NCMNet：用于两视图匹配剪枝的邻居一致性挖掘网络

刘鑫，秦荣，严骏驰，杨巨峰

摘要—匹配剪枝在各种基于特征匹配的任务中发挥着重要作用，其目的是从初始匹配中识别出正确的匹配（内点）。在坐标空间和特征空间中寻找一致的 k 最近邻是之前方法中的一个常用策略。然而，根据最近邻的相似性约束，内点附近包含的许多不规则的错误匹配（离群点）会错误地变成内点的邻居。为了解决这个问题，我们提出了一种全局图空间来寻找具有着相似图结构的一致性邻居。这是通过使用全局连通图来实现的，该图基于空间和特征的一致性，显式地呈现匹配之间的密切关系。此外，为了增强该方法在各种匹配场景中的鲁棒性，我们开发了一个邻居一致性块去充分利用三种类型邻居的潜力。邻居一致性可以通过依次提取邻居内部上下文和探索邻居之间的交互来逐步挖掘。最终，我们提出了一个邻居一致性挖掘网络（NCMNet），用于估计参数模型并去除离群点。大量实验结果表明，在多种两视图几何估计的基准测试中，我们提出的方法优于其他先进的方法。同时，我们还进行了四项扩展任务，包括遥感图像配准、点云配准、3D 重建和视觉定位，以此来测试方法的泛化能力。源代码可见于：https://github.com/xinliu29/NCMNet。

Index Terms—匹配剪枝，特征匹配，邻居一致性，全局图，参数化模型

1 引言

在图像对之间准确估计特征匹配对于许多计算机视觉任务至关重要，例如视觉同步定位与建图（SLAM）[2]、运动恢复结构（SfM）[3], [4]、图像配准 [5], [6] 和视觉定位 [7]。给定一对图像，可以通过使用现有的特征提取方法获取特征关键点及其对应的描述子，这些工作包括人工设计的方法 [8], [9], [10] 和基于学习的方法 [11], [12], [13]。然后，我们通过对描述子施加相似性约束或利用先进的深度学习算法 [8], [12], [14] 来建立初始匹配。然而，由于局部描述子的局限性 [15], [16], [17]，尤其是在面临严重的光照变化、视角变化、遮挡、模糊等情况时，不可避免地会存在大量错误的匹配（即，离群点）。这些离群点会显著影响基于特征匹配的下游任务的准确性。因此，为了减轻这个问题，匹配剪枝 [18], [19], [20] 被用来进一步从初始匹配中识别出正确的匹配（即，内点）。

作为开创者，RANSAC [21] 及其变体 [22], [23], [24] 采用假设-验证框架，通过迭代方式寻找一个具有最多支持内点的最优参数模型。但它们的稳定性随着内点的比例降低而逐渐降低，这主要是由于大量离群值对模型生成产生了不利影响。除了以贪心的方式搜索可能的模型外，还有许多工作利用了匹配的几何特性 [16], [18], [25]。如图 1 (a) 所示，内点和离群点在二维刚性变换下的分布存在显著差异。也就是说，内点通常符合一致的约束条件（比如，相似的长度、角度和运动），而离群点则是随机分布的。因此，匹配的一致性被视为重要的先验知识，并已被广泛研究用于区分内点和离群点 [15], [26]。

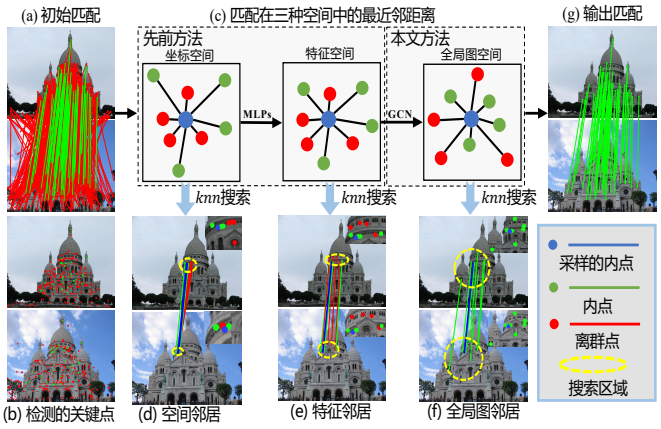


图 1. NCMNet 的处理流程。通过 SIFT [8] 特征关键点 (b) 建立的初始匹配 (a) 包含大量离群点。如 (c) 所示，我们的全局图空间可以缩短一些内点的最近邻距离，它们在其他两个空间中距离采样点较远，在全局图空间内这些内点也有可能成为邻居。我们展示了一个采样内点的三种邻居类型，包括 (d) 空间邻居、(e) 特征邻居和 (f) 全局图邻居。邻居搜索区域用黄色椭圆表示。如 (g) 所示，我们的 NCMNet 能够取得出色的结果。MLPs：多层感知器。GCN：改进的图卷积网络。

邻居一致性因其高效性而受到广泛关注，因为它仅关注每个匹配中的少量元素。为了明确定义邻居，早期的研究 [18], [27], [17] 在原始匹配的坐标空间内使用 k-最近邻（knn）搜索来寻找空间上一致的匹配，这些邻居称为空间邻居。近年来，一些基于学习的方法 [28], [29] 开始在通过网络学习得到的高

- 刘鑫，秦荣，杨巨峰：南开大学计算机学院，南开国际先进研究院（深圳福田），中国。杨巨峰还在中国深圳鹏城实验室工作。（邮件：*xinliu_0209@163.com*，*qinrong_nk@mail.nankai.edu.cn*，*yangjufeng@nankai.edu.cn*）。
- 严骏驰：上海交通大学人工智能研究院，中国。（邮件：*yanjunchi@sjtu.edu.cn*）。
- 本文的通讯作者为：杨巨峰。

- 本文为 *TPAMI2024* 论文 *[1]* 的中文译版。由刘鑫、王士博翻译，杨巨峰校稿。

维特征空间中使用 k-近邻（knn）搜索来寻找特征一致的匹配，这些邻居称为特征邻居。这些研究表明，邻居一致性在区分匹配方面具有良好的发展前景。然而，初始匹配在图像对中通常分布不均匀，这可能导致内点附近存在大量随机的离群点，特别是约 90% 离群点的宽基线场景中 [30], [28]。因此，如图 1 (c) 所示，一些匹配由于在上述两个欧式空间中彼此接近而错误地成为了邻居。如图 1 (d-e) 所示，在坐标空间和特征空间中，一个采样内点搜索到的邻居（蓝线）包含了一些意料之外的离群点（红线）。实际上，仅通过欧式空间中的相似性约束来处理这种情况是相当有挑战性的 [16], [31]。

为了解决上述问题，我们提出了一种非欧式全局图空间。具体地，内点在全局层面上往往具有一致性 [21], [16], [31]。也就是说，一个采样内点与其他内点之间具有强连接，而与离群点之间的连接则较弱或不存在。它们能够形成相似的图结构 [32], [33], [34], [35]，这些结构可以被图卷积网络（GCN）[36] 很好地识别。因此，我们通过构建一个图空间来捕捉这种全局一致性，其中邻居的定义取决于图结构的相似性。随后，我们采用改进的 GCN 来进一步探索这种一致性，并增强匹配之间的长距离依赖。与先前的欧式空间相比，内点在全局图空间中具有更强的的密切关系。因此，我们的全局图空间能够拉近具有相似图结构的匹配之间的距离。这些匹配在坐标和特征空间中可能难以成为邻居，因为它们的最近邻距离较大，如图 1 (c) 所示。更具体地，我们首先构建一个加权全局图，其中节点代表所有匹配，边表示使用一致性分数计算的成对亲和度。为了获得一个更具代表性的图，我们基于长度约束开发了一种空间一致性，以补充 [37] 中使用的特征一致性。接下来，我们利用改进的 GCN [36] 来获得我们的全局图空间。最终，我们在这个空间中使用 knn 搜索来识别全局一致的匹配，这些匹配称为全局图邻居。值得注意的是，全局图邻居在空间上和采样的匹配并不接近，如图 1(f) 所示。换句话说，由于我们采用了全局图操作，它具有更大的搜索区域（参见消融实验）

空间邻居和特征邻居分别通过低维的空间相似性和高维的特征相似性进行搜索，这些邻居集中在采样匹配的局部范围内。相比之下，我们的全局图邻居则关注具有相似图结构的全局一致邻居。如 [38], [28] 中所述，匹配剪枝需要丰富的局部和全局上下文信息。因此，为了增强在困难匹配场景中的鲁棒性，我们设计了一个邻居一致性（NC）块，以充分利用三种类型邻居的潜力。NC 块包含三个基本组件：邻居嵌入构建、自我上下文提取 (SCE) 层和交叉上下文交互 (CCI) 层。具体来说，我们首先根据不同的邻居构建三个有向图作为邻居嵌入。为了提取相应的邻居上下文特征，SCE 层以分组卷积的方式动态捕捉邻居内部的关系并聚合其上下文信息。CCI 层用于进一步探索它们之间的交互。由于 [37] 中使用的单一交叉注意力分支表现能力有限，我们设计了一种分层分组方式去有效融合和调整邻居间的交互信息。基于 NC 块，我们提出了邻居一致性挖掘网络（NCMNet） [37] 以及包含两项改进的 NCMNet+，以实现两视图匹配剪枝。

本文的贡献可以总结为以下三点：（1）通过空间和特征一致性在匹配之间构建显式连接，我们提出了一个全局图空间。该空间用于寻找具有相似图结构的一致邻居（2）利用三种类型邻居的潜力，我们开发了 NC 块，通过提取邻居内部上下文信息以及探索邻居之间的交互，渐进地挖掘邻居一致性。（3）我们通过一系列几何估计基准测试和扩展任务证明了我们方法的有效性和泛化能力。

本文是我们发表在 CVPR 2023 初步会议版本 [37] 的扩展版，其中改进如下：（1）在加权全局图的构建中，NCMNet 仅依赖于从网络中学习的特征一致性，但由于学习过程中的模糊性，这可能会不够准确。NCMNet+ 应用了匹配中固有的空间一致性来补充特征一致性，从而增强全局图空间的可靠性。（2）在 CCI 层中，NCMNet 使用单一的交叉注意力操作来探索邻居之间的交互。在此基础上，NCMNet+ 采用了有效的分层分组方式来丰富信息的整合，从而提高了匹配剪枝的准确性。（3）在验证方面，NCMNet 主要关注于通过估计本质矩阵来恢复相机姿态。在本研究中，我们增加了基础矩阵和单应矩阵的估计。同时，我们进行了广泛的实验和详细的消融分析，以全面理解我们的方法。（4）我们进一步将所提出的方法扩展到四个基于特征匹配的任务，包括遥感图像配准、点云配准、3D 重建和视觉定位。

2 相关工作

一般来说，特征匹配工作可以分为两大类：无检测器方法和基于检测器的方法。无检测器方法 [39], [40], [41], [42], [43] 直接处理图像对并生成像素级的稠密匹配。虽然这些方法效果很突出，但由于图像中像素数量极多，它们通常会带来巨大的计算开销。相比之下，基于检测器的方法 [8], [11], [44], [45] 在过去几十年中发挥了关键作用。这些方法通过检测显著的关键点来构建逐点的稀疏对应关系，然后进行匹配。然而，由于内点和离群点之间的不均衡分布，寻找准确的特征匹配仍然十分复杂 [46], [47]。这一问题可以通过进一步应用匹配剪枝方法来缓解。在接下来的章节中，我们将对相关背景材料进行详细回顾。

2.1 RANSAC 相关方法

作为近几十年来最著名的算法之一，RANSAC [21] 使用一种假设-验证框架来寻找最大的内点集。更具体地，它随机选择一个最小的数据子集来生成一个假设的参数模型（比如，用于本质矩阵估计的 5 个匹配 [48]）。然后，可以通过符合该模型的匹配数量来验证此模型的可靠性。这个过程将持续进行，直到达到预定义的迭代次数或阈值。遵循这一框架，后续的工作 [22], [23], [24], [49] 通过不同的采样和验证策略来提高效率或效果。MLESAC [22] 通过最大化匹配点的对数似然来确定最优模型，以增强算法的鲁棒性。USAC [24] 回顾了相关的变体，并基于一些重要的考量提出了一种通用结构，表

现出更好的改进效果。MAGSAC [49] 通过使用 σ-一致性来避免预定义的内点-离群点阈值，从而实现了更优的性能。此外，一些变体 [50], [51], [52], [53], [54], [55] 利用深度学习框架来提升参数模型的质量。这些 RANSAC 相关的工作仍然被视为离群点移除和参数模型估计的标准解决方案。然而，这种随机采样策略对离群点较为敏感 [46], [56], [57]。随着初始匹配中离群点比例的逐渐增加，它们的性能会显著下降。

2.2 基于学习的方法

深度学习技术发展催生了很多开创性的工作 [50], [58], [19]，这些工作利用神经网络来实现匹配剪枝。例如，DSAC [50] 基于概率选择设计了一个可微的 RANSAC 副本。近年来，PointNets [59], [60] 使用多层感知器（MLPs）来处理无序和不规则的点集，并受到了广泛关注。受到 PointNets 的启发，LFGC [19] 基于多层感知器（MLPs）训练一个排列不变结构，以估计匹配的内点权重并回归由本质矩阵编码的相机姿态。同样地，DFE [58] 也采用深度网络来预测用于基础矩阵估计的内点权重。

后续的方法将这种匹配分类范式作为标准，并通过不同方式提升性能。一方面，一些工作设计了多样的网络结构去捕捉丰富的上下文信息。为了获取局部上下文信息，OANet [38] 通过一个可微的池化层学习软分配矩阵，用于对输入的匹配进行聚类。然后，它通过上采样操作恢复匹配的原始尺寸。它还通过转置特征的维度来利用全局上下文信息。T-Net [61] 引入了一种 T 形网络架构，用于整合迭代子网络的输出。ConvMatch [62] 开发了一种规则的运动场，并探索了使用卷积神经网络来捕捉上下文信息的可能性。另一方面，一些研究人员利用注意力机制 [63] 来增强关键特征的表示能力。ACNe [64] 利用注意力权重从局部和全局方面对特征进行归一化。ANA-Net [65] 计算注意力权重的相似性来发现注意力一致的匹配。MSA-Net [66] 提供的多尺度注意力和 PGFNet [67] 使用的分组残差注意力能够进一步提高匹配剪枝的准确性。尽管上述工作表现出色，但它们仍然存在一些局限性。首先，在匹配学习过程中设计与数据无关的操作来隐式地捕捉上下文并不直观。其次，离群点会严重阻碍网络的学习和收敛。尽管注意力机制旨在缓解这个问题，但大多数方法在网络训练过程中仍然容易受到高比例离群点的负面影响。与这些方法不同，我们充分利用不同类型的邻居一致性，将匹配的内在几何和特征属性显式地整合到网络学习过程中。我们还采用了迭代剪枝策略 [28] 作为基本框架，以提取更可靠的候选匹配，从而改进网络学习。

2.3 匹配的一致性

在二维刚性变换下，内点通常具有一致性约束，而离群点的分布则是随机的 [15]。因此，匹配的一致性是区分内点和离群点的一个重要线索，过去对此已进行了广泛研究 [68], [46], [69]。例如，BF [16] 通过提出的双边函数来制定分段一致性约束，以实现全局建模。CODE [31] 在全局层面设计了一种一致性可分离的约束，用于过滤高噪声的匹配。GMS [18] 通过基于网格的分数估计器寻找一致的空间邻居，以确定匹配的可靠性。LPM [70] 也利用预定义的统计测量方法来探索邻居一致性。这些人工方法需要精细的参数调节才能取得满意的性能，同时，它们对视角大幅变化等具有挑战性的匹配场景较为敏感 [46], [47]。

受这些传统技术的启发，一些工作开始以可学习的方式探索匹配的一致性。NM-Net [30] 开发了一个层次化网络，基于局部仿射信息 [71] 来挖掘特定兼容性邻居的上下文。LMCNet [17] 将空间邻居的运动一致性重新表示为一个通过图拉普拉斯算子解决的平滑函数。CLNet [28] 在特征空间中搜索邻居，并设计了一个从局部到全局的一致性学习框架。MS^{2}DG-Net [29] 也通过语义动态图来利用特征邻居的局部拓扑结构。它们通过各种网络结构或学习范式来聚合邻居信息，以实现鲁棒的匹配剪枝。然而，如节 1所分析的那样，由于大量离群点的不规则分布，从上述空间中搜索到的邻居可能不一致。在本文中，我们提出了一种全局图空间，在全局层面上显式地捕捉具有强一致性的内点，使得具有相似全局图结构的匹配能够成为邻居。同时，为了增强方法在复杂匹配场景中的鲁棒性，我们有经验地设计了一个邻居一致性块。它通过我们提出的 SCE 层和 CCI 层渐近地提取并整合三种类型的邻居上下文。

3 方法

本节将详细描述我们的方法。首先描述如何定义双视图匹配剪枝的问题。接着介绍 NCMNet 的实现细节，包括全局图空间、邻居一致性块以及网络架构。最后我们会给出对于损失函数的描述。

3.1 问题定义

给定一对匹配的图像，我们可以利用人工定义的特征提取方法 [8], [10] 或基于学习的方法 [11], [12] 来获取特征关键点及其相应的描述子。初始匹配集合 $S = \{s_1, s_2, ..., s_N\} \in \mathbb{R}^{N\times 4}$ 可以通过描述子的相似性匹配策略或神经网络 [72] 来估计。这里，$s_i = (u_i, v_i)$ 表示第 i 个匹配，其中 u_i 和 v_i 分别是在两幅匹配图像中使用相机内参归一化后的关键点坐标。N 表示初始匹配的数量。但是，特征描述子的模糊性会不可避免地导致离群点的出现。因此，我们匹配剪枝的目的是从初始匹配中过滤掉离群点。

为了实现这一目的，匹配剪枝过程通常将初始匹配集合 S 作为**输入**，并**输出**所有匹配的标签（即，离群点或内点）。也就是说，集合 S 被分为一个内点集合 S_{in} 和一个离群点集合 S_{out}。同时，通过使用内点集合 S_{in} 来估计参数模型（比如，本质矩阵），以评估方法的性能。参数模型用于恢复匹配图像的相机姿态，包括相应的旋转和平移向量。总之，匹配剪枝方法的优化目标是寻找足够的内点以恢复精确的相机姿态。

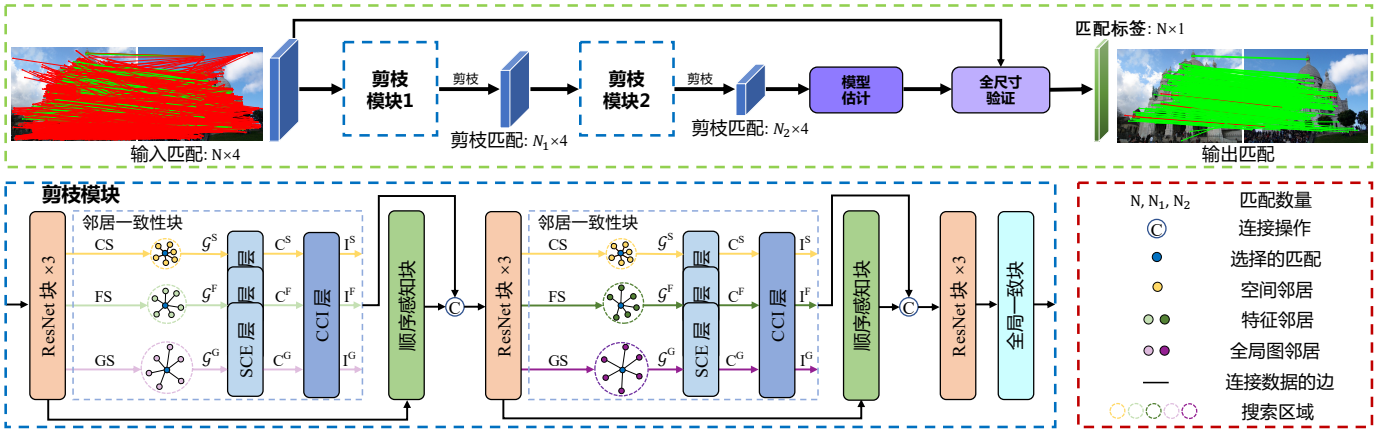


图 2. 我们提出的 NCMNet 的框架。$N \times 4$ 的初始匹配作为输入，然后估计参数模型和 $N \times 1$ 的内点概率。采用包含两个剪枝模块的迭代剪枝策略作为核心架构，以提取更可靠的候选者用于模型估计。每个剪枝模块包括一些现有的网络结构和所提出的邻居一致性（NC）块。NC 块主要由三个关键部分组成：三种邻居嵌入 $(\mathcal{G}^S, \mathcal{G}^F, \mathcal{G}^G)$ 的构建，自我上下文提取 (SCE) 层用于捕捉和聚合邻居内部的上下文 (C^S, C^F, C^G)，以及交叉上下文交互 (CCI) 层用于融合和调整邻居之间的信息 (I^S, I^F, I^G)。CS：坐标空间，FS：特征空间，GS：全局图空间。

更具体地，我们以图 2中所示的 NCMNet 为例。我们选择迭代剪枝策略 [28] 作为核心架构，以减轻离群点在网络学习过程中的负面影响。$f_{\theta 1}(S) = (S_1, o_1)$ 和 $f_{\theta 2}(S_1) = (S_2, o_2)$ 代表两个顺序的剪枝模块，其相关参数为 $\theta 1$ 和 $\theta 2$。$S_1 \in \mathbb{R}^{N_1 \times 4}$ 和 $S_2 \in \mathbb{R}^{N_2 \times 4}$ 是两个剪枝后的匹配集合，其中 $N > N_1 > N_2$。与 S 相比，S_1 和 S_2 预计会更可靠，它们是由学习到的逻辑值 o_1 和 o_2 决定的，并且会被用于参数模型估计。接下来，o_2 通过一个额外的 ResNet 块 [19] 和一个 MLP 层进行处理，以计算内点权重集合 w_2，具体如下：

$$w_2 = \tanh(\mathrm{ReLU}(o_2)) \in [0, 1), \tag{1}$$

其中，$\tanh(\cdot)$ 和 $\mathrm{ReLU}(\cdot)$ 表示激活函数。然后，我们利用 S_2 和 w_2，并使用加权八点算法 [19], [28] 来估计一个本质矩阵 $\hat{E} \in \mathbb{R}^{3\times3}$。最后，通过全尺寸验证操作，可以获得所有匹配的标签集合 $l \in \mathbb{R}^{N\times1}$。该架构表示如下：

$$\hat{E} = g(S_2, w_2), \tag{2}$$

$$l = v(\hat{E}, S), \tag{3}$$

其中，$g(\cdot)$ 指的是加权八点算法 [19]，相比于传统的八点算法 [15]，该算法由于考虑了内点权重，因此具有更强的鲁棒性。值得注意的是，在加权八点算法中，所采用的自伴随特征值分解操作相对于内点权重是可微的，这有助于本质矩阵的端到端回归。$v(\cdot)$ 是基于极线约束的全尺寸验证操作 [15]，用于避免一些内点被错误移除。还有值得注意的一点是，对于极线距离由 $\hat{E}$ 计算得出小于 10^{-4} 阈值的匹配，被视为**内点**：

$$S_{in} = \left\{ s_i \mid l_i < 10^{-4} \right\}, \tag{4}$$

其中，S_{in} 是保留的内点集合。类似地，**离群点**可以定义为：

$$S_{out} = \left\{ s_i \mid l_i > 10^{-4} \right\}, \tag{5}$$

其中，S_{out} 是离群点集合。这与确定真实匹配标签的方式相同。

3.2 增强的全局图空间

邻居一致性是用于区分匹配的一种有效线索，其利用了内点的邻居彼此兼容，而离群点则随机分布的事实。因此，为每个内点寻找可靠且一致的邻居是很重要的。在本文中，我们采用一种可微的方法，利用三种类型邻居的潜力来处理复杂的匹配情况。为了寻找不同类型的邻居，我们采用了三种不同的邻居搜索空间，包括坐标空间 $S \in \mathbb{R}^{N\times4}$、特征空间 $F \in \mathbb{R}^{N\times d}$ 和我们的全局图空间。d 是通道的数量。S 表示网络的输入，而 F 代表从多个 ResNet 块中学习到的中间特征映射。我们可以通过在 S 上执行 k-最近邻（knn）搜索来获得每个匹配的空间 k-近邻。同样地，可以在 F 上获取特征空间 k-近邻。空间和特征邻居关注的是具有相似低维坐标和高维特征的匹配。我们的全局图空间旨在通过改进的图卷积网络 [36], [28] 找到具有相似图结构的全局一致邻居。与 [37] 相比，我们在全局图构建中引入了匹配的空间一致性，以补充原有的特征一致性，这称为增强的全局图空间。

具体来说，我们首先构建一个加权全局图 $\mathcal{G}^g = \{\mathcal{V}^g, \mathcal{E}^g\}$，其中节点 $\mathcal{V}^g$ 表示所有匹配，无向边 $\mathcal{E}^g$ 使用增强的兼容性得分 s^c_{ij} 连接每两个匹配，公式如下：

$$s^c_{ij} = s^f_{ij} \odot s^s_{ij}, 1 \le i, j \le N. \tag{6}$$

它基于特征和空间一致性分数，表示匹配 s_i 和 s_j 之间的密切关系。与 [37] 类似，我们基于 F 估计初始内点权重 w^p：

$$w^p = \mathrm{ReLU}(\tanh(\mathrm{MLP}(F))), \tag{7}$$

其中，$\mathrm{MLP}(\cdot)$ 表示一个用于将通道维度减少到 1 的 MLP 层。然后，两个匹配之间的特征一致性分数计算如下：

$$s^f_{ij} = w^p_i \cdot w^p_j, \tag{8}$$

该公式用于衡量匹配之间的特征相似程度。此外，我们还使用匹配对之间的空间一致性分数来补充特征一致性分数，公式如下：

$$s_{ij}^{s} = \max(0, 1 - \frac{d_{ij}^2}{\epsilon_d^2}), \tag{9}$$

其中，$max(0,\cdot)$ 操作用于避免负值。$d_{ij} = \left| ||u_i - u_j|| - ||v_i - v_j|| \right|$ 是两个匹配的空间差异，即 $s_i = (u_i, v_i)$ 和 $s_j = (u_j, v_j)$，这是基于长度约束计算的。ϵ_d 表示一个距离超参数，用于控制长度约束的敏感性。对于两个匹配 s_i 和 s_j，如果 d_{ij} 大于 ϵ_d，则它们被视为空间上不兼容，其 s_{ij}^{s} 取值为零。相反，当 s_{ij}^{s} 值较大时，它们在空间上是兼容的，这可以作为特征一致性分数的可靠调节器。因此，我们可以构建一个加权全局图 $\mathcal{G}^g$ 的加权邻接矩阵 $A = s_{ij}^{c} \in \mathbb{R}^{N\times N}$，它描述了匹配之间的长距离依赖性。只有当两个匹配同时具有较高的特征和空间一致性分数时，才会形成强关联，否则，该连接将会很弱或不存在。最后，我们利用谱图卷积操作 [36] 进一步学习这种关联性：

$$L = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}, \tag{10}$$

$$F^g = \sigma(LFW^g), \tag{11}$$

其中，$\tilde{A} = A + I_N$ 表示邻接矩阵，并加入了对角单位矩阵 I_N 的自连接。$\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$ 是 $\tilde{A}$ 的对角度矩阵。图拉普拉斯矩阵 L 将 F 调整到谱域。W^g 表示学习到的权重。$\sigma(\cdot)$ 是 ReLU$(\cdot)$ 激活函数。$F^g \in \mathbb{R}^{N\times d}$ 是增强的全局图空间，它可以有效地从两个不同的方面反映匹配的全局一致性，特别是针对内点。这可以拉近具有相似图结构的匹配的最近邻距离，使它们在我们的增强全局图空间中成为邻居。然后，我们在 F^g 上执行 knn 搜索，以获取每个匹配的全局图 k 近邻。由于全局操作，全局图邻居的邻居搜索区域较大（参见消融实验）。

3.3 邻居一致性块

三种邻居搜索空间关注不同类型的邻居。因此，为了增强方法在应对具有挑战性的匹配情况时的鲁棒性，我们提出了一个邻居一致性（NC）块，以渐近地挖掘这三种邻居的一致性。作为 NCMNet 的核心结构，我们的 NC 块由三个关键部分组成：邻居嵌入构建、自我上下文提取（SCE）层和交叉上下文交互（CCI）层。

邻居嵌入构建。当三种类型的邻居被搜索到后，我们首先需要考虑如何为网络学习构建对应的邻居嵌入。图结构很适合对元素之间的复杂关系进行表征和建模，这让它成为了各领域 [73], [74], [75], [76] 中的重要工具。因此，在这一部分中，根据每个匹配 s_i 的不同邻居构建了三个独立的有向图，即 $\mathcal{G}_i^S = \{\mathcal{V}_i^S, \mathcal{E}_i^S\}$, $\mathcal{G}_i^F = \{\mathcal{V}_i^F, \mathcal{E}_i^F\}$, $\mathcal{G}_i^G = \{\mathcal{V}_i^G, \mathcal{E}_i^G\}$。以 $\mathcal{G}i^S$ 为例，节点 $\mathcal{V}_i^S = \{s_{i1}^S, ..., s_{ik}^S\}$ 表示 s_i 的空间 k-近邻，而有向边 $\mathcal{E}_i^S = \{e_{i1}^S, ..., e_{ik}^S\}$ 则将 s_i 与其在 $\mathcal{V}_i^S$ 中的空间邻居连接起来。参考 [77], [28]，边的构建如下：

$$e_{ij}^S = [f_i, f_i - f_{ij}^S], j = 1, 2, ..., k. \tag{12}$$

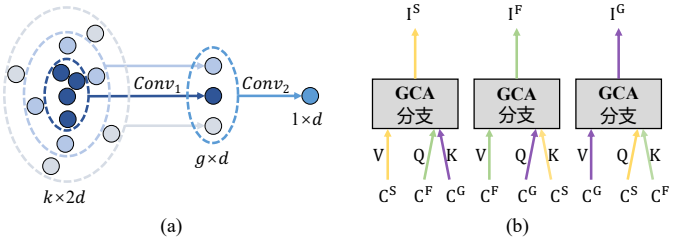


图 3. (a) 我们在 SCE 层中提出的分组卷积方式。它根据邻居节点与锚点的相关性将其分为 g 组。使用两个连续的卷积层（$Conv_1$ 和 $Conv_2$）动态提取邻居内部的上下文信息。(b) CCI 层的结构。**GCA**：分组交叉注意力。它包含三个并行的 GCA 分支，用于整合邻居之间的信息。从不同的邻居上下文特征中学习到的值（V）、查询（Q）和键（K）用于交叉注意力操作。

其中，f_i 和 f_{ij}^S 分别表示 s_i 及其第 j 个空间邻居 s_{ij}^S 在 $F = \{f_1, f_2, ..., f_N\}$ 中的特征映射。$f_i - f_{ij}^S$ 是它们的残差。$[\cdot,\cdot]$ 表示在通道维度上的特征连接操作。因此，我们可以得到所有匹配的空间邻居嵌入 $\mathcal{G}^S \in \mathbb{R}^{N\times k\times 2d}$。特征邻居嵌入 $\mathcal{G}^F \in \mathbb{R}^{N\times k\times 2d}$ 和全局图邻居嵌入 $\mathcal{G}^G \in \mathbb{R}^{N\times k\times 2d}$ 也可以通过相同的方法获取。

自我上下文提取（SCE）层。在构建了三个邻居嵌入之后，下一阶段涉及有效地挖掘邻居内部的上下文信息。一个简单的方法是使用常见的池化操作，例如最大池化和平均池化。然而，这些不加区分的方法会存在丢失图节点之间相关性的缺点。因此，为了充分利用我们邻居嵌入的图结构，SCE 层被提出用于邻居信息的聚合。考虑到图中的节点是根据相似性原则排序的，我们的 SCE 层采用了一种分组卷积方式 [28]，以动态获取邻居关系，并沿着图的边缘聚合邻居上下文信息。

更具体地，如图 3(a) 所示，给定 s_i 的一个邻居嵌入 $\mathcal{G}_i \in \mathbb{R}^{k\times 2d}$，节点根据它们与锚点的相关性被划分为 g 个子集，每个组包含 k/g 个节点。该嵌入通过两个连续的卷积层进行处理，之后是批量归一化（BN）[78] 和 ReLU 激活函数。这个处理过程可表示如下：

$$C_i = (Conv_2(Conv_1(\mathcal{G}_i)). \tag{13}$$

$Conv_1(\cdot)$ 和 $Conv_2(\cdot)$ 分别表示带有可学习核 $1 \times \frac{k}{g}$ 和 $1 \times g$ 的卷积层。为了简化，省略了 BN 和 ReLU。$C_i \in \mathbb{R}^{1\times d}$ 表示 $\mathcal{G}_i$ 的输出。在每个 NC 块中，使用三个并行的 SCE 层独立处理每个邻居嵌入，从而得到三个相应的邻居上下文特征，表示为 $\{C^S, C^F, C^G\} \in \mathbb{R}^{N\times d}$。

交叉上下文交互（CCI）层。一旦获得了这三个邻居上下文特征，我们的目标是协同地融合和调整邻居间的信息。在 [37] 中，我们的 CCI 层使用了一种交叉注意力操作，但由于其单一的顺序方式，其在探索邻居之间信息方面的能力有限 [63], [79], [67]。基于此，我们以分组的方式丰富了这三种特征的信息整合。如图 3(b) 所示，CCI 层由三个并行的分组交叉注意力（GCA）分支组成。在每个分支中，值 V 是从一个邻居上下文特征中学习得到的，而查询 Q 和键 K 则是从另外两个特征中得出的。GCA 分支的总览如图 4 所示。我们首

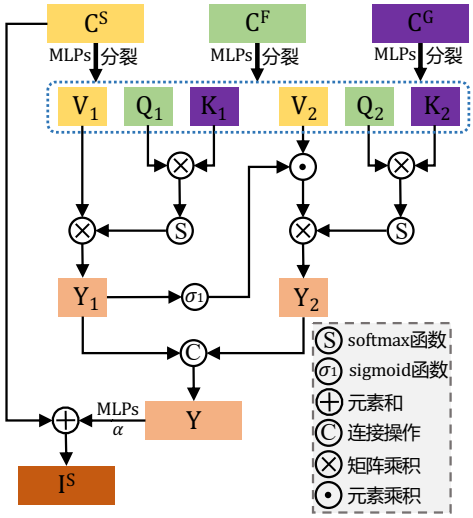


图 4. 提出的分组交叉注意力（GCA）分支。为了便于展示，分组大小设为 2。三个邻居上下文特征分别用于生成值（V）、查询（Q）和键（K）。然后，它们被平均分成 q 个特征组。除了交叉注意力操作外，每个特征组还接收前一个组的输出，以增加多样性和信息交流。

先将三个邻居上下文特征输入到一个独立的 MLP 层中，然后经过 BN 和 ReLU，生成三个新的特征 $\{\mathrm{Q},\mathrm{K},\mathrm{V}\}\in\mathbb{R}^{N\times d}$。然后，我们沿着通道维度将它们均匀分成 q 个组，表示为 $\{\mathrm{Q}_i,\mathrm{K}_i,\mathrm{V}_i\}\in\mathbb{R}^{N\times\frac{d}{q}}, i\in\{1,2,...,q\}$。对于第 i 组，执行 Q_i 和 K_i 的转置之间的矩阵乘法，然后使用 softmax 函数计算注意力权重矩阵：

$$A_i^w=\mathrm{softmax}(\mathrm{Q}_i{\mathrm{K}_i}^T), \tag{14}$$

其中，$A_i^w\in\mathbb{R}^{N\times N}$ 用于衡量匹配之间的相关性。接下来，执行 V_i 和 A_i^w 之间的矩阵乘法，以增强 V_i。同时，为了增强特征组之间的多样性和交流，我们采用了一种分层乘法操作来连接所有组。具体来说，除了第一组外，其他组可以利用前一组的输出信息。每组的输出定义如下：

$$Y_i=\begin{cases}\mathrm{V}_iA_i^w, & i=1;\\ \sigma_1(Y_{i-1})\odot\mathrm{V}_iA_i^w, & 1<i\le q,\end{cases} \tag{15}$$

其中，$\sigma_1(\cdot)$ 是 sigmoid 激活函数。在这一部分中，我们采用分组方式从不同的角度探索丰富的上下文，并通过分层乘法来增强特征组之间的交互。由于组合爆炸效应，这些操作有助于网络学习 [79], [67]。最后，将所有特征组的输出沿通道维度进行连接：

$$Y=\mathrm{concat}(Y_1,Y_2,...,Y_i), \tag{16}$$

其中，Y 是所有组的最终输出。这里，我们给出第一个 GCA 分支的示例：

$$I^S=\alpha(MLPs(Y))+C^S, \tag{17}$$

其中 $MLPs(\cdot)$ 由一个带有 BN 和 ReLU 的 MLP 层组成。学习得到的尺度参数 α 被初始化为 0。I^S 表示第一个 GCA 分支的输出，其中每个位置的响应是其他两个邻居特征所有位置与原始特征之间的加权结合。因此，内点可以通过选择性地聚合上下文，在三种邻居上下文特征中获得相互增益，这进一步提高了内点和离群点之间的区分能力。同样地，我们可以通过第二和第三个 GCA 分支分别生成结果 I^F 和 I^G。三种邻居交互特征 $\{I^S,I^F,I^G\}\in\mathbb{R}^{N\times d}$ 构成了所提出 NC 块的最终输出。

3.4 网络架构

我们使用 NC 块构建出了一个称为邻居一致性挖掘网络的匹配剪枝网络。NCMNet 的具体架构如图 2 所示。它以初始匹配作为输入，采用两个连续的剪枝模块渐近地提取可靠的候选项，这对精确预测参数模型和匹配标签至关重要。因此，为了提高候选者的可靠性，剪枝模块需要具备足够的能力来捕捉丰富的上下文信息。每个剪枝模块包括一些现成的网络结构 [19], [38], [28] 和我们的 NC 块用于匹配处理。作为基本结构，ResNet 块 [19] 包含用于匹配学习的两个 MLP 层和若干归一化技术。顺序感知块 [38] 旨在通过有序聚类操作隐式地捕捉局部和全局上下文。全局一致性块 [28] 编码特征的全局上下文信息，以估计用于剪枝匹配的全局分数。值得强调的是，特征空间和我们的全局图空间是可学习的。因此，我们引入了一个渐进式精炼处理（即，在每个剪枝模块中使用两个 NC 块），以提高邻域的可靠性并捕获全面的邻域上下文。进一步地，我们用 NCMNet+ 来表示使用了两个新改进的 NCMNet [37]，即增强的全局图空间和 CCI 层中的 GCA 分支。

3.5 损失函数

遵循 [38], [28]，神经网络通过分类损失和回归损失来进行优化：

$$\mathcal{L}=\mathcal{L}_c(o_m,y_m)+\beta\mathcal{L}_e(E,\hat{E}), \tag{18}$$

其中，β 代表平衡两项损失的权重系数。

分类损失 $\mathcal{L}_c(\cdot)$ 是一个二分类损失，定义如下：

$$\mathcal{L}_c(o_m,y_m)=\sum_{m=1}^{M}H(\tau_m\odot o_m,y_m), \tag{19}$$

其中，M 代表剪枝模块的数量。o_m 是第 m 个剪枝模块的相关逻辑值。y_m 代表通过默认阈值为 10^{-4} 的极线距离 d_{epi} 获得的弱监督下的地真标签。$\odot$ 是哈达玛积。$H(\cdot)$ 代表二分交叉熵函数。极线距离接近 d_{epi} 的内点可能会受到标签模糊的影响。这里，我们使用了一个自适应温度向量 τ_m [28] 来解决这个问题：

$$\tau_i=\exp(-\frac{||d_i-d_{epi}||_1}{d_{epi}}), \tag{20}$$

其中，d_i 是匹配 s_i 的极线距离。对于满足 $d_i>d_{epi}$ 的离群点，τ_i 会被设为 1。因此，有着更小极线距离的内点会对网络模型优化产生更强的影响。

对于回归损失 $\mathcal{L}_e(\cdot)$，我们应用了一种几何损失，定义如下：

$$\mathcal{L}_e(E,\hat{E}) = \frac{(p'^T\hat{E}p)^2}{\|Ep\|^2_{[1]} + \|Ep\|^2_{[2]} + \|E^Tp'\|^2_{[1]} + \|E^Tp'\|^2_{[2]}}. \tag{21}$$

虚拟匹配 (p, p') 是利用真实的本质矩阵 E 构造的，用于评估估计的 $\hat{E}$。$c_{[i]}$ 代表向量 c 的第 i 个元素。

4 实验

在接下来的部分，我们将 NCMNet/NCMNet+ 与最先进的匹配剪枝方法进行比较。实验在不同的基准测试上进行，以展示我们提出的网络的有效性和泛化能力。本文提供了实现细节、对比结果以及消融研究。

4.1 实现细节

如图 2 所示，我们的网络以不同特征提取方法生成的初始匹配作为输入，包括 SIFT [8]、ORB [10] 和 SuperPoint [12]。除非另有说明，我们默认使用 SIFT [8] 结合最近邻描述子匹配作为技术。在我们的实验中，匹配的数量 N 大约为 2000，通道维度 d 为 128。NCMNet 采用迭代剪枝策略 [28] 作为核心结构，该结构由两个连续的剪枝模块组成，剪枝率为 0.5。对于 NC 块，我们根据经验将邻居数量 k 设置为 9 和 6，分别用于两个剪枝模块。因此，我们将两个剪枝模块中 SCE 层的组数 g 分别设置为 3 和 2。在公式 9 中，距离超参数 ϵ_d 设置为 0.2。参考了 [67]，我们在 GCA 分支中选择了 $q = 4$ 作为组大小。顺序感知块中的簇数量设置为 250。**相关代码已经在https://github.com/xinliu29/NCMNet中提供，以保证我们结果的可被复现**。

训练细节。我们遵循之前的基准测试 [38]，使用 Pytorch 在给定的数据集上训练网络模型。优化过程中，我们采用了 Adam [80] 优化器，配置了批量大小为 32，学习率为 10^{-3}。网络训练共进行 500k 次迭代。在开始时，公式 18中的平衡参数 β 设为 0，然后在经过前 20k 次迭代后固定为 0.5。

4.2 对比结果

我们将提出的网络与一些最先进的匹配剪枝方法进行比较，包括传统方法和基于学习的方法。对于所有传统方法，我们首先采用固定阈值为 0.8 的比例测试 [8]，以去除大量质量较差的初始匹配，因为这些方法无法很好地处理高比例的离群点情况。对于所有基于学习的方法，我们使用整个初始匹配作为输入。我们将通过不同的任务来评估方法的性能和泛化能力。

4.2.1 几何估计

我们可以通过使用加权八点算法或 RANSAC 来估计本质矩阵，从而恢复两视图几何信息，包括旋转和位移。几何估计的质量对下游的特征匹配应用有着重要影响。因此，它是评估匹配剪枝算法性能的主要标准。

表 1
在 YFCC100M [81] 和 SUN3D [82] 上的定量比较结果。给出了已知和未知场景下的 mAP5° (%)。红色 表示最优结果，蓝色 表示次优结果。

方法	YFCC100M		SUN3D	
	已知	未知	已知	未知
RANSAC [21]	30.19	40.83	19.13	14.57
DEGENSAC [83]	21.00	27.65	16.01	11.01
GC-RANSAC [84]	30.43	41.58	18.86	14.14
MAGSAC [49]	32.80	41.61	20.35	16.24
MAGSAC++ [85]	30.48	40.95	18.90	14.19
AdaLAM [27]	32.37	45.40	21.02	15.94
LFGC [19]	16.87	25.95	11.55	09.30
DFE [58]	18.02	30.29	14.44	12.34
OANet++ [38]	33.96	38.95	20.86	16.18
ACNe [64]	29.17	33.06	18.86	14.12
LMCNet [17]	33.73	47.50	19.92	16.82
T-Net [61]	41.33	48.20	22.38	17.24
MS^{2}DG-Net [29]	39.68	48.20	22.20	17.84
MSA-Net [66]	39.53	50.65	18.64	16.86
CLNet [28]	39.16	53.10	20.35	17.03
PGFNet [67]	42.06	53.70	23.66	19.32
NCMNet [37]	52.33	63.43	26.12	20.66
NCMNet+	52.40	65.83	25.99	21.18

数据集。根据 [38]，我们使用雅虎的 YFCC100M [81] 作为室外场景数据集，SUN3D [82] 作为室内场景数据集来训练和测试网络模型。YFCC100M 包含了来自互联网的 1 亿张旅游图像，按照不同的地标分成了 71 个图像序列，其中选择了 4 个序列用于网络测试。SUN3D 包含大量从各种 RGBD 视频中采样的图像帧，它被分成了 254 个序列，其中 15 个图像序列用于测试。训练序列被分割成三个不重叠的部分，包括训练集 (60%)、验证集 (20%) 和已知测试集 (20%)。需要特别强调的是，室内场景尤其困难，因为它通常涉及许多无纹理区域和重复结构。

评估。图像对的相机姿态通过从估计的本质矩阵计算出的旋转和平移向量进行编码。我们选择地真向量与估计向量之间的角度差异作为误差度量。通过计算不同阈值下的平均准确率（mAP）作为两视图几何估计的评估标准，其中 5° 下的 mAP（即 mAP5°）为默认指标。

结果。在表 1 中，展示了在 YFCC100M 和 SUN3D 数据集上本质矩阵估计的定量对比结果。可以看到，使用比例测试可以显著提升 RANSAC 的性能，因为初始匹配中的大多数离群点已经提前被移除。这证明了 RANSAC 相关方法在处理高离群点比例时存在困难。使用比例测试的传统方法与一些基于学习的方法相比，能够展现具有竞争力的结果。显然，我们的方法在每个测试场景中都比所有传统和基于学习的基线有着明显的优势。例如，与排名第二的 PGFNet [67] 相比，NCMNet 分别在已知和未知的室外场景中获得了 10.27% 和 9.73% 的 mAP5° 巨大提升。我们提出的 NCMNet+ 通过使

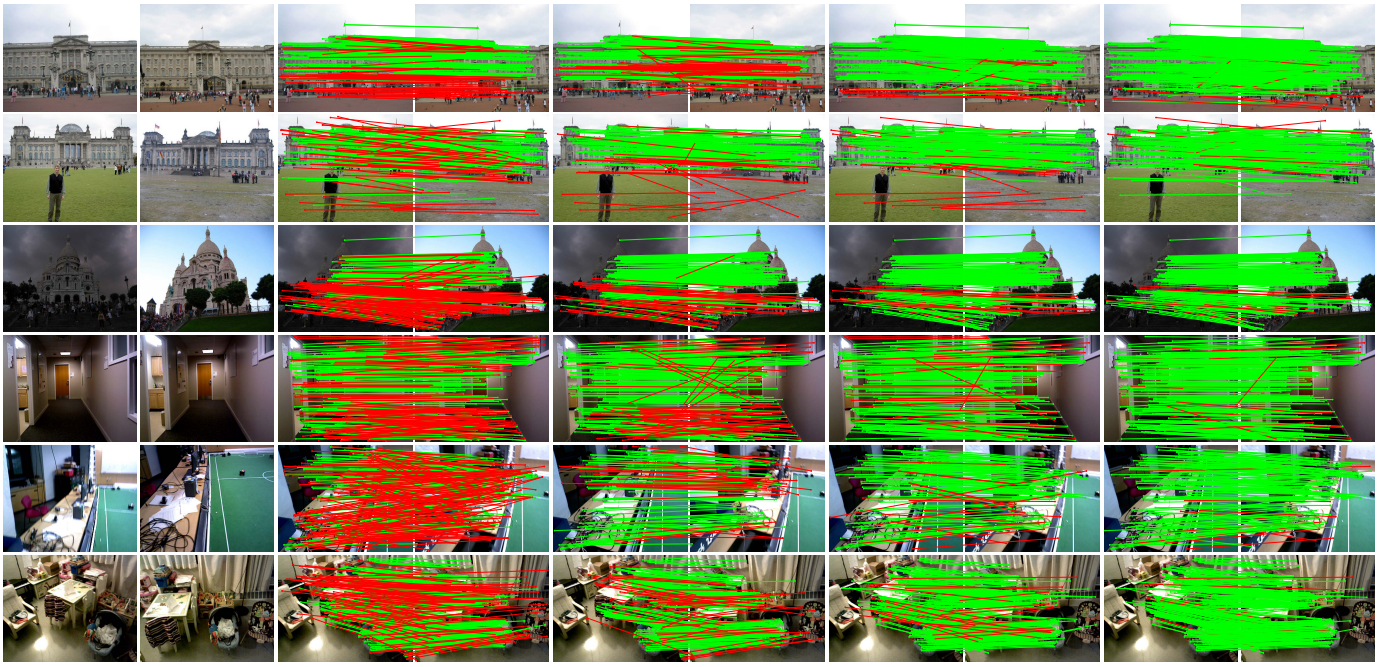


图 5. 匹配剪枝的可视化结果。从左到右列：匹配图像、OANet++[38]、CLNet[28]、NCMNet [37] 和 NCMNet+。前三个示例选自 YFCC100M [81] 上的未知场景，其余示例来自 SUN3D [82] 的未知场景。这些图像对涉及到大幅度的光照变化、视角变化、遮挡、重复结构、无纹理物体等情况。网络模型保留的内点（绿色和离群点（红色）被展示出来。

图 6. NCMNet+ 在 YFCC100M [81] 和 SUN3D [82] 上的失败案例。我们同样展示了地真内点。

表 2
使用 SIFT [8] 和 SuperPoint [12] 在 YFCC100M [81] 的未知场景上的性能比较。展示了在无或有 RANSAC [21] 作为后处理步骤的情况下的 mAP5°。

方法	SIFT [8]		SuperPoint [12]	
	-	RANSAC	-	RANSAC
RANSAC [21]	-	40.83	-	34.38
LFGC [19]	25.95	50.00	24.25	42.57
OANet++ [38]	38.95	52.59	35.27	45.45
MS^{2}DG-Net [29]	48.20	57.15	37.38	46.48
CLNet [28]	53.10	59.13	39.19	48.15
PGFNet [67]	53.70	57.83	42.03	47.30
ANA-Net [65]	31.55	59.10	-	-
NCMNet [37]	63.43	63.33	48.20	52.20
NCMNet+	65.83	64.15	49.80	53.35

用增强的全局图空间和改进的邻居间交互，能够进一步提升性能。图 5 展示了我们的方法与另外两个基线方法 [38], [28] 在匹配剪枝上的可视化对比结果。对于具有挑战性的室外和室内匹配场景，如大幅光照变化、视角变化、遮挡、重复结构和无纹理物体，我们提出的方法获得了可靠的剪枝结果。此外，图 6 展示了一些失败案例，其中离群点占主导地位。我们可以发现，在这些情况下，由于重叠区域有限、模糊以及低光照条件，真实内点相比于初始匹配（2000）非常稀疏。这使得匹配剪枝变得更加困难。根据表 3 和表 4结果，一个更强大的匹配估计器有可能缓解这一问题。另一个潜在的发现是，在这些情况下，这些弱监督的真实标签可能不可靠，这会干扰我们方法的学习和优化过程。我们认为提高模型对此类噪声标签的鲁棒性在未来的研究中是非常有价值和意义的。

此外，鲁棒的模型估计器 RANSAC [21] 被用作基于学习方法的后处理技术来估计本质矩阵。它将网络保留的匹配作为输入，并采用 0.001 的内点阈值，如 [38] 所述。我们还考虑使用学习到的特征提取方法来检测像素级关键点并构建相应的描述符。SuperPoint [12] 提出了一个自监督框架用于关键点检测和描述，在多视图几何问题中受到了广泛关注。在这里，我们采用 SuperPoint 结合最近邻匹配策略来构建两张图像的初始匹配。YFCC100M 数据集未知场景上的定量结果如表 2 所示。对于 ANA-Net [65]，由于训练过程不可用，我们直接使用了公开可用的网络模型。同样地，我们的 NCMNet 和 NCMNet+ 在所有情况下均优于所有基线方法。通过将

表 3
在 YFCC100M [81] 的未知场景上的对比结果。初始匹配由基于学习的匹配器估计，*i.e.*, SuperGlue [72] 和 LightGlue [14]。展示了 mAP5° 和 mAP10°。

Methods	SuperGlue [72]		LightGlue [14]	
	mAP5°	mAP10°	mAP5°	mAP10°
RANSAC [21]	59.90	71.14	63.23	74.04
LFGC [19]	58.88	70.79	62.05	73.10
OANet++ [38]	60.93	71.98	63.50	74.15
MS^{2}DG-Net [29]	59.95	71.30	62.63	73.25
CLNet [28]	63.10	74.00	68.65	78.18
PGFNet [67]	60.73	71.90	62.33	73.65
NCMNet [37]	66.33	76.33	70.38	79.24
NCMNet+	68.25	77.30	71.70	79.69

表 4
在 YFCC100M [81] 的未知场景上，网络对于不同稠密匹配器的泛化能力比较，包括 LoFTR [42] 和 DKM [86]。展示了 mAP5° 和 mAP10° 的结果。

方法	LoFTR [42]		DKM [86]	
	mAP5°	mAP10°	mAP5°	mAP10°
RANSAC [21]	68.58	77.58	74.85	82.13
LFGC [19]	64.93	74.64	73.45	81.43
OANet++ [38]	64.85	74.68	73.75	81.28
MS^{2}DG-Net [29]	66.90	76.09	73.15	80.76
CLNet [28]	69.78	78.35	75.28	82.30
PGFNet [67]	63.53	73.31	73.83	81.48
NCMNet [37]	70.25	78.55	75.55	82.51
NCMNet+	70.75	79.00	75.80	82.53

RANSAC 作为后处理步骤，可以进一步提高相机姿态的准确性，尤其是对于那些使用加权八点算法表现较差的方法（例如，LFGC 和 OANet++）。所有带有 RANSAC 后处理的学习方法都超过了基础的 RANSAC，表明基于学习的匹配剪枝比简单的比例测试更为有效。然而，我们发现当使用 SIFT 时，NCMNet 和 NCMNet+ 结合 RANSAC 的性能有所下降。这是因为 RANSAC 很难从网络保留的匹配中进一步提取合适的内点，而加权八点算法可以充分利用内点及其权重。同时，我们发现使用 SuperPoint 的方法表现不如使用 SIFT 的方法。这归因于学习到的特征提取方法具有有限的泛化能力，因此仅使用最近邻搜索来估计初始匹配的质量难以被保证。

基于学习的匹配器。最近，SuperGlue [72] 将特征方法生成的关键点作为输入，利用图神经网络来增强关键点的区分能力并学习最佳的匹配。它可以看作是描述子最近邻匹配的替代方案。在这里，我们首先使用 SuperPoint [12] 作为关键点检测器。接着，选择更具优势的 SuperGlue 和 LightGlue [14] 作为关键点匹配器，后者是 SuperGlue 的高效替代方案，用于估计初始匹配，网络模型由作者提供。最后，我们采用基于学习的匹配剪枝方法并结合 RANSAC 来估计本质矩阵，其中网络模型在 SuperGlue 生成的数据集上进行重新训练此外，我们发现全尺寸验证步骤中的极线距离阈值对性能有显著影响，因此在此实验中将其设为 1*e*-7。在表 3 中，我们给出了 YFCC100M 数据集的未知场景上的对比结果。可以看到，RANSAC 可以完成显著的性能收益。这是因为基于学习的 SuperGlue 和 LightGlue 通过提升关键点的表示能力，能够生成更准确的初始匹配。例如，当使用 SuperGlue 时，mAP5° 从表 2 中展示的 34.38% 提升到了 59.90%。因此，许多基于学习的剪枝方法难以进一步提高恢复的相机姿态的准确性。相比之下，我们的 NCMNet 和 NCMNet+ 在这两种情况下都能够获得合适的性能收益，因为我们的剪枝方法能够为几何估计提供更合适的内点。这进一步证明了我们的方法与更先进的关键点匹配器 [72], [14] 的兼容性。

泛化性。在本节中，我们采用不同的匹配器和数据集来评估网络模型的泛化能力。最近，一些稠密匹配器 [42], [88], [89], [86] 以图像对作为输入，不需要关键点来建立像素级的稠密匹配。在这里，我们使用最先进的匹配器构建初始匹配，包括半稠密匹配器 LoFTR [42] 和稠密匹配器 DKM [86]，其中模型是公开可用的。YFCC100M 数据集上未知场景的定量结果如表 4 所示。我们使用在 SuperGlue 上训练的剪枝网络模型，并根据经验将极线距离阈值设置为 1*e*-2。与稀疏匹配器相比，这些密集方法能够获得更好的性能，因为它们不受限于关键点检测器。我们的方法仍然带来了得体的提升，表明匹配剪枝作为一个补充模块的可用性和潜力。

表 5
网络在室外 PhotoTourism 和 Pragueparks [47] 数据集以及室内 ScanNet [87] 数据集上的泛化能力。展示了无/有 RANSAC 后处理时的 mAP5°。

Methods	PhotoTourism	Pragueparks	ScanNet
LFGC [19]	13.62/43.13	02.42/49.51	02.07/11.93
OANet++ [38]	30.35/48.39	07.37/49.17	04.73/14.27
MS^{2}DG-Net [29]	36.79/52.52	09.68/57.54	04.93/15.33
CLNet [28]	38.43/51.49	17.27/59.52	06.53/15.93
PGFNet [67]	41.22/52.34	09.90/56.00	04.87/14.93
NCMNet [37]	52.62/56.54	24.09/63.15	09.00/16.87
NCMNet+	52.93/56.84	27.17/63.48	10.13/17.47

同时，我们还分析了网络模型在不同数据集上的泛化能力，包括室外的 PhotoTourism 和 Pragueparks [47] 数据集，以及室内的 ScanNet [87] 数据集。PhotoTourism 是一个包含 9 个场景用于测试的旅游照片数据，Pragueparks 是一个小规模的视频序列，包含 3 个场景用于测试，其来自图像匹配挑战赛 [47]。ScanNet 是一个大型的 RGB-D 视频数据集，其中 [72] 提供了 1500 对测试图像。我们采用 SIFT [8] 结合最近邻匹配策略来建立匹配。网络模型分别在使用 SIFT 特征的室外 YFCC100M 或室内 SUN3D 数据集上进行训练。如表 5 所示，NCMNet 和 NCMNet+ 在不同的匹配场景中展现了比其他方法更好的泛化能力，进一步证明了我们方法的鲁棒性。

基础矩阵估计。在前面的实验中，我们通过估计本质矩阵来获取相对姿态，这假设相机内参是已知的。在运动恢复结构（SfM）流程中，估计基础矩阵是一种更广泛使用的方法 [3]。

表 6
YFCC100M [81] 未知场景上基础矩阵的估计结果。展示了不同误差阈值下的 mAP 结果。

Methods	mAP5°	mAP10°	mAP15°	mAP20°
LFGC [19]	19.90	30.96	39.11	45.68
OANet++ [38]	30.95	42.63	50.79	56.82
MS^{2}DG-Net [29]	36.63	48.56	56.68	62.76
CLNet [28]	47.98	57.51	63.88	68.50
PGFNet [67]	40.20	51.13	58.35	63.54
NCMNet [37]	53.03	62.89	68.83	73.18
NCMNet+	54.60	63.79	69.53	73.70

表 7
HEB [90] 数据集上单应矩阵估计的结果。展示了重投影误差（RPE）、角度姿态误差（APE）、旋转误差（RE）和绝对平移误差（ATE）的平均准确率。

Method	*RPE*	*APE*	*RE*	*ATE*
RANSAC [21]	27.89	2.71	19.25	27.34
LFGC [19]	39.02	3.74	25.47	31.06
OANet++ [38]	39.65	3.98	26.41	31.55
MS^{2}DG-Net [29]	39.22	2.96	26.01	30.59
CLNet [28]	40.08	3.98	26.44	31.56
PGFNet [67]	38.34	3.84	25.53	31.00
NCMNet [37]	42.14	4.35	27.38	32.17
NCMNet+	43.86	4.51	28.32	32.64

它们之间的主要区别在于，后者使用原始图像坐标作为输入，而不是归一化坐标。因此，我们在 YFCC100M 数据集重新训练网络模型，并使用加权八点算法来估计基础矩阵，并使用与本质矩阵估计相同的评估指标。如表 6 所示，NCMNet 和 NCMNet+ 在不同误差阈值下的 mAP 继续显示出比其他最先进方法的显著优势。此外，我们可以观察到，与其他使用本质矩阵的竞争对手相比，我们的方法可以获得相当或更优的从基础矩阵中恢复的相机姿态估计结果。

4.2.2 单应矩阵估计

单应矩阵能够表示两个平面或视角之间的变换，在计算机视觉应用中起着关键作用 [15]。寻找图像对的可靠单应矩阵同样需要准确的特征匹配。在本次实验中，我们评估用于单应矩阵估计的匹配剪枝方法。

数据集。 HEB [90] 是一个大规模的单应矩阵数据集，包含从 Pi3D 数据集 [90] 中采样的 226,260 个图像对之间的单应矩阵。测试集由九个视角和光照变化显著的场景组成，因此内点比例较低。鉴于训练数据不足，我们遵循 [90] 中的评估策略，使用在 YFCC100M 数据集上预训练的网络模型进行评估。

评估。根据基准测试 [90]，我们展示了重投影误差（RPE）、角度姿态误差（APE）、旋转误差（RE）和绝对平移误差（ATE）的平均准确率（mAA），以评估模型在过滤离群点用于单应矩阵估计中的表现。四种误差的 mAA 按以下阈值计算：APE 和 RE 的阈值为 1 到 10 度，ATE 的阈值为 0.1 到 5 米，RPE 的阈值为 1 到 20 像素。

表 8
关于每个剪枝模块中关键组件性能收益的消融研究。**IPS**: 迭代剪枝策略。**SCE**: 自我上下文提取层。**CCI**: 交叉上下文交互层。**PNR**: 渐进邻居精化处理。**OA**: 有序感知块。**粗体**表示最佳结果。

IPS	SCE	CCI	PNR	OA	mAP5°	mAP20°
✓					53.10	76.11
✓	✓				56.50	78.34
✓	✓	✓			58.63	80.03
✓	✓	✓	✓		61.73	81.46
✓	✓	✓	✓	✓	**63.43**	**82.46**

表 9
同时使用三种类型邻居的有效性。**SN**: 空间邻居。**FN**: 特征邻居。**GN**: 全局图邻居。

	Three SN	Three FN	Three GN	SN+FN+GN
mAP5°	61.40	62.60	61.73	**63.43**
mAP20°	81.26	81.74	81.31	**82.46**

结果。匹配剪枝方法在单应矩阵估计中的定量对比结果在表 7中展示。初始匹配由 [90] 提供。对于基于学习的模型，它们被用于过滤离群点，同时 RANSAC [21] 估计器用于最终的单应矩阵估计。可以看出，我们的方法在所有情况下均优于所有传统和基于学习的方法。例如，与次优的 CLNet 相比，NCMNet 在四个指标上分别获得了 2.06%、0.37%、0.94% 和 0.61% 的 mAA 提升。同时，实验结果表明，我们的 NCMNet+ 可以进一步提升性能。

4.3 消融实验

在本节中，我们通过构建消融实验来检验所提出的 NCMNet 在 YFCC100M [81] 未知场景上不同组件的性能。我们使用 mAP5° 和 mAP20° 作为评估方法的指标。

主要组件。在所提出的 NCMNet 中，采用了迭代剪枝策略 [28] 作为网络框架。为了验证剪枝模块中主要组件的效果，我们评估了它们相较于基线 [28] 的性能收益。SCE 层用于提取邻居内部的上下文信息，同时 CCI 层旨在探索邻居之间的交互。为了提高动态邻居的可靠性，我们使用了渐进邻居精化处理。而有序感知块有助于隐式地获取局部和全局上下文。各剪枝模块中主要组件的性能提升如表 8 所示。显而易见，随着 SCE 层和 CCI 层的逐步增加，模型性能逐渐提升。此外，从第 4 行和第 5 行的结果可以看出，采用渐进邻居精化处理和有序感知块是有效的。我们的 NCMNet（IPS + SCE + CCI + PNR + OA）能够实现最佳性能，验证了每个主要组件的有效性和合理性。

三种类型的邻居。在图 1 (c) 中，我们展示了三种类型邻居的可视化对比。为了进一步展示三种邻居的搜索区域，图 7 给出了所有内点的平均邻居搜索区域的定量结果。该结果是所有邻居覆盖的矩形区域与整幅图像面积之间的平均比例。

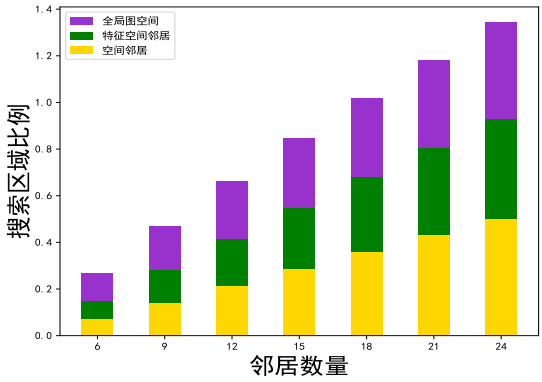


图 7. 关于不同邻居数量 k 的所有内点的平均邻居搜索区域比例（%）示意图。由于考虑了长距离依赖，全局图空间能够在更远的距离找到邻居。

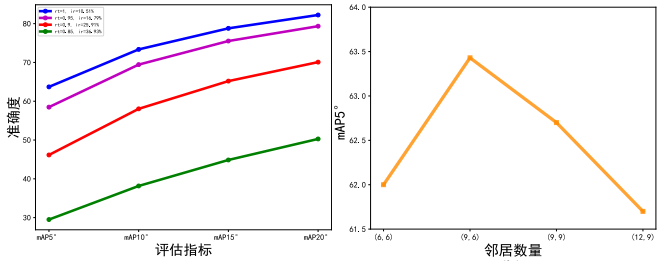


图 8. (a) 不同初始匹配中内点比例的影响。使用不同阈值的比例测试（rt）来获取具有不同内点比例的输入集。我们报告了在不同误差阈值下的 mAP。(b) 邻居嵌入中不同邻居数量的参数分析。$(\cdot,\cdot)$ 分别表示第一个和第二个剪枝模块中每个邻居嵌入中的邻居数量 k。展示了 mAP5°。

在这里，由于考虑了匹配之间的长距离依赖，内点的全局图邻居的搜索区域在不同的 k-近邻数量下相比其他邻居表现出更大的搜索范围。此外，为了展示三种邻居之间的互补性，我们在 NC 块中使用相同的邻居嵌入进行评估。对比结果如表 9 所示。显然，当同时使用三种类型的邻居时，网络能够实现最佳性能。

邻居上下文聚合。为了动态提取每个邻居嵌入的邻居上下文信息，我们的 SCE 层采用了一种分组卷积方法。因此，我们将其与一些经典的聚合方式进行对比，包括平均池化层、最大池化层和使用 $1\times k$ 核的卷积层，以证明这一设计的有效性。对比结果如表 10 所示，其中分组卷积策略优于其他竞争者，表明了其有效性。

输入的内点比例。传统方法的性能，例如 RANSAC [21] 及其变体 [83], [84], [49]，高度依赖于初始匹配中的内点比例（ir）。因此，我们分析了内点比例对 NCMNet 的影响，如图 8 (a) 所示。在描述子匹配过程中使用带有不同阈值的 Lowe 的比例测试（rt） [8]，我们构建了具有不同内点比例的初始匹配作为网络输入，并在相应的训练集上重新训练了 NCMNet。与传统方法相比，我们的方法即使在低内点比例下也表现出有效性。尽管比例测试在减少输入离群点方面表现出一定优势，但它也会导致许多重要的内点被舍弃，从而削弱整体性能。结果还表明，我们的网络在具有挑战性的条件下更具鲁棒性，即在初始匹配中内点足够多但离群点也很多的情况下。

表 10
SCE 层中不同上下文聚合方式的定量比较。“**Avg-pooling & MLPs**”使用平均池化层和两个带有 BN 和 ReLU 的连续 MLP 层来聚合邻居上下文。“**Max-pooling**”表示一个最大池化层。

	mAP5°	mAP20°
Avg-pooling & MLPs	61.48	81.53
Max-pooling & MLPs	62.75	81.86
$1\times k$ kernels Conv.	62.88	81.91
Grouped Conv.	**63.43**	**82.46**

表 11
三种类型邻居的不同组合策略。**SN**: 空间邻居。**FN**: 特征邻居。**GN**: 全局图邻居。**CCI**: 交叉上下文交互层。**RT**(ms): 平均运行时间。**FLOPs**(G): 每秒浮点运算次数。

baseline	SN	FN	GN	CCI	mAP5°	mAP20°	RT	FLOPs
✓					25.95	54.63	**5.01**	**0.86**
✓	✓				30.17	58.13	5.77	1.55
✓		✓			31.15	59.66	5.83	1.55
✓			✓		30.73	60.97	10.18	1.58
✓	✓	✓			33.22	61.66	6.51	2.31
✓	✓		✓		33.70	62.98	11.00	2.34
✓		✓	✓		33.83	62.60	10.89	2.34
✓	✓	✓	✓		36.03	63.96	11.63	3.10
✓	✓	✓	✓	✓	**37.83**	**65.94**	13.90	3.35

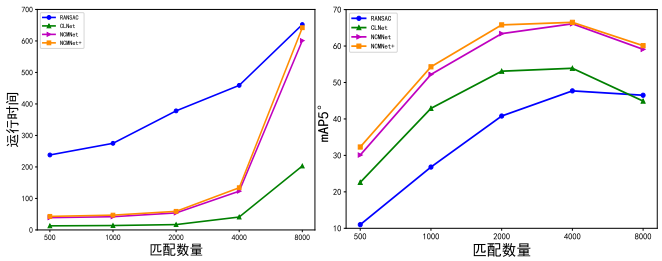
邻居的组合。在本研究中，我们提出为每个匹配寻找三种类型的邻居，以适应复杂的匹配情况。在此，我们通过比较三种类型邻居的不同组合来验证该设计的有效性和效率。我们选择 LFGC [19] 作为对比基线，该方法包含 12 个连续的 ResNet 块。我们将不同的邻居嵌入组合和 CCI 层插入基线的中间，以挖掘不同的邻居一致性信息。对于两种或三种类型的邻居上下文特征，我们直接采用通道维度上的连接操作来融合它们的信息。表 11 展示了性能收益和计算开销。可以看出，当使用一种或两种类型的邻居时，获得了较好的性能提升。当同时使用三种类型的邻居时，相比于基线的性能收益最佳，这进一步表明三种类型邻居是互补的。然而，我们发现全局图邻居的构建比其他两种邻居需要更多的运行时间。这是因为 GCN 操作需要较高的计算成本，特别是对于更多的输入匹配。根据上述观察，对于某些实时特征匹配应用，采用更高效的 GCN 策略是必要的，例如图聚类 [91]、拓扑采样 [92] 和流水线并行 [93]。未来，如何提高全局图构建的效率将是我们的主要关注点。

邻居数量 k 的分析。在 NC 块中，我们在不同的邻居搜索空间中为每个匹配寻找 k 最近邻，以构建邻居嵌入。一个合适的邻居数量 k 对于提取邻居上下文至关重要。图. 8 (b) 显示了不同邻居数量 k 组合的结果。我们的 NCMNet 在 $k=(9,6)$ 的组合下，相较于其他设置达到了最佳性能。因此，两个剪枝模块中的 k 分别设置为 9 和 6。

改进的有效性。基于之前的版本 [37]，我们提出了增强的全局图空间以及 CCI 层中的分组交叉注意力分支。前者利用匹配中固有的空间一致性来补充 [37] 中使用的特征一致性，其旨在通过空间和特征方面的密切关系构建一个全局连

表 12
所提出的两项改进的有效性。**EGS**: 增强的全局图空间。**GCA**: 分组交叉注意力分支。

	baseline	EGS	GCA	EGS+GCA
mAP5°	63.43	64.85	64.40	**65.83**
mAP20°	82.46	82.60	82.68	**83.14**



(a) (b)

图 9. 不同输入匹配数量带来的影响。采用 SIFT 和最近邻搜索获得 500/1000/2000/4000/8000 个匹配。展示了 (a) 平均运行时间 (ms) 和 (b) mAP5° (%)。

通图，以增强长距离依赖性。后者基于有效的分层分组方式探索丰富的邻居间信息交互，去置换 [37] 中的单一交叉注意力操作。我们选择 NCMNet [37] 作为基线，并在其基础上添加两个提出的操作。如表 12 所示，当配备了增强的全局图空间或分组交叉注意力分支时，基线模型的性能得到了进一步提升。增强的全局图空间能够提供更可靠的全局一致邻居。分组交叉注意力分支可以丰富三种邻居特征的信息整合。所提出的 NCMNet+ 实现了最佳性能，证明了这两项改进的有效性。

不同数量的输入。输入的匹配数量可能会因不同的估计方法而变化，这会影响剪枝工作的有效性和效率。因此，我们测试了由 SIFT 和最近邻搜索估计的不同数量匹配对性能和运行时间的影响。剪枝模型使用 2000 个匹配进行训练。如图 9 所示，随着输入数量的增加，由于潜在内点的增加，网络可能会获得更好的性能。同时，由于多次 GCN 操作带来的昂贵计算开销，运行时间也会显著增加。然而，如前述消融实验所提到的，我们认为这一问题在未来可以通过使用更高效的 GCN 策略 [91], [92], [93] 得到进一步缓解。

5 扩展任务

在许多基于特征匹配的任务中，两幅图像之间的准确特征匹配是重要的前提条件。在本节中，我们将把方法扩展到几个基于特征的重要任务中，包括遥感图像配准、点云配准、3D 重建和视觉定位。

5.1 遥感图像配准

图像配准旨在估计几何变换，并对齐源图像和目标图像之间的重叠区域。遥感图像配准 [94] 是一些任务（如图像融合、多光谱分类和变化检测）的关键过程，同样需要准确的特征匹配 [95], [96]。匹配剪枝可以为精确的图像配准提供可靠的内点。

表 13
遥感图像对的定量配准结果。使用平均 $RMSE$、MAE、MEE 和运行时间 (RT) 进行评估。↓ 表示值越低越好。

Method	$RMSE$↓	MEE↓	MAE↓	$RT(ms)$↓
RANSAC [21]	50.60	55.94	164.29	291.69
LFGC [19]	10.40	8.80	43.68	45.08
OANet++ [38]	7.17	8.79	34.10	73.80
MS^{2}DG-Net [29]	6.72	5.11	42.48	72.84
CLNet [28]	10.90	11.07	47.87	73.76
NCMNet [37]	1.55	0.01	23.88	73.47
NCMNet+	1.39	0.01	23.03	72.83

表 14
在 3DMatch [99] 数据集上的点云配准对比结果。使用平均 RR、IP 和 IR 进行评估。

Descriptor	FCGF			FPFH		
Method	RR	IP	IR	RR	IP	IR
RANSAC [21]	86.57	76.86	77.45	40.05	51.52	34.31
LFGC [19]	91.56	77.49	80.85	73.66	64.60	58.67
OANet++ [38]	91.87	77.76	80.49	72.51	62.62	55.96
MS^{2}DG-Net [29]	92.05	77.76	84.35	78.54	67.99	72.10
CLNet [28]	91.81	77.99	82.72	77.94	68.26	67.12
NCMNet [37]	92.54	78.28	84.70	78.60	69.64	70.13
NCMNet+	92.98	78.69	85.92	79.71	70.19	72.92

数据集。我们选择了由 [97], [98] 提供的 57 对低空遥感图像，这些图像涵盖了不同类型和场景。这些图像对提供了使用 SIFT 特征 [8] 生成的初始匹配，其中内点和离群点被手动标注。这些图像对涉及大幅度的视角变化以及极端的图像模式，因此不可避免地会遇到大量离群点。

评估。由于缺乏足够的数据用于训练网络，我们直接采用在 YFCC100M 数据集上训练的网络模型来测试其泛化能力。我们使用均方根误差（$RMSE$）、中位数误差（MEE）以及最大误差（MAE）作为评估指标，以衡量方法的配准性能。这些指标的公式如下：

$$RMSE = \sqrt{\frac{1}{L}\sum_{i=1}^{L}\|r_i - \mathcal{F}(s_i)\|_2^2}, \tag{22}$$

$$MEE = \text{median}\left\{\|r_i - \mathcal{F}(s_i)\|_2\right\}_{i=1}^{L}, \tag{23}$$

$$MAE = \max\left\{\|r_i - \mathcal{F}(s_i)\|_2\right\}_{i=1}^{L}, \tag{24}$$

其中，r_i 和 s_i 分别表示参考图像和感知图像的关键点坐标。$\mathcal{F}$ 是通过方法估计出的两幅匹配图像之间的变换函数。L 表示关键点坐标的数量。$|\cdot|_2$ 表示向量的欧式范数。median$(\cdot)$ 和 max$(\cdot)$ 分别计算对应的中位数和最大值。

结果。表 13 给出了一些匹配剪枝方法的定量结果，其中运行时间表示在获得精炼的匹配后进行配准的时间。使用

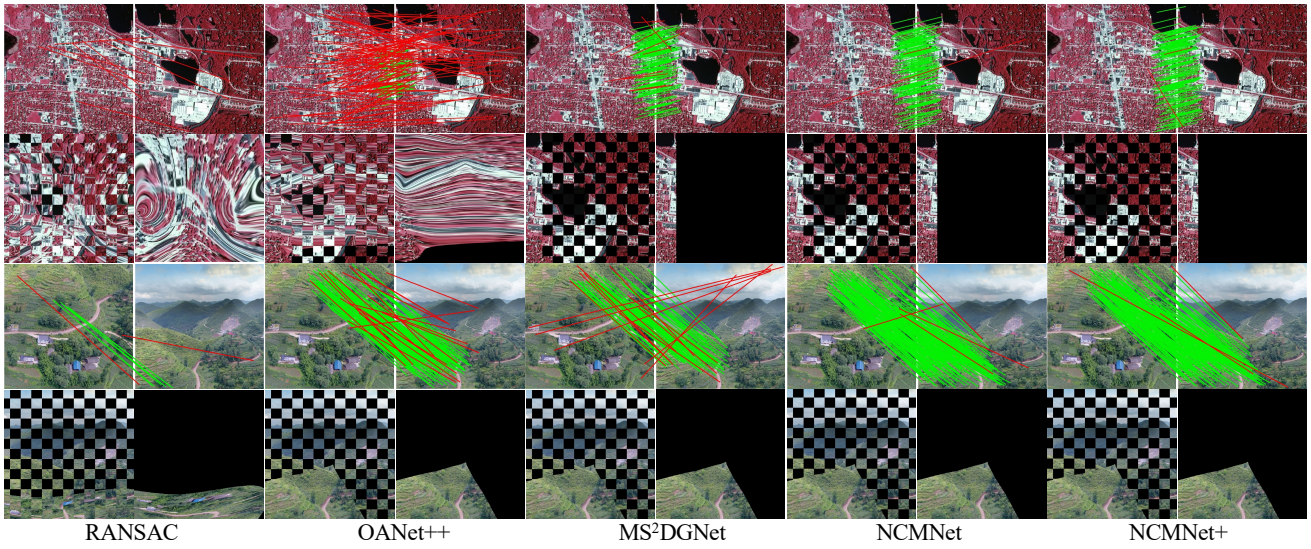


图 10. 匹配剪枝和图像配准的可视化结果。从左到右列依次为：RANSAC [21]、OANet++[38]、MS^{2}DG-Net[29]、NCMNet [37] 和 NCMNet+。在第 1 和第 3 行中，红色线表示离群点，绿色线表示各方法保留的内点。在第 2 和第 4 行中，左侧展示棋盘图像，右侧展示变换后的图像。

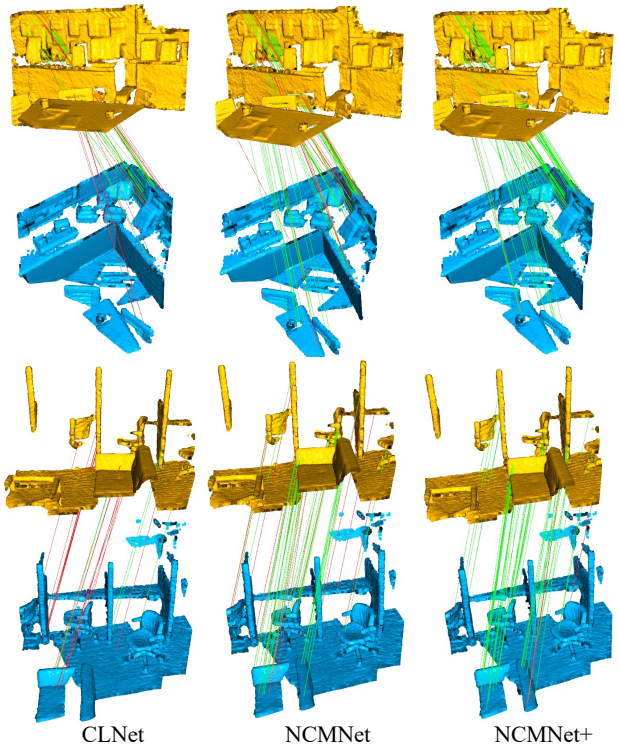


图 11. 3DMatch[99] 上匹配剪枝的可视化结果。从左到右列分别为：CLNet [28]、NCMNet [37] 和 NCMNet+。绿色和红色线分别表示识别出的内点和离群点。

了 1000 次迭代的 RANSAC 进行对比。对于基于学习的方法，我们首先利用网络模型过滤离群点，然后采用 50 次迭代的 RANSAC 来进一步处理保留的匹配。与这些竞争者相比，我们的 NCMNet 和 NCMNet+ 展现了更优越的性能。NCMNet+ 在 $RMSE$、MAE 和 MEE 方面取得了最优结果。此外，图 10 中展示了两个典型图像对上的匹配剪枝和图像配准的可视化结果。可以看到，所提出的方法能够获得更

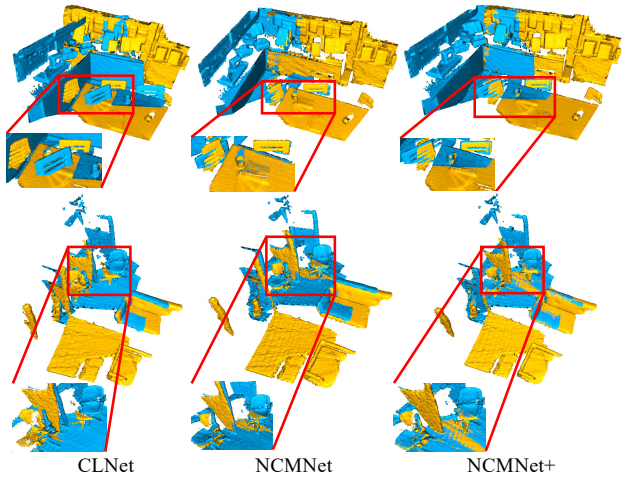


图 12. 3DMatch[99] 上点云配准的可视化结果。从左到右列分别为：CLNet [28]、NCMNet [37] 和 NCMNet+。

准确的结果，并且比其他方法保留更多的内点。

5.2 点云配准

点云配准 [100], [101], [102] 需要确定用于对齐一对点云的最优姿态变换，这是点云处理中的一个关键问题。类似于图像特征匹配，它可以通过建立可靠的点对点特征匹配来解决，但由于 3D 特征提取方法的局限性和重叠区域的有限性，离群点是不可避免的。因此，对于处理含有大量离群点的 3D 匹配，剪枝匹配是不可或缺的步骤之一。

数据集。由于匹配维度的差异，我们使用室内 3DMatch [99] 数据集来重新训练和测试网络模型。在测试集中，包含来自八个不同场景的 1,623 个部分重叠的点云片段。

评估。我们使用可学习的全卷积几何特征（FCGF）[103] 和传统的快速点特征直方图（FPFH）[104] 作为特征提取方

表 15
不同规模数据集上 3D 重建的定量对比结果。R+M: 比例测试 + 相互检查

数据集	方法	Reg	Sparse	Dense	TL	Obs	Reproj↓
Fountain	OANet++ [38]	11	12456	306598	4.88	5535	0.45px
	MS^{2}DG-Net [29]	11	10584	313808	4.86	4681	0.45px
	CLNet [28]	11	12029	320722	4.87	5326	0.46px
	NCMNet [37]	11	12319	327812	4.85	5609	0.46px
	NCMNet+	11	14354	373133	5.01	6331	0.43px
Herzjesu	OANet++ [38]	8	7406	278702	4.24	3944	0.49px
	MS^{2}DG-Net [29]	8	7229	254100	4.22	3815	0.48px
	CLNet [28]	8	7498	245425	4.25	3981	0.49px
	NCMNet [37]	8	7539	261289	4.27	3998	0.50px
	NCMNet+	8	7773	282455	4.31	4146	0.49px
South-Building	OANet++ [38]	126	128245	2350753	5.36	5456	0.58px
	MS^{2}DG-Net [29]	127	110953	2297649	5.65	4940	0.57px
	CLNet [28]	127	119784	2195264	5.53	5216	0.58px
	NCMNet [37]	128	136360	2199067	5.26	5608	0.59px
	NCMNet+	128	121082	2386528	5.68	5277	0.58px
Gendarmenmarkt	OANet++ [38]	958	334503	1512576	5.56	1942	0.79px
	MS^{2}DG-Net [29]	976	324324	1358179	5.73	1882	0.79px
	CLNet [28]	973	333665	1238846	5.71	1950	0.78px
	NCMNet [37]	970	349821	1391622	5.45	2057	0.80px
	NCMNet+	1006	368701	1635901	5.71	2001	0.78px
Alamo	OANet++ [38]	806	219696	1701979	11.34	3020	0.72px
	MS^{2}DG-Net [29]	858	220835	2036959	11.29	2907	0.72px
	CLNet [28]	859	221539	2954842	11.13	2857	0.70px
	NCMNet [37]	838	240389	2833031	10.61	3034	0.74px
	NCMNet+	865	257033	3147980	11.68	3064	0.72px

法来构建初始匹配。按照 [105] 的训练设置，网络在 FCGF 上训练 50 轮，然后在 FCGF 和 FPFH 上进行测试。我们采用点云配准中最重要的标准——配准召回率（RR）来评估性能。它表示成功对齐的比例，要求旋转误差必须低于 30cm，平移误差必须小于 20°。同时，内点精度（IP：识别出的内点与保留匹配的比例）和内点召回率（IR：识别出的内点与实际内点的比例）用于评估匹配剪枝的性能。

结果。两种设置下的定量对比结果如表 14 所示。这里，RANSAC 使用了 1000 次迭代进行对比。可以看到，我们的 NCMNet+ 在两种设置下都能够获得最佳的 IP 和 IR，这表明了出色的匹配剪枝性能。在 RR 方面，由于更好的匹配结果，所提出的 NCMNet+ 超过了所有对比方法。此外，图 11 和图 12 分别展示了一些方法在匹配剪枝和配准上的可视化结果，进一步证明了我们方法的有效性。

5.3 3D 重建

3D 重建的过程需要使用多张 2D 图像来恢复物体或场景的 3D 模型 [106]。通常，重建的性能很大程度上依赖于 2D 图像对中匹配的质量。因此，我们评估了网络在 3D 重建任务中的泛化能力。

数据集。按照 [107] 的方法，我们使用 ***COLMAP*** [3] 平台进行一系列 3D 重建实验。测试数据集包含 YFCC100M [81] 中的一些小规模和中等规模子集。具体来说，小规模子集包括 Fountain（11 张图像）、Herzjesu（8 张图像）和 South-Building（128 张图像），这些图像通过枚举方式选择相似图像。根据 [108]，在中等规模子集中，使用词袋模型（BoW）搜索前 20 张相似图像，这些子集包括 Gendarmenmarkt（1,463 张图像）和 Alamo（2,915 张图像）。

评估。所有基于学习的模型均使用 YFCC100M 数据集进行训练，并使用 SIFT 提取的 4,000 个关键点。我们报告了

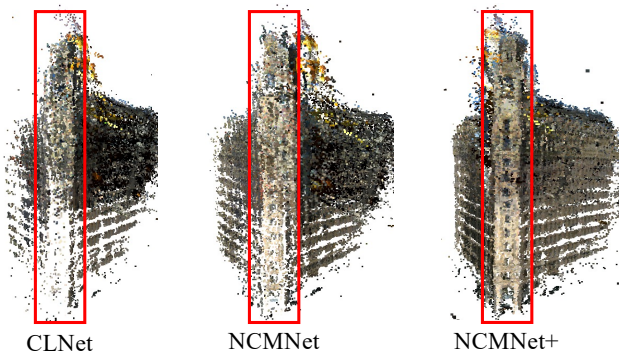


图 13. Alamo 数据集的 3D 重建可视化结果。从左到右分别为：CLNet [28]、NCMNet [37] 和 NCMNet+ 的结果。

表 16
Aachen Day-Night v1.1 数据集 [110], [111] 上视觉定位的定量对比结果。展示了在不同阈值下正确定位的查询图像的百分比。

Methods	Day (0.25m, 2°) / (0.5m, 5°) /(5m, 10°)	Night (0.25m, 2°) / (0.5m, 5°) /(5m, 10°)
SIFT [8]	65.3 / 72.0 / 78.9	16.8 / 19.9 / 27.7
LFGC [19]	77.4 / 82.8 / 86.9	30.9 / 34.6 / 41.4
OANet++ [38]	79.1 / 84.8 / 89.0	33.0 / 37.2 / 45.5
MS^{2}DG-Net [29]	76.8 / 83.3 / 86.9	26.2 / 31.4 / 42.4
CLNet [28]	82.8 / 89.4 / 93.4	39.8 / 50.8 / 61.8
NCMNet [37]	82.8 / 91.1 / 95.0	43.5 / 53.9 / 69.1
NCMNet+	84.2 / 92.5 / 96.0	48.2 / 59.7 / 75.4

配准图像数量（Reg）、稀疏点数量（$Sparse$）、稠密点数量（$Dense$）、平均轨迹长度（TL）和重投影误差（$Reproj$），以评估模型在提升 3D 重建匹配质量方面的表现。其中，$Sparse$ 和 $Dense$ 是主要的评估指标。

结果。3D 重建的定量对比结果如表 15 所示。这些基于学习的匹配剪枝方法被用于提升匹配的质量。与其他方法相比，我们的方法始终表现出更好的性能。同时，如图 13 所示，Alamo 数据集上的可视化结果进一步证明了我们方法的有效性。尤其是在红框部分，我们的 NCMNet+ 提供了比其他两种方法更完整的重建结果。

5.4 视觉定位

视觉定位旨在根据参考场景（如 3D 场景模型）的视觉表示来估计查询图像的 6 自由度（DoF）相机位姿 [7], [109]。基于 3D 结构的视觉定位过程需要利用局部特征生成 2D-3D 匹配，以进行位姿估计。为了进一步识别足够且准确的匹配，剪枝匹配是必要的，并且这对视觉定位的成功与否也是至关重要的。

数据集。我们在 Aachen Day-Night v1.1 数据集 [110], [111] 上进行实验，该数据集主要关注光照剧烈变化下的定位问题。数据集包括 6,697 张白天参考图像、824 张白天查询图像和 191 张夜间查询图像，这些图像均由移动设备拍摄。我们在长期视觉定位基准测试 [112] 上评估性能。

评估。通过网络模型获得的匹配结果被集成到开源定位管道 HLoc [113] 中。具体来说，我们在查询图像和参考图像之间使用 SIFT 构建 2,000 个初始匹配，然后使用匹配剪枝方法获取可靠的匹配。我们利用 ***COLMAP*** [3] 对参考的 3D SfM 模型进行三角化并恢复其位姿。网络模型在 YFCC100M 数据集上使用 2,000 个 SIFT 关键点进行训练。. 我们使用在不同阈值下正确定位查询的比例作为评估指标。

结果。视觉定位的定量对比结果如表 16 所示。可以看到，我们的 NCMNet+ 在不同阈值下的白天和夜间场景中均显著优于所有基线方法。在光照剧烈变化条件下，我们的方法能够提供比竞争对手更优越的结果，展示了其在具有挑战性的视觉定位任务中的优越性和鲁棒性。

6 结论

在这项工作中，我们提出了一种名为邻居一致性挖掘网络（NCMNet）的有效架构，用于应对具有挑战性的匹配剪枝任务。我们开发了一个全局图空间，通过全局连通图显式地建模匹配之间的长距离密切关系来搜索一致的邻居。同时，我们设计了一个邻居一致性块，渐近地挖掘三种类型邻居的一致性，以增强方法的鲁棒性。此外，我们引入了空间一致性来提高全局图空间的可靠性，并采用分层分组方式丰富邻居间信息的融合。我们在不同基准测试和扩展任务上进行了全面的实验，验证了 NCMNet 和 NCMNet+ 的有效性和泛化能力，证明了其相较于现有方法的显著优势。

致谢

这项工作得到了如下机构的资助：天津市自然科学基金（编号：20JCJQJC00020），国家自然科学基金（编号：U22B2049, 62302240），中央高校基本科研业务费专项基金，以及南开大学超算中心（NKSC）。

参考文献

[1] X. Liu, R. Qin, J. Yan, and J. Yang, "Ncmnet: Neighbor consistency mining network for two-view correspondence pruning," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1–19, 2024.

[2] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "Orb-slam: a versatile and accurate monocular slam system," IEEE Transactions on Robotics, vol. 31, no. 5, pp. 1147–1163, 2015.

[3] J. L. Schonberger and J.-M. Frahm, "Structure-from-motion revisited," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4104–4113.

[4] G. Xiao, H. Wang, J. Ma, and D. Suter, "Segmentation by continuous latent semantic analysis for multi-structure model fitting," International Journal of Computer Vision, vol. 129, no. 7, pp. 2034–2056, 2021.

[5] Z. Yang, Y. Yang, K. Yang, and Z.-Q. Wei, "Non-rigid image registration with dynamic gaussian component density and space curvature preservation," IEEE Transactions on Image Processing, vol. 28, no. 5, pp. 2584–2598, 2018.

[6] G. Xiao, J. Ma, S. Wang, and C. Chen, "Deterministic model fitting by local-neighbor preservation and global-residual optimization," IEEE Transactions on Image Processing, vol. 29, no. 4, pp. 8988–9001, 2020.

[7] C. Toft, W. Maddern, A. Torii, L. Hammarstrand, E. Stenborg, D. Safari, M. Okutomi, M. Pollefeys, J. Sivic, T. Pajdla et al., "Long-term visual localization revisited," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 4, pp. 2074–2088, 2020.

[8] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.

[9] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in Proceedings of the European Conference on Computer Vision, 2006, pp. 404–417.

[10] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in Proceedings of the IEEE International Conference on Computer Vision, 2011, pp. 2564–2571.

[11] K. M. Yi, E. Trulls, V. Lepetit, and P. Fua, "Lift: Learned invariant feature transform," in Proceedings of the European Conference on Computer Vision, 2016, pp. 467–483.

[12] D. DeTone, T. Malisiewicz, and A. Rabinovich, "Superpoint: Self-supervised interest point detection and description," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 224–236.

[13] J. Chang, J. Yu, and T. Zhang, "Structured epipolar matcher for local feature matching," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2023, pp. 6176–6185.

[14] P. Lindenberger, P.-E. Sarlin, and M. Pollefeys, "Lightglue: Local feature matching at light speed," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 17 627–17 638.

[15] R. Hartley and A. Zisserman, Multiple view geometry in computer vision. Cambridge University Press, 2003.

[16] W.-Y. D. Lin, M.-M. Cheng, J. Lu, H. Yang, M. N. Do, and P. Torr, "Bilateral functions for global motion modeling," in Proceedings of the European Conference on Computer Vision, 2014, pp. 341–356.

[17] Y. Liu, L. Liu, C. Lin, Z. Dong, and W. Wang, "Learnable motion coherence for correspondence pruning," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 3237–3246.

[18] J. Bian, W.-Y. Lin, Y. Matsushita, S.-K. Yeung, T.-D. Nguyen, and M.-M. Cheng, "Gms: Grid-based motion statistics for fast, ultra-robust feature correspondence," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4181–4190.

[19] K. M. Yi, E. Trulls, Y. Ono, V. Lepetit, M. Salzmann, and P. Fua, "Learning to find good correspondences," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 2666–2674.

[20] G. Wang and Y. Chen, "Local consensus transformer for correspondence learning," in IEEE International Conference on Multimedia and Expo, 2023, pp. 1151–1156.

[21] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," Communications of the ACM, vol. 24, no. 6, pp. 381–395, 1981.

[22] P. H. Torr and A. Zisserman, "Mlesac: A new robust estimator with application to estimating image geometry," Computer Vision and Image Understanding, vol. 78, no. 1, pp. 138–156, 2000.

[23] P. J. Rousseeuw and A. M. Leroy, Robust regression and outlier detection. John wiley & sons, 2005.

[24] R. Raguram, O. Chum, M. Pollefeys, J. Matas, and J.-M. Frahm, "Usac: A universal framework for random sample consensus," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 8, pp. 2022–2038, 2012.

[25] L. Zhou, S. Zhu, Z. Luo, T. Shen, R. Zhang, M. Zhen, T. Fang, and L. Quan, "Learning and matching multi-view descriptors for registration of point clouds," in Proceedings of the European Conference on Computer Vision, 2018, pp. 505–522.

[26] H.-Y. Chen, Y.-Y. Lin, and B.-Y. Chen, "Co-segmentation guided hough transform for robust feature matching," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 12, pp. 2388–2401, 2015.

[27] L. Cavalli, V. Larsson, M. R. Oswald, T. Sattler, and M. Pollefeys, "Handcrafted outlier detection revisited," in Proceedings of the European Conference on Computer Vision, 2020, pp. 770–787.

[28] C. Zhao, Y. Ge, F. Zhu, R. Zhao, H. Li, and M. Salzmann, "Progressive correspondence pruning by consensus learning," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 6464–6473.

[29] L. Dai, Y. Liu, J. Ma, L. Wei, T. Lai, C. Yang, and R. Chen, "Ms2dg-net: Progressive correspondence learning via multiple sparse semantics dynamic graph," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 8973–8982.

[30] C. Zhao, Z. Cao, C. Li, X. Li, and J. Yang, "Nm-net: Mining reliable neighbors for robust feature correspondences," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 215–224.

[31] W.-Y. Lin, F. Wang, M.-M. Cheng, S.-K. Yeung, P. H. Torr, M. N. Do, and J. Lu, "Code: Coherence based decision boundaries for feature correspondence," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 1, pp. 34–47, 2017.

[32] M. Cho and K. M. Lee, "Progressive graph matching: Making a move of graphs via probabilistic voting," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 398–405.

[33] C. Liu, S. Zhang, X. Yang, and J. Yan, "Self-supervised learning of visual graph matching," in Proceedings of the European Conference on Computer Vision, 2022, pp. 370–388.

[34] P. C. Lusk, K. Fathian, and J. P. How, "Clipper: A graph-theoretic framework for robust data association," in IEEE International Conference on Robotics and Automation, 2021, pp. 13 828–13 834.

[35] R. Wang, Z. Guo, S. Jiang, X. Yang, and J. Yan, "Deep learning of partial graph matching via differentiable top-k," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 6272–6281.

[36] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," arXiv preprint arXiv:1609.02907, 2016.

[37] X. Liu and J. Yang, "Progressive neighbor consistency mining for correspondence pruning," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 9527–9537.

[38] J. Zhang, D. Sun, Z. Luo, A. Yao, L. Zhou, T. Shen, Y. Chen, L. Quan, and H. Liao, "Learning two-view correspondences and geometry using order-aware network," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 5845–5854.

[39] C. Liu, J. Yuen, and A. Torralba, "Sift flow: Dense correspondence across scenes and its applications," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 5, pp. 978–994, 2010.

[40] C. B. Choy, J. Gwak, S. Savarese, and M. Chandraker, "Universal correspondence network," Advances in Neural Information Processing Systems, pp. 2406–2414, 2016.

[41] I. Rocco, M. Cimpoi, R. Arandjelović, A. Torii, T. Pajdla, and J. Sivic, "Neighbourhood consensus networks," Advances in Neural Information Processing Systems, pp. 1658–1669, 2018.

[42] J. Sun, Z. Shen, Y. Wang, H. Bao, and X. Zhou, "Loftr: Detector-free local feature matching with transformers," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 8922–8931.

[43] Q. Zhou, T. Sattler, and L. Leal-Taixe, "Patch2pix: Epipolar-guided pixel-level correspondences," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 4669–4678.

[44] R. Wang, J. Yan, and X. Yang, "Combinatorial learning of robust deep graph matching: an embedding based approach," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 6, pp. 6984–7000, 2020.

[45] J. Lee, B. Kim, S. Kim, and M. Cho, "Learning rotation-equivariant features for visual correspondence," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 21 887–21 897.

[46] J. Ma, X. Jiang, A. Fan, J. Jiang, and J. Yan, "Image matching from handcrafted to deep features: A survey," International Journal of Computer Vision, vol. 129, no. 1, pp. 23–79, 2021.

[47] Y. Jin, D. Mishkin, A. Mishchuk, J. Matas, P. Fua, K. M. Yi, and E. Trulls, "Image matching across wide baselines: From paper to practice," International Journal of Computer Vision, vol. 129, no. 2, pp. 517–547, 2021.

[48] D. Nistér, "An efficient solution to the five-point relative pose problem," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 6, pp. 756–770, 2004.

[49] D. Barath, J. Matas, and J. Noskova, "Magsac: marginalizing sample consensus," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 10 197–10 205.

[50] E. Brachmann, A. Krull, S. Nowozin, J. Shotton, F. Michel, S. Gumhold, and C. Rother, "Dsac-differentiable ransac for camera localization," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 6684–6692.

[51] E. Brachmann and C. Rother, "Neural-guided ransac: Learning where to sample model hypotheses," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 4322–4331.

[52] D. Barath, L. Cavalli, and M. Pollefeys, "Learning to find good models in ransac," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 15 744–15 753.

[53] T. Wei, Y. Patel, A. Shekhovtsov, J. Matas, and D. Barath, "Generalized differentiable ransac," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 17 649–17 660.

[54] T. Wei, J. Matas, and D. Barath, "Adaptive reordering sampler with neurally guided magsac," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 18 163–18 173.

[55] C. Nie, G. Wang, Z. Liu, L. Cavalli, M. Pollefeys, and H. Wang, "Rlsac: Reinforcement learning enhanced sample consensus for end-to-end robust estimation," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 9891–9900.

[56] C. Zhao, Z. Cao, J. Yang, K. Xian, and X. Li, "Image feature correspondence selection: a comparative study and a new contribution," IEEE Transactions on Image Processing, vol. 29, no. 2, pp. 3506–3519, 2020.

[57] X. Jiang, J. Ma, G. Xiao, Z. Shao, and X. Guo, "A review of multimodal image matching: Methods and applications," Information Fusion, vol. 73, pp. 22–71, 2021.

[58] R. Ranftl and V. Koltun, "Deep fundamental matrix estimation," in Proceedings of the European Conference on Computer Vision, 2018, pp. 284–299.

[59] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 652–660.

[60] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," in Advances in Neural Information Processing Systems, 2017, pp. 5099–5108.

[61] Z. Zhong, G. Xiao, L. Zheng, Y. Lu, and J. Ma, "T-net: Effective permutation-equivariant network for two-view correspondence learning," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 1950–1959.

[62] S. Zhang and J. Ma, "Convmatch: Rethinking network design for two-view correspondence learning," in AAAI Conference on Artificial Intelligence, 2023, pp. 3472–3479.

[63] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in Advances in Neural Information Processing Systems, 2017, pp. 5998–6008.

[64] W. Sun, W. Jiang, E. Trulls, A. Tagliasacchi, and K. M. Yi, "Acne: Attentive context normalization for robust permutation-equivariant learning," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020, pp. 11 286–11 295.

[65] X. Ye, W. Zhao, H. Lu, and Z. Cao, "Learning second-order attentive context for efficient correspondence pruning," in AAAI Conference on Artificial Intelligence, 2023, pp. 3250–3258.

[66] L. Zheng, G. Xiao, Z. Shi, S. Wang, and J. Ma, "Msa-net: Establishing reliable correspondences by multiscale attention network," IEEE Transactions on Image Processing, vol. 31, pp. 4598–4608, 2022.

[67] X. Liu, G. Xiao, R. Chen, and J. Ma, "Pgfnet: Preference-guided filtering network for two-view correspondence learning," IEEE Transactions on Image Processing, vol. 32, pp. 1367 – 1378, 2023.

[68] A. Myronenko, X. Song, and M. Carreira-Perpinan, "Non-rigid point set registration: Coherent point drift," Advances in Neural Information Processing Systems, pp. 1009–1016, 2006.

[69] Y. Liu, B. N. Zhao, S. Zhao, and L. Zhang, "Progressive motion coherence for remote sensing image matching," IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1–13, 2022.

[70] J. Ma, J. Zhao, J. Jiang, H. Zhou, and X. Guo, "Locality preserving matching," International Journal of Computer Vision, vol. 127, no. 5, pp. 512–531, 2019.

[71] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," International journal of computer vision, vol. 60, pp. 63–86, 2004.

[72] P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich, "Superglue: Learning feature matching with graph neural networks," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 4938–4947.

[73] M. Leordeanu and M. Hebert, "A spectral technique for correspondence problems using pairwise constraints," in Proceedings of the IEEE International Conference on Computer Vision, 2005, pp. 1482–1489.

[74] J. You, J. Leskovec, K. He, and S. Xie, "Graph structure of neural networks," in International Conference on Machine Learning, 2020, pp. 10 881–10 891.

[75] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip, "A comprehensive survey on graph neural networks," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 1, pp. 4–24, 2020.

[76] L. Wu, Y. Chen, K. Shen, X. Guo, H. Gao, S. Li, J. Pei, B. Long et al., "Graph neural networks for natural language processing: A survey," Foundations and Trends® in Machine Learning, vol. 16, no. 2, pp. 119–328, 2023.

[77] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph cnn for learning on point clouds," Acm Transactions on Graphics, vol. 38, no. 5, pp. 1–12, 2019.

[78] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in International Conference on Machine Learning, 2015, pp. 448–456.

[79] S.-H. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. Torr, "Res2net: A new multi-scale backbone architecture," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 2, pp. 652–662, 2019.

[80] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.

[81] B. Thomee, D. A. Shamma, G. Friedland, B. Elizalde, K. Ni, D. Poland, D. Borth, and L.-J. Li, "Yfcc100m: The new data in multimedia research," Communications of the ACM, vol. 59, no. 2, pp. 64–73, 2016.

[82] J. Xiao, A. Owens, and A. Torralba, "Sun3d: A database of big spaces reconstructed using sfm and object labels," in Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 1625–1632.

[83] O. Chum, T. Werner, and J. Matas, "Two-view geometry estimation unaffected by a dominant plane," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, 2005, pp. 772–779.

[84] D. Barath and J. Matas, "Graph-cut ransac," in Proceedings

of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6733–6741.

[85] D. Barath, J. Noskova, M. Ivashechkin, and J. Matas, "Magsac++, a fast, reliable and accurate robust estimator," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 1304–1312.

[86] J. Edstedt, I. Athanasiadis, M. Wadenbäck, and M. Felsberg, "Dkm: Dense kernelized feature matching for geometry estimation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 2023, pp. 17 765–17 775.

[87] A. Dai, A. X. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Nießner, "Scannet: Richly-annotated 3d reconstructions of indoor scenes," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 5828–5839.

[88] P. Truong, M. Danelljan, L. Van Gool, and R. Timofte, "Learning accurate dense correspondences and when to trust them," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 5714–5724.

[89] H. Chen, Z. Luo, L. Zhou, Y. Tian, M. Zhen, T. Fang, D. Mckinnon, Y. Tsin, and L. Quan, "Aspanformer: Detector-free image matching with adaptive span transformer," in Proceedings of the European Conference on Computer Vision, 2022, pp. 20–36.

[90] D. Barath, D. Mishkin, M. Polic, W. Förstner, and J. Matas, "A large-scale homography benchmark," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 21 360–21 370.

[91] W.-L. Chiang, X. Liu, S. Si, Y. Li, S. Bengio, and C.-J. Hsieh, "Cluster-gcn: An efficient algorithm for training deep and large graph convolutional networks," in Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, pp. 257–266.

[92] H. Li, M. Wang, S. Liu, P.-Y. Chen, and J. Xiong, "Generalization guarantee of training graph convolutional networks with graph topology sampling," in International Conference on Machine Learning, 2022, pp. 13 014–13 051.

[93] C. Wan, Y. Li, C. R. Wolfe, A. Kyrillidis, N. S. Kim, and Y. Lin, "PipeGCN: Efficient full-graph training of graph convolutional networks with pipelined feature communication," in The International Conference on Learning Representations, 2022, pp. 1–12.

[94] Y. Wu, J.-W. Liu, C.-Z. Zhu, Z.-F. Bai, Q.-G. Miao, W.-P. Ma, and M.-G. Gong, "Computational intelligence in remote sensing image registration: A survey," International Journal of Automation and Computing, vol. 18, pp. 1–17, 2021.

[95] B. Zitova and J. Flusser, "Image registration methods: a survey," Image and Vision Computing, vol. 21, no. 11, pp. 977–1000, 2003.

[96] S. Chen, J. Chen, Y. Rao, X. Chen, X. Fan, H. Bai, L. Xing, C. Zhou, and Y. Yang, "A hierarchical consensus attention network for feature matching of remote sensing images," IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1–11, 2022.

[97] X. Jiang, J. Jiang, A. Fan, Z. Wang, and J. Ma, "Multiscale locality and rank preservation for robust feature matching of remote sensing images," IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 9, pp. 6462–6472, 2019.

[98] X. Jiang, J. Ma, A. Fan, H. Xu, G. Lin, T. Lu, and X. Tian, "Robust feature matching for remote sensing image registration via linear adaptive filtering," IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 2, pp. 1577–1591, 2020.

[99] A. Zeng, S. Song, M. Nießner, M. Fisher, J. Xiao, and T. Funkhouser, "3dmatch: Learning local geometric descriptors from rgb-d reconstructions," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1802–1811.

[100] H. Wang, Y. Liu, Q. Hu, B. Wang, J. Chen, Z. Dong, Y. Guo, W. Wang, and B. Yang, "Roreg: Pairwise point cloud registration with oriented descriptors and local rotations," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, pp. 10 376 – 10 393, 2023.

[101] X. Zhang, J. Yang, S. Zhang, and Y. Zhang, "3d registration with maximal cliques," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 17 745–17 754.

[102] Y. Wu, Y. Zhang, W. Ma, M. Gong, X. Fan, M. Zhang, A. Qin, and Q. Miao, "Rornet: Partial-to-partial registration network with reliable overlapping representations," IEEE Transactions on Neural Networks and Learning Systems, 2023.

[103] C. Choy, J. Park, and V. Koltun, "Fully convolutional geometric features," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 8958–8966.

[104] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (fpfh) for 3d registration," in IEEE International Conference on Robotics and Automation, 2009, pp. 3212–3217.

[105] X. Bai, Z. Luo, L. Zhou, H. Chen, L. Li, Z. Hu, H. Fu, and C.-L. Tai, "Pointdsc: Robust point cloud registration using deep spatial consistency," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 15 859–15 869.

[106] A. Schmied, T. Fischer, M. Danelljan, M. Pollefeys, and F. Yu, "R3d3: Dense 3d reconstruction of dynamic scenes from multiple cameras," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 3216–3226.

[107] J. Zhang, D. Sun, Z. Luo, A. Yao, H. Chen, L. Zhou, T. Shen, Y. Chen, L. Quan, and H. Liao, "Oanet: Learning two-view correspondences and geometry using order-aware network," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 6, pp. 3110–3122, 2022.

[108] M. Dusmanu, J. L. Schönberger, and M. Pollefeys, "Multi-view optimization of local feature geometry," in Proceedings of the European Conference on Computer Vision, 2020, pp. 670–686.

[109] V. Panek, Z. Kukelova, and T. Sattler, "Visual localization using imperfect 3d models from the internet," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 13 175–13 186.

[110] Z. Zhang, T. Sattler, and D. Scaramuzza, "Reference pose generation for long-term visual localization via learned features and view synthesis," International Journal of Computer Vision, vol. 129, pp. 821–844, 2021.

[111] T. Sattler, T. Weyand, B. Leibe, and L. Kobbelt, "Image retrieval for image-based localization revisited." in British Machine Vision Conference, vol. 1, no. 2, 2012, pp. 4–15.

[112] C. Toft, W. Maddern, A. Torii, L. Hammarstrand, E. Stenborg, D. Safari, M. Okutomi, M. Pollefeys, J. Sivic, T. Pajdla, F. Kahl, and T. Sattler, "Long-term visual localization revis-

ited," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 4, pp. 2074–2088, 2022.

[113] P.-E. Sarlin, C. Cadena, R. Siegwart, and M. Dymczyk, "From coarse to fine: Robust hierarchical localization at large scale," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 12 716–12 725.