为减少计算量，提升模型效率，本文借鉴了基准模型 E-LoFTR 的特征聚合策略，对输入的粗特征进行下采样以实现信息聚合。对图 I^A 的粗特征 $feature_A$ 采用深度卷积进行相邻令牌(Token)信息聚合，对图 I^B 的粗特征 $feature_B$ 采用最大池化层提取上下文特征信息，得到的聚合特征作为后续注意力的输入 x，即

$$feature_A = \mathrm{Conv2D}\left(feature_A\right) \tag{2}$$

$$feature_B = \mathrm{MaxPool}\left(feature_B\right) \tag{3}$$

标准注意力层的输入包含查询 $\boldsymbol{Q}$、键 $\boldsymbol{K}$ 和值 $\boldsymbol{V}$。其中，$\boldsymbol{Q}$，$\boldsymbol{K}$ 和 $\boldsymbol{V}$ 通过对输入 x 进行线性投影得到。查询 $\boldsymbol{Q}$ 的计算过程如式(4)所示，键/值同理，即

$$Q = \mathrm{Linear}\left(x\right) \tag{4}$$

而在大视角变换的应用场景中，局部特征的方向会发生显著变化，传统的线性投影难以保持特征一致性。为增强注意力机制对方向敏感特征的捕获效率，抑制无关区域的噪声影响，本文提出将线性投影层替换为多方向卷积结合门控的级联结构，在特征投影阶段显示引入空间方向感知能力。具体如下：

1）对输入的特征 x 采用标准的 1×1 卷积完成基础特征变换，得到变换后的特征为

$$x' = \mathrm{Conv}_{1\times1}\left(x\right) \tag{5}$$

2）对 x' 使用 3 个不同卷积核的深度可分离卷积分别捕捉水平、垂直和对角线方向的特征，分别用 H，V 和 D 表示，无论图像如何变化，至少有一种卷积核能有效响应原始方向模式，即

$$\begin{aligned} H &= \mathrm{Conv}_{1\times3}\left(x'\right) \\ V &= \mathrm{Conv}_{3\times1}\left(x'\right) \\ D &= \mathrm{Conv}_{3\times3}\left(x'\right) \end{aligned} \tag{6}$$

3）通过一个 1×1 的卷积以及 Sigmoid 激活函数生成动态门控权重 G，以实现自适应强化与当前视角方向最相关的方向特征，抑制无关区域的信息，即

$$G = \sigma\left(\mathrm{Conv}_{1\times1}\left(x'\right)\right) \tag{7}$$

式中：1×1 卷积的作用是隐式建模通道间的依赖关系，Sigmoid 将权重压缩到[0,1]范围，实现自适应特征选择。

4）将得到的多方向特征与门控融合输出，得到空间感知能力增强的特征作为查询 $\boldsymbol{Q}$，提高注意力机制对局部几何特征的捕获能力，如式(8)所示，键/值同理，即

$$\boldsymbol{Q} = G \odot \left(H + V + D\right) \tag{8}$$

式中：$\odot$ 表示点积。

将 $\boldsymbol{Q}$，$\boldsymbol{K}$，$\boldsymbol{V}$ 依次传入自注意力模块和交叉注意力进行计算，得到的输出 M 与输入 x 进行残差连接，经过 L 层后，最终输出 2 幅图的重构粗特征 $feature_c^A$ 和 $feature_c^B$。

1.4 粗匹配

粗匹配过程如下：

1）通过对 I^A 的重构特征 $feature_c^A$ 和 I^B 的重构特征 $feature_c^B$ 计算点积相似度得到相似度矩阵 $\boldsymbol{S}$，表示 2 图之间各个特征的相似度得分，即

$$\boldsymbol{S}\left(a,b\right) = \frac{1}{\tau}\left\langle feature_c^A\left(a\right), feature_c^B\left(b\right)\right\rangle \tag{9}$$

式中：τ 表示温度参数，通过该参数调整相似度矩阵的分布平滑度。

2）引入双软最大(Dual-Softmax)算法在相似度矩阵 $\boldsymbol{S}$ 的行和列上分别应用 Softmax 归一化后得到初步匹配概率矩阵，即

$$\boldsymbol{P}_c = \mathrm{softmax}\left(\boldsymbol{S}\left(a,\cdot\right)_b\right) \cdot \mathrm{Softmax}\left(\boldsymbol{S}\left(\cdot,b\right)_a\right) \tag{10}$$

3）使用预定义阈值 θ_c 筛除低置信度的匹配点对，并采用互最近邻(Mutual Nearest Neighbors, MNN)准则后得到更可靠的匹配点对集合，即

$$M_c=\left\{(a,b)\mid\forall(a,b)\in MNN\left(\boldsymbol{P}_c\right),\boldsymbol{P}_c(a,b)>\theta_c\right\} \quad (11)$$

1.5 特征融合与精匹配

首先对粗特征 $feature_c$ 进行双线性上采样，依次与 1/4 尺度和 1/2 尺度的细特征融合，最终输出全分辨率特征 $feature_f$。精匹配阶段利用粗匹配得到的 M_c 剪裁固定网格，经像素-亚像素两级优化得到最终匹配 M_f。

其中，融合网络采用的传统双线性插值上采样方法虽然计算开销相对较小，但在处理复杂场景时由于难以捕捉多尺度和多风格的特征信息导致匹配精度下降。因此，本文方法基于 LIU 等[28]提出的基于深度学习的轻量化上采样方法，设计了一个融合动态采样与多风格特征的混合上采样器 Fusion- DySample，如图 3 所示。

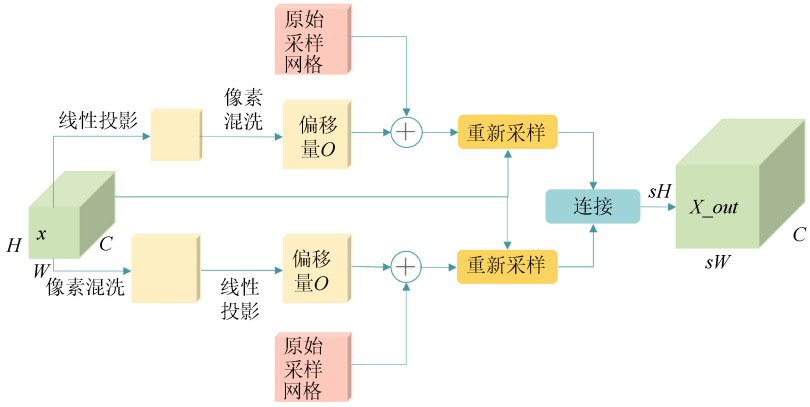

图 3 Fusion-DySample 结构图

Fig. 3 Architecure of Fusion-DySample

通过引入可学习的偏移量和采样范围，动态调整上采样过程中的采样位置，避免了复杂的卷积网络带来的计算复杂度，且超越了传统上采样方法。该采样方法结合了局部感知 LP(Linear + Pixel Shuffle)和全局感知 PL(Pixel Shuffle + Linear)2 种上采样策略，并通过自适应权重模块融合其输出，从而在保证计算效率的同时，提高匹配精度和鲁棒性。其中，局部感知通过优先对输入特征进行多通道线性投影生成偏移量，再进行像素混洗分配到空间位置，以增强细节表达能力；而全局感知则先通过像素混洗分解特征，再通过线性投影生成轻量偏移量，以降低计算量和参数量。Fusion-DySample 通过在复杂场景中灵活适配不同特征模式，兼顾了局部细节捕捉和全局信息保留，提升精匹配阶段的特征表达能力。

1.6 逆变换

由于双阶段 SIFT 视角矫正的引入，图 I^B 通过单应性变换矫正为 $\tilde{I}^B$，导致后续的流程在图像对 I^A 和 $\tilde{I}^B$ 上进行匹配。故在得到最终精细匹配预测结果 M_f 后，需要将矫正后的图像的匹配点集 M_f ($\tilde{I}^B$)通过单应性逆变换映射回矫正前的图像上 I^B，如式(12)所示。最终得到输入图像对 I^A 和 I^B 的大量高精度匹配特征点，即

$$M_f\left(I^B\right)=warp\left(M_f\left(\tilde{I}^B\right),\boldsymbol{H}^{-1}\right) \quad (12)$$

2 实验

实验所使用的训练数据集为公开可获得的 MegaDepth 数据集，其中包含 196 个场景共 100 万张图像。实验过程中，双阶段 SIFT 视角矫正模块采用的局部增强区域半径设置为 80，RANSAC 阈值为 5.0，自注意力和交叉注意力的循环迭代次数 L=4，使用 AdamW 优化器，初始学习率为 0.001，共训练 30 个 Epoch，Batch 大小设置为 1。

实验在 16 GB 显存的 RTX 4070 Ti GPU 和 32 GB 内存组成的计算平台上进行。软件平台为 PyCharm，深度学习框架为 PyTorch。

2.1 单应性估计

本实验在数据集 HPatches 上进行单应性估计实验。该数据集包含 52 个光照变化序列和 56 个视角变化序列，每个序列中包含 6 张图像和 5 个单应性变换矩阵。其中，第 1 张图像作为参考图像，分别与其余 5 张图像配对构成 5 对匹配对。

遵循 E-LoFTR 的方案，本实验将所有图像的短边统一缩放至 480 像素，对于所有对比方法都使用 OpenCV 中的 RANSAC 算法作为鲁棒单应性估计器。在评估过程中，将采用角点在变换前后的重投影误差作为评估指标，报告在 3 像素、5 像素和 10 像素阈值下角误差的累积分布函数曲线下的面

积(Area Under The Curve, AUC)。

由表 1 可知，本文方法在不同像素阈值下的单应性估计 AUC 指标均优于基线方法，由于视角矫正的引入，使模型能够适应大视角变换；以及方向感知门控注意力机制和 Fusion-DySample 上采样模块均能捕获到更多的有效空间信息，使得模型在各种场景中的表现更为稳定，显示出良好的几何对齐能力。

表 1　单应性估计结果对比/%

Table 1　Comparison of homography estimation results/%

方法	@3px	@5px	@10px
DISK+NN	52.3	64.9	78.9
SP+SuperGlue	53.9	68.3	81.7
SP+LightGlue	54.5	68.7	82.1
LoFTR	65.9	75.6	84.6
MatchFormer	66.2	76.1	85.6
E-LoFTR	66.5	76.4	85.5
Ours	**67.0**	**77.1**	**86.2**

注：加粗数据表示最优值。

2.2　基于 NewMega 的图像匹配

本文实验基于 MegaDepth 测试数据集选取最具代表性的图像构建了一个包含旋转变化与单应性扰动的全新数据集 NewMega，旨在模拟真实世界中因拍摄角度、视角差异和透视畸变等引起的非刚性匹配挑战。其中包含 1 张源图像和 5 张变换后的图像组成了 5 组图像对，如图 4 所示。

首先对图像进行 60°和 90°旋转处理用于构建子集 NewMega(r = 60)和 NewMega(r = 90)。其次引入单应性变换(Homography)，通过随机单应性矩阵对原始图像进行变换生成目标图像，单应性矩阵由随机偏移 4 个角点的方式构建，其扰动变换强度控制匹配的难度，生成的图像匹配对用于构建子集 NewMega (H-easy)和 NewMega (H-hard)。为实现匹配性能的定量评估，采用 3 种评价指标：正确匹配点对数量 m、总匹配点对数量 p 和匹配成功率(Match Success Rate, MSR)。其中，正确匹配点对数量表示满足一定误差阈值的匹配点对数量，即

$$m = \sqrt{\left(x_i - \hat{x}_i\right)^2 + \left(y_i - \hat{y}_i\right)^2} \leqslant \delta \tag{13}$$

式中：$\left(x_i, y_i\right)$表示预测的匹配点对；$\left(\hat{x}_i, \hat{y}_i\right)$表示真实对应点对；$\delta$ 表示误差阈值，设置为 1 px。

匹配成功率表示在所有检测出的匹配点对中被正确匹配的点对比例，即

$$MSR = \frac{m}{p} \tag{14}$$

(a)　(b)　(c)　(d)　(e)

图 4　NewMega 示例图((a) 原始图像；(b) 旋转 60°后的图像；(c) 旋转 90°后的图像；(d) 弱单应性变换图像；(e) 强单应性变换图像)

Fig. 4　Example images from NewMega dataset ((a) Original image; (b) Image rotated by 60°; (c) Image rotated by 90°; (d) Image with weak homography transformation; (e) Image with strong homography transformation)

不同匹配算法在 NewMega (r = 60)，NewMega (r = 90)，NewMega (H-easy)和 NewMega (H-hard)上的实验结果见表 2，可视化结果如图 5 所示，其中绿色连线表示正确匹配，红色连线表示错误匹配。方法性能由 3 个指标联合评估，两者越高，表示方法性能越好。在旋转变换场景下，传统的 SIFT 匹配算法依赖局部方向直方图来抵抗旋转干扰，对图像旋转变换具备较好的适应能力，但特征提取主要集中在纹理丰富的区域，对于天空这类弱纹理区域无法提取到有效的特征点。而深度学习算法 LightGlue 和 E-LoFTR 在大角度旋转变换条件下几乎完全失效，正确匹配点对数量趋近于零。在单应性扰动场景下，LightGlue 对单应性变换有一定的适应能力，但仍不如 SIFT 和 E-LoFTR；其中 E-LoFTR 凭借其密集匹配的优势，匹配点对数量大幅提升。

本文方法结合了 SIFT 和 E-LoFTR 的优势，通过在预处理阶段加入双阶段 SIFT 视角矫正模块，显著提升了算法对旋转变换和单应性变换场景的适应能力。此外，在特征重构阶段中，方向感知门控注意力的引入增加了模型对空间的感知能力；特征融合阶段的 Fusion-DySample 上采样模块进一步

减少特征还原过程中的信息损失。综上所述，本文方法在匹配点对数量、匹配精度以及弱纹理区域的特征提取能力均有显著地提升，验证了所提的改进方法在各种场景中的有效性。

表 2　NewMega 上的图像匹配结果对比/(m/p/MSR)

Table 2　Image matching results on NewMega dataset/(m/p/MSR)

方法	NewMega (r = 60)			NewMega (r =90)			NewMega (H-easy)			NewMega (H-hard)		
	m	p	MSR/%	m	p	MSR/%	m	p	MSR/%	m	p	MSR/%
SIFT	5 995	6 340	94.6	7 446	7 643	97.4	2 558	2 877	88.9	571	759	75.2
LightGlue	0	15	0	0	14	0	790	1 449	54.5	649	1 207	53.7
E-LoFTR	0	0	0	0	2	0	5 602	6 542	85.6	1 755	3 063	57.2
Ours	**13 310**	**13 366**	**99.6**	**12 404**	**12 515**	**99.1**	**16 330**	**16 341**	**99.9**	**14 531**	**15 879**	**91.5**

注：加粗数据表示最优值。

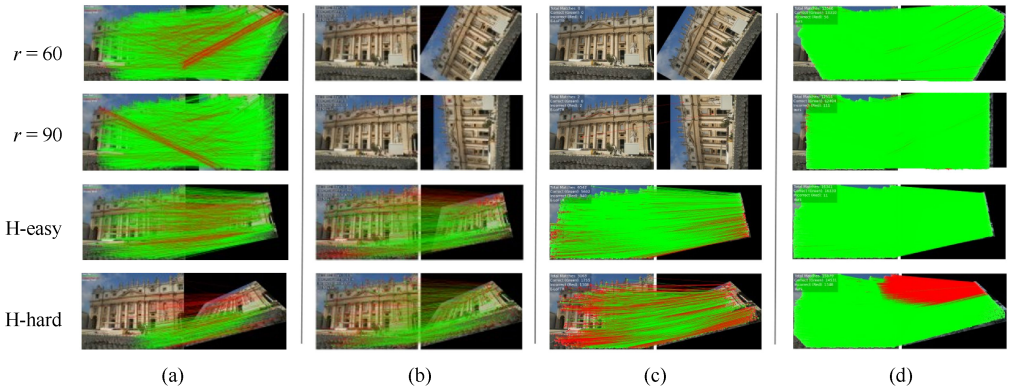

图 5　NewMega 上的图像匹配可视化对比结果

Fig. 5　Qualitative comparison of image matching on NewMega dataset ((a) SIFT; (b) LightGlue; (c) E-LoFTR; (d) Ours)

2.3　基于工业场景的图像匹配

为验证本方法在工业场景下的泛化能力，在私有汽车零件数据集上进行图像匹配实验。

该实验不仅面临大视角变换的挑战，而且还包含汽车零件结构复杂(大量大小不一的螺丝孔洞)、表面纹理稀少(大面积光滑金属表面)以及反射光强烈(明显高光区域)等困难条件。参照上一节的实验方法和设置构建 OurData (r = 60)，OurData (r = 90)，OurData (H-easy)和 OurData (H-hard)，示例图像如图 6 所示。不同匹配算法在 OurData (r = 60)，OurData (r = 90)，OurData (H-easy)和 OurData (H-hard)上的实验结果见表3，可视化结果如图7所示，其中绿色连线表示正确匹配，红色连线表示错误匹配。与在 NewMega 数据集上的图像匹配实验相比，本文方法在工业场景下的匹配点对数量 m 和匹配成功率 MSR 指标虽有所下降，但整体性能仍远超其他 3 种对比方法。在旋转角度为 90°的场景中，MSR 比 SIFT 低 5.5%，但正确点对数量比 SIFT 多出 6 615。与 E-LoFTR 相比，正确点对数量提升最高达 7 911，MSR 提升最高达 89.8%。综上，此实验成功验证了本方法在工业场景下的泛化能力。

图 6　OurData 示例图((a) 原始图像；(b) 旋转 60°后的图像；(c) 旋转 90°后的图像；(d) 弱单应性变换图像；(e) 强单应性变换图像)

Fig. 6　Example images from OurData dataset ((a) Original image; (b) Image rotated by 60°; (c) Image rotated by 90°; (d) Image with weak homography transformation; (e) Image with strong homography transformation)

表 3　OurData 上的图像匹配结果对比

Table 3　Image matching results on OurData dataset

方法	OurData (r=60)			OurData (r=90)			OurData (H-easy)			OurData (H-hard)		
	m	p	MSR/%	m	p	MSR/%	m	p	MSR/%	m	p	MSR/%
SIFT	856	959	89.3	1 305	1 349	**96.7**	329	503	65.4	117	251	46.6
LightGlue	1	20	5.0	0	35	0	485	789	61.4	373	691	53.9
E-LoFTR	6	71	8.5	9	647	1.4	3 352	4 107	81.6	1 432	2 847	50.3
Ours	**7 843**	**8 640**	**90.8**	**7 920**	**8 682**	91.2	**9 575**	**10 093**	**94.9**	**9 286**	**9 623**	**96.5**

注：加粗数据表示最优值。

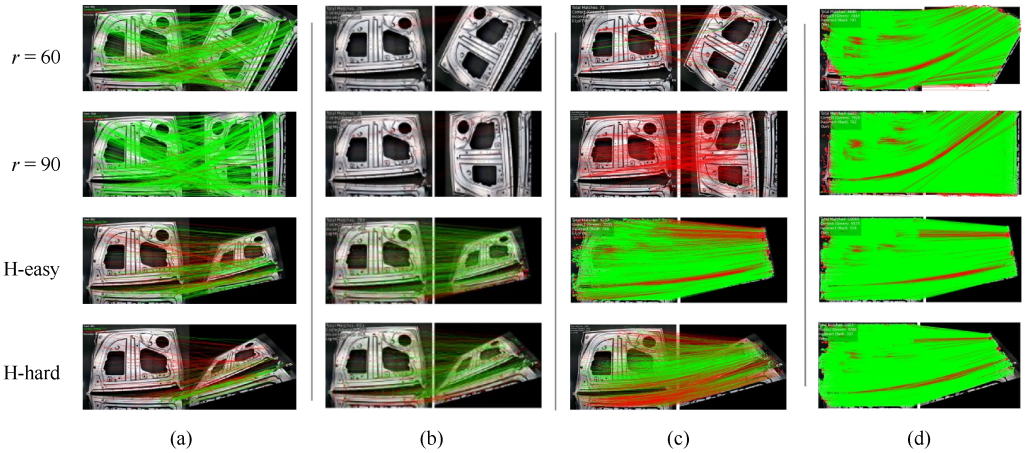

图 7　OurData 上的图像匹配可视化对比结果

Fig. 7　Qualitative comparison of image matching on OurData dataset ((a) SIFT; (b) LightGlue; (c) E-LoFTR; (d) Ours)

2.4　消融实验

为验证本方法的有效性，特别是双阶段 SIFT 视角矫正模块、方向感知门控注意力模块以及 Fusion-DySample 上采样模块在整体架构中的性能贡献。在公开数据集 MegaDepth 的小样本数据集(包含 1 500 个场景)上进行消融实验。通过匹配过程恢复的相对位姿精度作为衡量指标，其中位姿误差定义为旋转和平移误差中的最大角度误差，分别计算在 5°，10°和 20°阈值下的相对位姿估计 AUC 值。消融结果见表 4。

表 4　消融实验

Table 4　Ablation study

方法	@5°	@10°	@20°
基础模型 E-LoFTR	56.4	72.2	83.5
A	56.9	72.5	83.8
B	56.7	72.4	83.7
C	56.6	72.4	83.6
D	**57.1**	**72.7**	**83.9**

注：A 表示在 E-LoFTR 中加入双阶段 SIFT 视角矫正模块；B 表示将 E-LoFTR 中的标准注意力替换为方向感知门控注意力模块；C 表示将 E-LoFTR 中的双线性插值上采样替换为 Fusion-DySample 上采样模块；D 表示完整模型(A+B+C)；加粗数据表示最优值。

为验证双阶段 SIFT 视角矫正模块的有效性，在基础模型中加入该模块作为预处理步骤，AUC 值分别提升了 0.5%，0.3%和 0.3%，实验结果表明，本文提出的双阶段 SIFT 视角矫正在大视角变换场景中能通过视角对齐提高匹配精度。为验证方向感知门控注意力模块的有效性，在基础模型的特征重构阶段将标准注意力替换为此模块，实验结果表明，该模块增强了特征表征能力，AUC 值分别提升 0.3%，0.2%和 0.2%。为验证 Fusion-DySample 上采样模块的有效性，在基础模型的特征融合阶段将原有的双线性插值上采样替换为此模块，AUC 值分别提升 0.2%，0.2%和 0.1%，结果表明该上采样方法能够更准确地还原特征信息，提升细匹配阶段的精度。最后，同时引入提出的 3 个模块作为完整框架进行实验，在 5°，10°和 20°阈值下计算相对位姿估计的 AUC 值分别为 57.1%，72.7%和 83.9%，相比基线 E-LoFTR，分别提升了 0.7%，0.5%和 0.4%，实现了最佳性能，验证了所提出的方法的有效性。

3　结论

针对现有匹配方法在大视角变换场景下存在匹配点对数量不足、匹配正确率较低的问题，基于匹配网络 E-LoFTR 进行改进，提出一种大视角变换场景下的图像匹配方法。首先，采用双阶段 SIFT

视角矫正模块对图像对进行几何对齐，该模块极大程度提升了大视角匹配的准确性；其次，设计的方向感知门控注意力机制增强了模型对空间关键信息的提取能力；最后，Fusion-DySample 上采样模块保证了特征上采样过程中的信息完整性。本文方法在多个数据集上均展现出优越性能，匹配点对数量和匹配正确率显著提升。然而，由于引入传统 SIFT 算法，整体效率有所下降。未来工作将致力于提升方法的计算效率，以满足实际应用的实时性需求。

参考文献 (References)

[1] MUR-ARTAL R, TARDÓS J D. ORB-SLAM2: an open-source SLAM system for monocular, stereo, and RGB-D cameras[J]. IEEE Transactions on Robotics, 2017, 33(5): 1255-1262.

[2] BROWN M, LOWE D G. Automatic panoramic image stitching using invariant features[J]. International Journal of Computer Vision, 2007, 74(1): 59-73.

[3] CADENA C, CARLONE L, CARRILLO H, et al. Past, present, and future of simultaneous localization and mapping: toward the robust-perception age[J]. IEEE Transactions on Robotics, 2016, 32(6): 1309-1332.

[4] 石虹, 徐伟, 刘少清. 基于 LW-LoFTR 的增强现实三维注册算法[J]. 西安工程大学学报, 2025, 39(2): 75-83.
SHI H, XU W, LIU S Q. Augmented reality 3D registration algorithm based on LW-LoFTR[J]. Journal of Xi ' an Polytechnic University, 2025, 39(2): 75-83 (in Chinese).

[5] 舒军, 王江舸, 杨莉, 等. 改进 R-LoFTR++的智能巡检特征匹配算法[J]. 重庆理工大学学报(自然科学), 2025, 39(2): 86-96.
SHU J, WANG J G, YANG L, et al. Intelligent inspection feature matching algorithm of R-LoFTR++[J]. Journal of Chongqing University of Technology (Natural Science), 2025, 39(2): 86-96 (in Chinese).

[6] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.

[7] BAY H, ESS A, TUYTELAARS T, et al. Speeded-up robust features (SURF)[J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359.

[8] RUBLEE E, RABAUD V, KONOLIGE K, et al. ORB: an efficient alternative to SIFT or SURF[C]//2011 International Conference on Computer Vision. New York: IEEE Press, 2011: 2564-2571.

[9] ROSTEN E, DRUMMOND T. Machine learning for high-speed corner detection[C]//The 9th European Conference on Computer Vision. Cham: Springer, 2006: 430-443.

[10] CALONDER M, LEPETIT V, STRECHA C, et al. BRIEF: binary robust independent elementary features[C]//The 11th European Conference on Computer Vision. Cham: Springer, 2010: 778-792.

[11] SARLIN P E, DETONE D, MALISIEWICZ T, et al. SuperGlue: learning feature matching with graph neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2020: 4937-4946.

[12] DETONE D, MALISIEWICZ T, RABINOVICH A. SuperPoint: self-supervised interest point detection and description[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. New York: IEEE Press, 2018: 337-349.

[13] LINDENBERGER P, SARLIN P E, POLLEFEYS M. LightGlue: local feature matching at light speed[C]//2023 IEEE/CVF International Conference on Computer Vision. New York: IEEE Press, 2023: 17581-17592.

[14] SHI Y, CAI J X, SHAVIT Y, et al. ClusterGNN: cluster-based coarse-to-fine graph neural network for efficient feature matching[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2022: 12507-12516.

[15] JIANG H W, KARPUR A, CAO B Y, et al. OmniGlue: generalizable feature matching with foundation model guidance[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2024: 19865-19875.

[16] ZHENG F J, CAO C Q, ZHANG Z Y, et al. Ada-Matcher: a deep detector-based local feature matcher with adaptive weight sharing[J]. Knowledge-Based Systems, 2025, 316: 113350.

[17] SUN J M, SHEN Z H, WANG Y, et al. LoFTR: detector-free local feature matching with transformers[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2021: 8918-8927.

[18] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//The 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 6000-6010.

[19] WANG Q, ZHANG J M, YANG K L, et al. MatchFormer: interleaving attention in transformers for feature matching[C]//The 16th Asian Conference on Computer Vision. Cham: Springer, 2023: 256-273.

[20] CHEN H K, LUO Z X, ZHOU L, et al. Aspanformer: detector-free image matching with adaptive span transformer[C]//The 17th European Conference on Computer Vision. Cham: Springer, 2022: 20-36.

[21] YU J H, CHANG J H, HE J F, et al. Adaptive spot-guided transformer for consistent local feature matching[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2023: 21898-21908.

[22] 张震宇, 杨小冈, 卢瑞涛, 等. Ada-LoFTR: 自适应图像块增强的局部特征匹配算法[EB/OL]. 电光与控制. (2024-07-14) [2025-03-25]. https://link.cnki.net/urlid/41.1227.TN.20240711.1543.005.
ZHANG Z Y, YANG X G, LU R T, et al. Ada-LoFTR: local feature matching algorithm for adaptive image block enhancement[EB/OL]. Electronics Optics & Control. (2024-07-14) [2025-03-25]. https://link.cnki.net/urlid/41.1227.TN.20240711.1543.005 (in Chinese).

[23] 郭印宏, 王立春, 李爽. 基于重复性和特异性约束的图像特征匹配[J]. 图学学报, 2023, 44(4): 739-746.
GUO Y H, WANG L C, LI S. Image feature matching based on repeatability and specificity constraints[J]. Journal of Graphics, 2023, 44(4): 739-746 (in Chinese).

[24] WANG Y F, HE X Y, PENG S D, et al. Efficient LoFTR: semi-dense local feature matching with sparse-like speed[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2024: 21666-21675.

[25] DAI A, CHANG A X, SAVVA M, et al. ScanNet: richly-annotated 3D reconstructions of indoor scenes[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2017: 2432-2443.

[26] LI Z Q, SNAVELY N. MegaDepth: learning single-view depth prediction from internet photos[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2018: 2041-2050.

[27] DING X H, ZHANG X Y, MA N N, et al. RepVGG: making VGG-style convnets great again[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2021: 13728-13737.

[28] LIU W Z, LU H, FU H T, et al. Learning to upsample by learning to sample[C]//2023 IEEE/CVF International Conference on Computer Vision. New York: IEEE Press, 2023: 6004-6014.