引文格式：张小宇,康金,李敏. 利用神经网络描述符改进的山区影像匹配方法[J]. 地理空间信息, 2025, 23(12):117-120. doi:10. 3969/j.issn. 1672-4623.2025.12.026

利用神经网络描述符改进的山区影像匹配方法

张小宇[1]，康　金[2]，李　敏[3]
（1. 陕西农林职业技术大学交通与测绘工程学院，陕西 杨凌 712100；2. 陕西农林职业技术大学信息工程学院，陕西 杨凌 712100；
3. 自然资源部第一大地测量队，陕西 西安 710054）

摘　要：针对低空无人机影像匹配过程中存在的匹配速率慢、鲁棒性差等问题，提出了一种利用神经网络描述符改进的影像匹配方法。首先利用SURF进行特征点检测提取，然后采用浅卷积神经网络学习的局部特征描述符TFEAT对特征点进行描述，为验证本方法的效果，采用弱纹理环境山区影像进行实验及利用RANSAC算法剔除误匹配结果。将该方法与SIFT、SURF、ORB算法在不同条件下进行了对比分析，得出在尺度变化的山区影像匹配中相比SIFT提高了17.56%，增强了影像匹配的鲁棒性。

关键词：影像匹配；特征描述符；卷积神经网络；深度学习
中图分类号：P231　　**文献标志码：**B　　**文章编号：**1672-4623（2025）12-0117-04

Improved Mountains Image Matching Method Based on Neural Network Descriptor

ZHANG Xiaoyu[1], KANG Jin[2], LI Min[3]
(1. School of Traffic and Surveying Engineering, Yangling Vocational &Technical College, Yangling 712100, China;
2. School of Information Engineering, Yangling Vocational &Technical College, Yangling 712100, China;
3. The First Geodetic Surveying Brigade, Ministry of Natural Resources, Xi'an 710054, China)

Abstract: Aiming at the problems of slow matching rate and poor robustness of low-altitude UAV images in the matching process, we proposed an improved image matching method based on neural network descriptor. Firstly, we used SURF to detect and extract feature points. Then, we used TFEAT, a local feature descriptor learned by shallow convolutional neural network, to describe feature points. Finally, we used random sampling consistency (RANSAC) algorithm to reject the mis-matching results. In order to verify the effectiveness of method, taking mountainous image data for example, we compared and analyzed this method with SIFT, SURF, and ORB algorithms under different conditions. The result shows that compared with SIFT, the method improved by 17.56% in mountainous image matching with scale changes, enhancing the robustness of image matching.

Key words: image matching, feature descriptor, convolutional neural network, deep learning

影像匹配技术是通过数学方法在2幅具有重叠区域的影像中建立对应关系的过程，其核心在于识别并匹配同名点，该技术主要包含3个关键阶段：特征点提取（从影像中检测显著的关键点）、特征点描述（对提取的点生成可区分的特征向量）和特征点匹配（通过相似性度量确定2幅影像中的对应点）。近年来随着低空无人机和计算机视觉技术的不断进步，低空无人机影像匹配成为了摄影测量领域的关键环节，尤其是在山区、露天采矿区等地形情况较复杂的弱纹理环境下，匹配依然存在相当大的挑战。目前，国内外普遍研究的是基于特征的影像匹配算法，该算法的速度主要取决于特征检测、描述、搜索配对算法[1]。由于低空无人机的影像有时存在一些复杂的几何和物理变化，在实际应用中，影像匹配算法面临着更高的要求，包括匹配点数量、正确率、匹配精度以及匹配耗时等方面[2]。2004年，高莎等[3]对其提出的SIFT（scale invariant feature transform）局部特征方法进行了全面总结。但是在视角差较大时，存在算法鲁棒性低等问题。2011年鲁杰等[4]提出了ORB（oriented fast and rotated bRIEF）算法，这是在当前计算机视觉领域应用广泛的方法，却缺少方向的抵抗能力。2015年曹南等[5]提出基于深度学习的关键点检测器TILDE（temporally invariant learned dEtector），但在视角变化显著时检测性能受限。本文主要结合工程实际，选择云南文山山区的地貌影像进行实验研究，利用神经网络描述符改进的匹配方法进

项目来源：陕西省教育科学“十四五”规划2023年度课题（SGH23Y2970）；杨凌职业技术学院科创项目（ZK22-43、ZK24-14、ZK24-30）。
第一作者简介：张小宇（1991—），硕士，讲师，研究方向为无人机摄影测量，E-mail：824292972@qq.com。

行处理，与传统算法匹配结果对比分析，可以发现本文的方法相比较传统方法在弱纹理的山区有一定的精度提高。

1 神经网络描述符改进的算法

通常来说，影像匹配方法分三步：先由特征检测器提取关键点，再用局部描述子描述，最后建立特征点对应关系。本文以SURF方法检测关键点，并引入浅层卷积网络学习得到的TFEAT描述子[6-7]，以提升低空无人机山区影像的匹配性能。训练阶段，将山区影像中每个SURF关键点对应的64 × 64影像块作为输入，构造三重样本数据集，通过反向传播迭代更新网络参数，使网络最终输出128维的局部学习型描述符，从而实现对弱纹理山区场景的稳健匹配。

SURF算法是一种高效的特征点检测方法，它巧妙地结合积分图像和盒子滤波技术，在响应图像上应用非极大值抑制技术，筛选出不同尺度空间中具有显著特征的点。这一过程中Hessian矩阵行列式的响应图像，由不同大小的盒子滤波模板与积分图像结合生成，而积分图像的加减则用于近似实现高斯滤波，这是SURF算法的处理步骤[8-10]。此外，还通过简化高斯二阶微分模板，将原本复杂的高斯卷积运算转化为更高效的盒子滤波运算[8]，进一步优化了计算过程，为后续的图像匹配和识别任务奠定了坚实的基础。

TFEAT是利用三重样本中的训练样本，以及不相关样本的挖掘信息。在低空无人机航测山区影像数据中，选取不同地物对应的不相关影像块，可以负训练其参数，如不同地物的影像块可能因旋转、尺度等几何变换在光学影像中可能表现一致，而相同地物的影像块却可能因季节、尺度、光线、角度等差异，在光学影像中表现完全不同。低空无人机航测中大量存在的“难负样本”正是提升网络性能的关键，反向传播利用这些易混淆对，可显著优化浅层神经网络，从而获得更具判别力的学习型描述符。

本文所使用的训练数据集包含了3个区域的低空无人机影像数据，其中3个区域的数据覆盖场景分别对应城镇、平原、山区等。本文在每个区域选取100张影像并进行裁剪、旋转、缩放等操作形成了10 000对64×64的训练数据集，训练数据集信息如表1所示。

本次研究的浅卷积网络学习，是一种基于对比学习的深度特征优化方法，涉及到3种形式的样本的相关关系，从而实现模型学习具有判别性的特征表示。

表1 训练数据集信息表

训练区域1	训练区域2	训练区域3
城镇	平原	山区
		

训练样本中 a 和 p 是同一特征点的不同视角、尺度的样本，而 n 是不同特征点的样本。在优化网络参数计算时，拉近同类样本，使得特征空间的 a 和 p 接近，推远异类样本，使得特征空间的 a 和 n 远离。以上欧氏距离关系，可以用数学表达为公式（1）表示：

$$\begin{cases}\delta_+ = \left\| f(a) - f(p) \right\|_2 \\ \delta_- = \left\| f(a) - f(n) \right\|_2 \end{cases} \tag{1}$$

式中，δ_+ 和 δ_- 分别为正相关样本的相似性和负相关样本的相似性，优化网络参数，使得 a 和 p 接近。给定一个参数值 μ，当 $\delta_- > \delta_+ + \mu$ 时，值下降至0，停止更新网络参数，损失函数定义如（2）所示：

$$\lambda(\delta_+, \delta_-) = \max(0, \mu + \delta_+ - \delta_-) \tag{2}$$

此外，还可以将三重样本中的难负样本定义为：

$$\begin{cases}\delta_* = \min(\delta_-, \delta_-') \\ \delta_-' = \left\| f(p) - f(n) \right\|_2 \end{cases} \tag{3}$$

当 $\delta_* = \delta_-'$ 时，交换三重样本公式计算中的基准样本 a 和相关样本 p，即让相关样本成为新的基准样本，反向传播不相关样本。这种处理方式确保 δ_* 始终距离较大，从而在优化过程中更有效地利用这些难以区分的样本。相应的损失函数计算，如公式（4）所示：

$$\lambda(\delta_+, \delta_*) = \max(0, \mu + \delta_+ + \delta_*) \tag{4}$$

在浅卷积训练中采用三重样本的方式，其详细参数设置如下：初始学习率设定为0.01，在每次迭代之后，学习率会更新为原来的0.9倍，直至学习率下降到10^{-6}时停止训练。

对于影像匹配来说，浅层神经网络的特征描述符能否适应山区影像匹配需通过实验验证。具体流程是先用SURF提取特征点，再以64像素×64像素邻域裁剪图像块作为训练样本，经网络学习后输出128维特征向量，用于度量特征点间相似度。

2 实验数据

本文采用云南文山的低空无人机影像数据，该数据是山地开采区域，地表植被稀疏，主要为裸露土

地，区域内无建筑物分布，整体呈现弱纹理特征，地物特征不明显，给影像匹配带来一定挑战（图 1）。



图 1　实验影像数据

3　处理与分析

为系统评估 SIFT、SURF、ORB 以及改进型 TFEAT（基于浅层神经网络）4 种特征描述符在山区影像中的稳健性，本文设计并实施了两组实验：一组关注影像旋转变化的影响；另一组关注尺度变化的影响。实验结果统一采用特征点提取重复率、匹配正确数以及匹配正确率 3 项指标进行量化对比，最终形成了 4 种方法在山区场景下的匹配分析结果。

重复率定义为在具有重叠区域的 2 张影像中，均能成功检出的特征点数目，与任一张影像检出的特征点总数之比[11-13]。该比值越高，表明检测算法对影像变化的适应性越强，其数学表达式见公式（5）：

$$R=\frac{N}{\min(N_1,N_2)}h_{\text{overlap}} \tag{5}$$

式中，R 为重复率；N 为左影像和右影像的共同特征点数量；N_1 为左影像特征点数量；N_2 为右影像特征点数量；h_{overlap} 为重叠度。

影像匹配正确率[13-15]在本实验中定义为，在单应变换约束下，2 幅影像中同名像点误差小于 1.5 像素的正确匹配数与总匹配数之比，该指标可以衡量特征匹配的精度与可靠性。匹配正确率计算公式：

$$P=\frac{T}{N} \tag{6}$$

式中，N 为影像匹配总数量；T 为影像匹配的正确数量；P 为影像匹配的正确率。

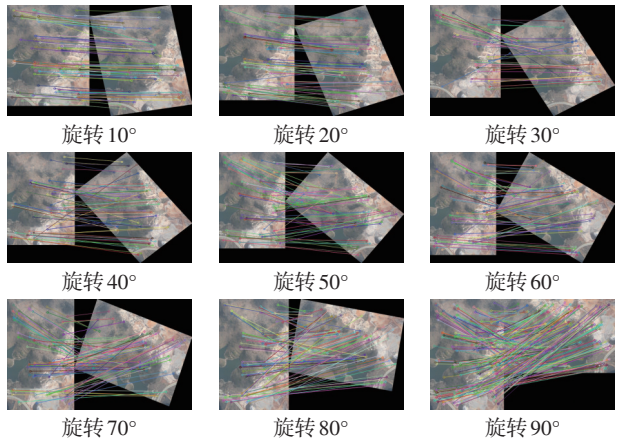


图 2　旋转不变性匹配结果

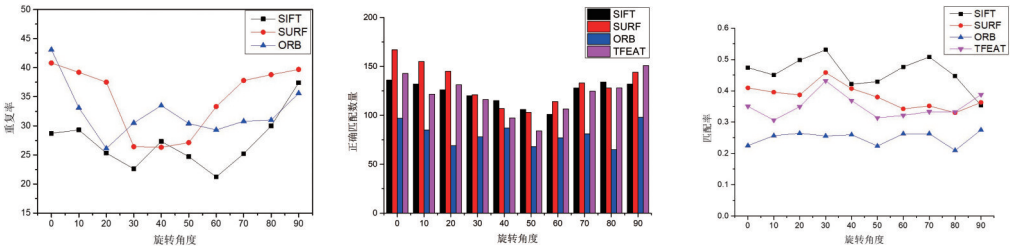
3.1　旋转不变性实验

基准影像保持不变，将参考影像数据按顺时针从 0°旋转至 90°，会得到旋转后的九张影像，然后利用本文所提匹配方法以及 3 种对比方法分别进行检测，具体旋转不变性匹配结果如图 2 所示。

针对旋转变化场景，对 4 种方法（SIFT、SURF、ORB、TFEAT）结果进行定量统计与分析。图 3 展示了 4 种匹配算法随着参考影像旋转角度增大，重复率、正确匹配数量、匹配率的变化情况。鉴于 TFEAT 描述符仍沿用 SURF 提取特征点，其重复率与 SURF 完全相同，故在特征点重复率对比图中不再另行绘制 TFEAT 曲线。在重复率检测方面，SURF 算法和 SIFT 算法在 0~30°的重复率检测均存在明显下降，ORB 算法亦在一定角度下有明显下降趋势。在正确匹配数量方面，SURF 算法、SIFT 算法、以及 TFEAT 算法均高于 ORB 算法。在匹配正确率方面，SIFT 算法的正确率最高，本文所提的 TFEAT 方法只是优于 ORB 方法，但从表 2 的统计结果中发现相比 SIFT 算法其匹配速度有很大的优势。

3.2　尺度不变性实验

以基准影像为参考，按 0.25 为步长对另一幅影像进行尺度缩放，使其尺度系数由 0.5 递增至 2.0，共生成 7 张不同尺度的待匹配影像。然后分别利用本文所



a 旋转角度的大小和重复率　　b 旋转的角度和影像匹配的数量　　c 影像匹配正确率

图 3　重复率检测对比分析结果如图

表2 不同角度下的匹配正确率和效率统计

匹配方法	不同旋转角度下的匹配正确率										平均用时/s
	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°	
SIFT	0.47	0.45	0.50	0.53	0.42	0.43	0.48	0.51	0.45	0.35	3.51
SURF	0.41	0.40	0.39	0.46	0.41	0.38	0.34	0.35	0.33	0.36	0.74
ORB	0.23	0.26	0.26	0.26	0.26	0.22	0.26	0.26	0.21	0.28	0.52
SURF+TFEAT	0.35	0.31	0.35	0.44	0.37	0.31	0.32	0.33	0.33	0.38	0.46

提匹配方法以及3种对比方法分别进行检测。

图4给出了在尺度变化条件下，4种方法（SIFT、SURF、ORB、TFEAT）特征点检测与匹配数量的定量统计结果。鉴于TFEAT描述符仍沿用SURF提取特征点，其重复率与SURF完全相同，故在重复率对比图中不再另行绘制TFEAT曲线。

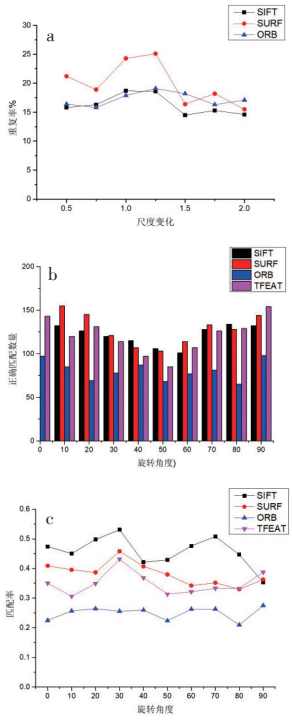


图4 影像特征点重复率对比分析结果

在本次实验中的重复率方面，当尺度范围介于0.5~1.5时， SURF的重复率表现出更好的效果，且3种算法的重复率整体较低。在正确匹配数量方面，由于针对山区影像数据匹配的训练模型，采用了多分辨率的影像数据，实验结果表明，改进的TFEAT方法在正确匹配数量上显著优于SIFT、SURF和ORB，在匹配正确率方面，4种算法均对尺度变化表现出一定适应性，但ORB因正确匹配数量偏低而整体正确率最低。尤为突出的是，TFEAT在矿山弱纹理区域仍能保持高正确率，优势最为明显。

4 结 语

本研究设计一种基于神经网络描述符改进的山区影像匹配方法，在山区影像匹配中有一定的实践应用价值。该算法从山区影像纹理弱、高差变化大等特点出发，以SURF检测的特征点为基准，进一步利用三重样本的浅层卷积神经网络学习局部特征描述符。进而改善了山区影像的匹配正确率。实验结果表明，该方法虽然在影像旋转角度不变性上与SURF算法的匹配率几乎一致，但是尺度变化上的匹配正确率比表现最好的SIFT算法平均高出17.56个百分点，比表现最差的ORB算法高出了24.71个百分点。

参考文献

[1] 张占平,武风英,包鹏章. 改进SIFT的倾斜无人机影像匹配方法[J]. 地理空间信息,2021,19(8):106-109

[2] 姜三,江万寿,郭丙轩. 词汇树索引约束的无人机影像快速特征匹配算法[J]. 武汉大学学报(信息科学版),2024,49(9):1 597-1 610

[3] 高莎,袁希平,甘淑,等. 集成SIFT算法与检测模型优化的UAV影像匹配方法[J]. 光谱学与光谱分析,2022,42(5):1 497-1 503

[4] 鲁杰,陈建,门宝霞等. 基于深度学习算法的高分辨率无人机遥感图像自动分割[J]. 自动化与仪器仪表,2023,42(8):5-9

[5] 曹南,蔡扬扬,李旭洋,等. 高分辨率遥感影像特征点自动化匹配方法研究[J]. 地理空间信息,2022,20(11):9-13

[6] 王晓红,邓仕雄,何志伟,等. 结合SURF算法和单应性矩阵的无人机影像匹配[J]. 测绘通报,2018(7):38-42

[7] 马伟乐. 基于空间一致性的影像匹配方法研究[D]. 武汉:武汉大学,2023

[8] 唐颖复,王忠静,张子雄. 基于改进SIFT和SURF算法的沙丘图像配准[J]. 清华大学学报(自然科学版),2021,61(2):161-169

[9] 胡佳怡. 高分辨率航空影像特征匹配技术研究 [D]. 南昌:东华理工大学, 2018

[10] 代友林,王晓红,李闯,等. 基于BRISK-LATCH的喀斯特山区无人机影像匹配算法[J]. 地理与地理信息科学,2021,37(1):26-32

[11] 邓仕雄,闫星光,刘继庚,等. 一种基于AKAZE算法的山区无人机影像匹配[J]. 矿山测量,2020,48(5):105-109

[12] 姚国标,黄鹏飞,龚健雅,等. 抗视差的宽基线弱纹理影像自动拼接算法[J]. 测绘学报,2023,52(9):1 515-1 527

[13] 康俊华. 基于全卷积神经网络的端到端立体密集匹配研究[J]. 测绘学报,2022,51(5):785

[14] 盘贻峰,李洋,黄丽霞,等. 一种多视角无人机影像匹配对优选方法[J]. 测绘科学,2023,48(3):171-180

[15] 何婧媛,谢生龙,田原,等. 卷积神经网络和特征融合的遥感图像场景分类[J]. 电子设计工程,2022,30(16):60-64