

热点追踪

基于极大团的三维配准方法

西北工业大学 张曦予 杨佳琪* 张世坤 张艳宁

本文是西北工业大学团队解读其在 CVPR 2023 获得最佳学生论文奖的工作 MAC。论文研究的问题是三维点云配准 (PCR)，旨在寻找最佳姿态以对齐点云对。由于现有的 3D 关键点检测器和描述符的局限性，点云之间的有限重叠以及数据噪声，特征匹配通常包含外点，对精确的点云配准产生极大挑战。我们提出了一种基于极大团 (MAC) 的三维配准方法。首先，我们提出了一种名为 MAC 的假设生成方法。与以前的最大团约束相比，我们的 MAC 方法能够在图中挖掘更多的局部信息。即使存在严重的异常值，由 MAC 生成的假设也具有很高的准确性。其次，基于 MAC 我们提出了一种新的点云配准方法，该方法在 U3M、3DMatch、3DLoMatch、KITTI 和 ETH 数据集上实现了最先进的性能。值得注意的是，我们的仅凭几何的 MAC 方法优于几种最先进的深度学习框架^[1-4]。MAC 还可以作为模块插入多个深度学习框架^[5-9]中，以提高它们的性能。

一、研究背景

我们研究的是计算机视觉中的一个经典问题——三维点云配准：寻找最佳姿态以对齐点云对。为了解决三维点云配准问题，有传统几何与深度学习两类方法。

传统几何方法^[10-14] RANSAC 从初始匹配集中迭代地采样匹配，为每个子集生成和评估几何估计，直到得到最佳的变换。尽管基于 RANSAC 的方法简单高效，但在高外点率情况下，它们的性能非常脆弱，并且需要大量迭代才能获得可接受的结果。为了解决当前评估指标耗时长以及对噪声敏感的问题，Yang 等人^[15]在计算过程中分析了内点和外点的贡献，并提出了几种可以有效提高 RANSAC 配准性能的指标。最近也出现了一系

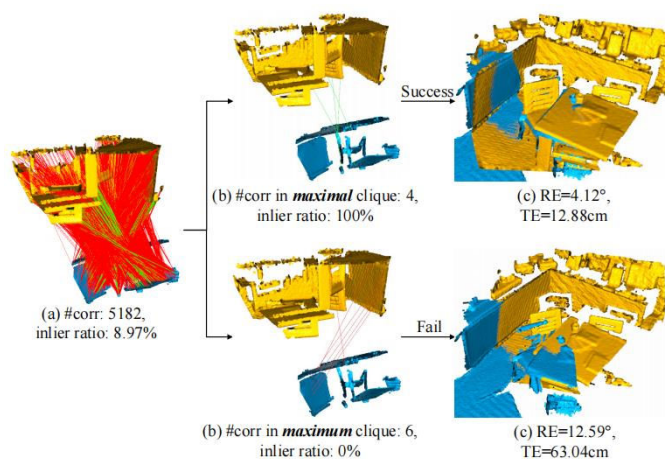


图 1 极大团与最大团在低重叠率点云数据上的区别

列基于 BnB 的全局配准方法。Bustos 和 Chin^[16]提出了保证的异常值去除 (GORE)，它计算每个匹配的紧密下界和紧密上界，并通过拒绝真正的外点来减少匹配集的大小。但这些方法的计算复杂度较高，尤其是当匹配集很大并且外点率极高时。

深度学习方法 除了仅使用几何方法外，最近的工作还采用了深度学习技术来进行三维点云配准。一些方法旨在检测更多可重复的关键点^[17, 18]，并提取更多的描述性特征^[5, 6]。Choy 等人^[6]提出的 FCGF 在没有关键点检测的情况下，通过全卷积神经网络在单次通过中计算特征。Bai 等人^[17]提出的 D3Feat 使用全卷积网络来获得点云的局部信息，并使用联合学习框架来实现 3D 局部特征检测和描述。为了对点云提取描述能力更强的特征，Ao 等人^[5]提出 SpinNet 提取局部特征，这些局部特征是旋转不变的并且具有足够的信息量以实现准确的配准；为了有效区分匹配中的内点和外点，Bai 等人^[1]提出 PointDSC 明确探索空间一致性，以消除异常匹配并使点云配准成功；为了遵循无检测方法，并实现端

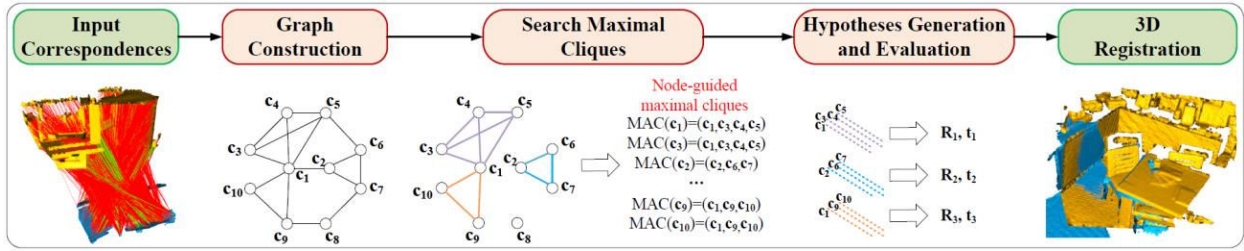


图 2 MAC 的流程图。a) 将初始匹配建模成兼容性图；b) 从兼容性图中搜索极大团作为兼容匹配集合；c) 根据兼容匹配集合生成并评估假设；d) 最佳假设作为配准结果。

到端方式的估计变换，Qin 等人^[8]提出 GeoTransformer 学习几何特征以进行强大的超点匹配，在低重叠情况下具有鲁棒性并且对刚性变换具有不变性。但是基于深度学习的方法需要大量的数据进行训练，并且通常在不同的数据集上缺乏泛化能力。

二、MAC方法介绍

2.1. 方法框架介绍

为了使两个点云 P^s 和 P^t 对齐，我们首先使用几何或深度学习描述子为它们提取局部特征。让 p^s 和 p^t 分别表示 P^s 和 P^t 中的点。通过匹配特征描述符，形成初始匹配集 $C_{initial} = \{c\}$ ，其中 $c = (p^s, p^t)$ 。MAC 从输入的 $C_{initial}$ 估计 P^s 和 P^t 之间的 6-DoF 姿态变换。方法的框架如图 2 所示。

2.2. 构建兼容性图

由于图空间 $G = (V, E)$ 相较于欧几里得空间可以更准确地描述匹配之间的亲和关系，因此我们将初始匹配建模为兼容性图，其中匹配由节点表示，几何上兼容的点由边连接。兼容性图的构建方式有以下两种：

1) 一阶图 (FOG)：FOG 是基于匹配对之间的刚性距离约束构架的，这种约束可以定量测量为：

$$S_{dist}(c_i, c_j) = ||p_i^s - p_j^s| - |p_i^t - p_j^t||. \quad (1)$$

c_i 和 c_j 兼容性分数的计算方式为：

$$S_{cmp}(c_i, c_j) = \exp\left(-\frac{S_{dist}^2(c_i, c_j)}{2d_{cmp}^2}\right), \quad (2)$$

其中 d_{cmp} 是一个距离参数。若 $S_{cmp}(c_i, c_j)$ 大于阈值， c_i 和 c_j 之间形成边 e_{ij} 并且 $S_{cmp}(c_i, c_j)$ 是 e_{ij} 的权值；否则， $S_{cmp}(c_i, c_j)$ 被设为 0。由于兼容性图是一个无向图，图的权重矩阵是一个对称矩阵。

2) 二阶图 (SOG)：之前的研究^[11]提出了二阶兼容性度量，它与全局集中的共同兼容匹配的数量有关。SOG 由 SOG 演变而来，其权重矩阵 W_{SOG} 的计算方式为：

$$W_{SOG} = W_{FOG} \odot (W_{FOG} \times W_{FOG}), \quad (3)$$

其中 \odot 表示矩阵按元素相乘。

如图 3 所示：SOG 考虑了匹配对全局集中的共同兼容性，而不仅仅是几何一致性，使其在高外点率的情况下更一致和鲁棒；SOG 比 FOG 更稀疏，因此有助于团搜索更加快速。基于以上两点，MAC 构建 SOG 来表示匹配的兼容性关系。

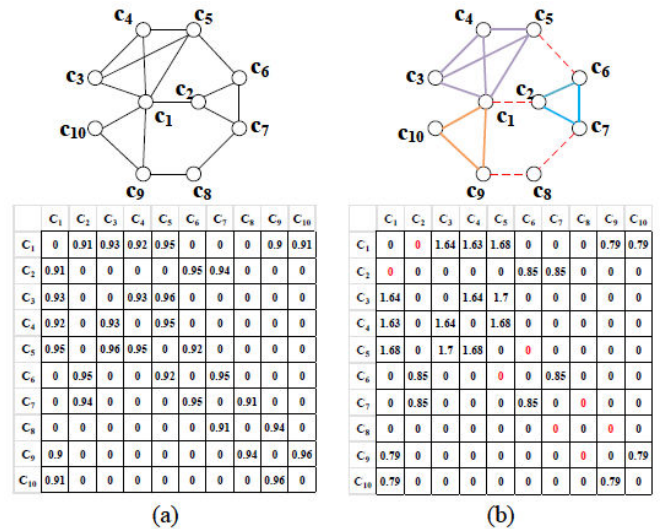


图 3 阐述一阶图 FOG 与二阶图 SOG 之间联系的示例。

(a) FOG 及其邻接矩阵； (b) SOG 及其邻接矩阵。

2.3. 搜索极大团

给定一个无向图 G ，团 $C = (V', E')$ ， $V' \subseteq V, E' \subseteq E$ 是 G 的子集，其中任意两个节点都由边连接。极大团是不能通过添加任何节点来扩展的团。特别地，具有最多节点数的极大团是图的最大团。理论上，内点会在图中形成团，因为内点通常在几何上彼此兼容。之前的研究^[23]专

注于在图中搜索最大团，但是最大团是非常严格的约束，只关注图中的全局共识信息。相反，我们放宽了约束，并利用极大团来挖掘更多的局部图信息。极大团的搜索使用 igraph 完成。

执行极大团搜索后，我们获得了极大团集合 $MAC_{initial}$ 。在实践中， $MAC_{initial}$ 通常包含数万个极大团，如果我们考虑所有极大团将会非常耗时。本节中，我们引入了一种节点引导的团筛选方法来减少 $|MAC_{initial}|$ 。首先，我们计算集合中每个团的权重。给定一个团 $C_i = (V_i, E_i)$ ，权重 $w(C_i)$ 计算如下：

$$w(C_i) = \sum_{e_j \in E_i} w(e_j), \quad (4)$$

其中 $w(e_j)$ 表示 W_{SOG} 中边 e_j 的权重。一个节点可能被周围的多个极大团包含，我们只保留局部权重最大的团与节点对应。然后，对剩余的团去重，获得极大团集合 $MAC_{selected}$ 。此步骤的动机是使用图节点周围局部几何结构的信息来找到最佳一致性的对应节点集。很明显， $|MAC_{selected}|$ 不会超过 $|V|$ 。我们可以将这些极大团直接送到下一阶段进行三维点云配准。

2.4. 假设生成与假设

从上一步筛选出的每个极大团都代表了一组一致的匹配。通过对每个一致性集应用 SVD 算法，我们可以获得一组 6-DoF 姿态假设。MAC 的最终目标是估计最大化如下目标函数的最优 6-DoF 刚性变换（由旋转姿态 $R^* \in SO(3)$ 和平移姿态 $t^* \in \mathbb{R}^3$ 组成）：

$$(R^*, t^*) = \arg \max_{R, t} \sum_{i=1}^N s(c_i), \quad (5)$$

其中 $c_i \in C_{initial}$ ， $N = |C_{initial}|$ 。我们在这里考虑了几个 RANSAC 假设评估指标^[15]，包括 MAE，MSE 和内点数。 $s(c_i)$ 表示一致性集中每个 c_i 的指标得分。拥有最高评估分数的假设被视为最佳假设，并作为最终变换以执行点云配准。

三、实验结果

3.1. 在模型数据集 U3M 上的结果

从图 4 中广泛比较的结果可以看出：MAC 性能最好，显著优于所有测试的 RANSAC 方式估计器，如 SAC-COT、OSAC、SAC-IA 和 RANSAC。基于 MAE 评估度

量的 MAC 在 U3M 上的配准性能最好。

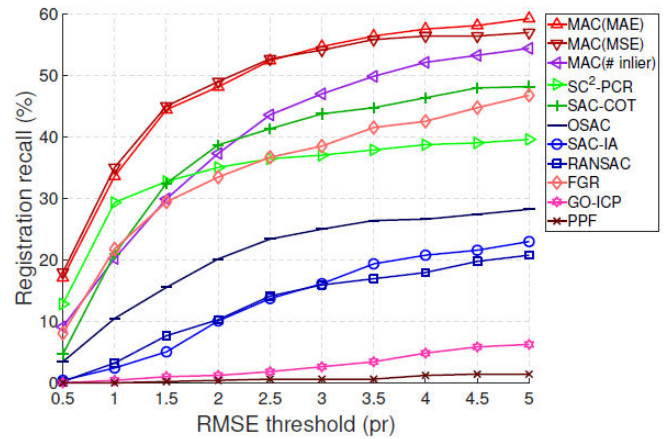


图 4 各方法在 U3M 数据集上的配准结果

3.2. 在室内场景数据集 3DMatch 和 3DLoMatch 上的结果

表 1 3DMatch 数据集上的配准结果

	FPFH			FCGF		
	RR(%)	RE(°)	TE(cm)	RR(%)	RE(°)	TE(cm)
<i>i) Traditional</i>						
SM [20]	55.88	2.94	8.15	86.57	2.29	7.07
FGR [45]	40.91	4.96	10.25	78.93	2.90	8.41
RANSAC-1M [13]	64.20	4.05	11.35	88.42	3.05	9.42
RANSAC-4M [13]	66.10	3.95	11.03	91.44	2.69	8.38
GC-RANSAC [5]	67.65	2.33	6.87	92.05	2.33	7.11
TEASER++ [36]	75.48	2.48	7.31	85.77	2.73	8.66
CG-SAC [30]	78.00	2.40	6.89	87.52	2.42	7.66
SC ² -PCR [8]	83.73	2.18	6.70	93.16	2.09	6.51
<i>ii) Deep learned</i>						
3DRegNet [27]	26.31	3.75	9.60	77.76	2.74	8.13
DGR [9]	32.84	2.45	7.53	88.85	2.28	7.02
DHVR [19]	67.10	2.78	7.84	91.93	2.25	7.08
PointDSC [3]	72.95	2.18	6.45	91.87	2.10	6.54
MAC	84.10	1.96	6.18	93.72	1.89	6.03

表 2 3DLoMatch 数据集上的配准结果

	FPFH			FCGF		
	RR(%)	RE(°)	TE(cm)	RR(%)	RE(°)	TE(cm)
<i>i) Traditional</i>						
RANSAC-1M [13]	0.67	10.27	15.06	9.77	7.01	14.87
RANSAC-4M [13]	0.45	10.39	20.03	10.44	6.91	15.14
TEASER++ [36]	35.15	4.38	10.96	46.76	4.12	12.89
SC ² -PCR [8]	<u>38.57</u>	<u>4.03</u>	<u>10.31</u>	<u>58.73</u>	<u>3.80</u>	<u>10.44</u>
<i>ii) Deep learned</i>						
DGR [9]	19.88	5.07	13.53	43.80	4.17	10.82
PointDSC [3]	20.38	4.04	10.25	56.20	3.87	10.48
MAC	40.88	3.66	9.45	59.85	3.50	9.75

配准方法对比 从表 1 和表 2 可以得出以下结论：

1) 无论使用哪种描述符，MAC 在 3DMatch 和 3DLoMatch 数据集上都优于所有比较方法，表明其强大的室内场景点云配准能力；2) 即使与深度学习方法进行比较，MAC 仍然在没有任何数据训练的情况下实现了更好的性能；3) 除了配准召回率 (RR) 指标外，MAC 实现了最佳的 RE 和 TE 指标。这表明 MAC 的配准非常准确，并且 MAC 能够对齐重叠度低的数据。

使用 MAC 增强深度学习性能 如表 3 所示, MAC 在 3DMatch 和 3DLoMatch 数据集上的所有测试方法中都显著提高了配准召回率。值得注意的是, SpinNet, Predator 和 CoFiNet 在 MAC 的增强下的表现超过了 GeoTransformer。MAC 与 GeoTransformer 配合使用, 在 3DMatch / 3DLoMatch 上实现了 95.7%/ 78.9% 的最新配准召回率。结果表明: 1) MAC 可以极大地提高现有的深度学习性能; 2) MAC 对样本数量不敏感。

表 3 3DMatch 与 3DLoMatch 数据集上深度学习方法与 MAC 结合后的性能提升

# Samples	3DMatch RR(%)					3DLoMatch RR(%)				
	5000	2500	1000	500	250	5000	2500	1000	500	250
FCGF [10]	85.1	84.7	83.3	81.6	71.4	40.1	41.7	38.2	35.4	26.8
SpinNet [11]	88.6	86.6	85.5	83.5	70.2	59.8	54.9	48.3	39.8	26.8
Predator [18]	89.0	89.9	90.6	88.5	86.6	59.8	61.2	62.4	60.8	58.1
CoFiNet [43]	89.3	88.9	88.4	87.4	87.0	67.5	66.2	64.2	63.1	61.0
GeoTransformer [29]	92.0	91.8	91.8	91.4	91.2	75.0	74.8	74.2	74.1	73.5
FCGF+MAC	91.3	92.2	91.6	90.4	85.6	57.2	56.0	52.6	42.4	32.1
SpinNet+MAC	6.2 \uparrow	7.5 \uparrow	8.3 \uparrow	8.8 \uparrow	14.2 \uparrow	17.1 \uparrow	14.3 \uparrow	14.4 \uparrow	7.0 \uparrow	5.3 \uparrow
Predator+MAC	95.3	95.1	93.3	91.4	81.2	72.8	69.9	59.2	54.8	32.1
CoFiNet+MAC	6.7 \uparrow	8.5 \uparrow	7.8 \uparrow	7.9 \uparrow	11.0 \uparrow	13.0 \uparrow	15.0 \uparrow	10.9 \uparrow	15.0 \uparrow	5.3 \uparrow
GeoTransformer+MAC	94.6	94.4	94.0	93.5	92.3	70.9	70.4	69.8	67.2	64.1
	5.6 \uparrow	4.5 \uparrow	3.4 \uparrow	5.0 \uparrow	5.7 \uparrow	11.1 \uparrow	9.2 \uparrow	7.4 \uparrow	6.4 \uparrow	6.0 \uparrow
	94.1	94.4	94.5	93.8	92.7	71.6	71.5	70.6	69.2	68.1
	4.8 \uparrow	5.5 \uparrow	6.1 \uparrow	6.4 \uparrow	5.7 \uparrow	4.1 \uparrow	5.3 \uparrow	6.4 \uparrow	6.1 \uparrow	7.1 \uparrow
	95.7	95.7	95.2	95.3	94.6	78.9	78.7	78.2	77.7	76.6
	3.7 \uparrow	3.9 \uparrow	3.4 \uparrow	3.9 \uparrow	3.4 \uparrow	3.9 \uparrow	3.9 \uparrow	4.0 \uparrow	3.6 \uparrow	3.1 \uparrow

3.3. 在室外场景数据集 KITTI 上的结果

表 4 KITTI 数据集上的配准结果

	FPFH			FCGF		
	RR(%)	RE(°)	TE(cm)	RR(%)	RE(°)	TE(cm)
<i>i) Traditional</i>						
FGR [45]	5.23	0.86	43.84	89.54	0.46	25.72
TEASER++ [36]	91.17	1.03	17.98	94.96	0.38	13.69
RANSAC [13]	74.41	1.55	30.20	80.36	0.73	26.79
CG-SAC [30]	74.23	0.73	14.02	83.24	0.56	22.96
SC ² -PCR [8]	<u>99.28</u>	<u>0.39</u>	8.68	97.84	0.33	20.58
<i>ii) Deep learned</i>						
DGR [9]	77.12	1.64	33.10	96.90	0.34	21.70
PointDSC [3]	<u>98.92</u>	0.38	8.35	97.84	0.33	20.32
MAC	99.46	0.40	<u>8.46</u>	97.84	<u>0.34</u>	<u>19.34</u>

从表 4 可以看出: 就配准召回性能而言, MAC 表现最好。MAC 也具有比现有技术的仅几何方法 SC²-PCR 更低的 TE。值得注意的是, 室外点云非常稀疏且分布不均匀。在物体、室内场景和室外场景数据集上的配准实验一致验证了 MAC 在不同应用场景下具有良好的泛化能力。

3.4. 在室外场景数据集 ETH 上的结果

表 5 不同方法在 ETH 数据集上生成匹配的内点率

	Gazebo		Wood		Avg.
	Autumn	Summer	Autumn	Summer	
FPFH [7]	0.42	0.24	0.21	0.26	0.29
FCGF [2]	2.34	1.25	1.35	1.68	1.62
Spinnet [11]	16.67	13.73	12.20	14.67	14.40

表 6 ETH 数据集各方法与 MAC 结合后的性能提升

	Gazebo		Wood		Avg.
	Autumn	Summer	Autumn	Summer	
FPFH [7]	16.85	10.03	10.43	10.40	11.92
FCGF [2]	54.35	28.03	52.17	51.20	42.78
Spinnet [11]	98.37	83.05	100.00	99.20	92.57
FPFH+MAC	46.74	27.68	33.04	43.20	36.12
FCGF+MAC	29.89 \uparrow	17.65 \uparrow	22.61 \uparrow	32.80 \uparrow	24.20 \uparrow
Spinnet+MAC	75.54	42.91	71.30	73.60	61.29
	21.19 \uparrow	14.88 \uparrow	19.13 \uparrow	22.40 \uparrow	18.51 \uparrow
	98.91	87.54	100.00	100.00	94.67
	0.54 \uparrow	4.49 \uparrow	-	0.80 \uparrow	2.10 \uparrow

与 3DMatch 相比, ETH^[19] 包含更复杂的几何形状。FPFH^[20]、FCGF 和 SpinNet 用于生成匹配, 然后 RANSAC-50K 和 MAC 将根据这些匹配执行配准。采样点或匹配的数量统一设置为 5000。当 RE $\leq 15^\circ$ 和 TE ≤ 30 cm 时, 配准被视为成功。表 5 和表 6 分别报告了生成匹配的质量和配准结果。

结果表明, 当描述符中的缺陷导致生成匹配集的内点率非常低时, MAC 仍然可以有效地从中找到准确的一致子集, 从而大大提高配准召回率。当与 FPFH 组合时, 使用 MAC 获得的配准召回率比 RANSAC 高 24.2%, 与 FC 组合时高 18.51%。MAC 与 SpinNet 结合在 ETH 上实现了 94.67% 的配准召回率。

3.5. MAC 生成假设的质量

表 7 RANSAC 和 MAC 生成正确假设数量的比较

# hypotheses	3DMatch				3DLoMatch			
	RANSAC		MAC		RANSAC		MAC	
	FCGF	FPFH	FCGF	FPFH	FCGF	FPFH	FCGF	FPFH
100	10.45	0.76	61.94	50.67	1.25	0.05	30.47	12.22
200	20.76	1.50	119.20	89.27	2.52	0.09	55.57	17.59
500	51.74	3.68	269.06	162.41	6.21	0.21	109.32	23.32
1000	103.65	7.39	456.18	217.32	12.43	0.41	156.11	26.02
2000	208.24	14.90	669.32	254.13	24.80	0.81	202.12	29.31

我们通过将 RANSAC 和 MAC 的假设与真值变换进行比较来评估生成的假设的质量。结果如表 7 所示。

与 RANSAC 在没有任何几何约束的情况下随机选择匹配并从中生成假设相比, MAC 有效地从兼容性图中的极大团中生成了更令人信服的假设, 充分利用了图中的一致性信息。

3.6. MAC 的性能上限

我们通过改变生成的正确假设数量的判断阈值来测试 MAC 的性能上限, 结果如表 8 所示。令人印象深刻的是, MAC-1 在 3DMatch / 3DLoMatch 上实现了 98.46%/ 91.24% 的配准召回率。这表明, 即使在重叠

度较低的数据集上，MAC 也能够为大多数点云对生成正确的假设。此外，我们可以推断出，使用更好的假设评估度量可以进一步提高 MAC 的性能。

表 8 性能上限探究。MAC-n 表示至少生成 n 个假设即认为可成功配准。

	3DMatch RR(%)	3DLoMatch RR(%)
MAC-1	98.46	91.24
MAC-5	97.10	83.32
MAC-10	96.43	77.93
MAC-20	94.70	70.47
MAC-50	91.13	56.37
MAC-origin	93.72	59.85

3.7. 时间和空间效率分析

表 9 平均耗时对比

# Corr.	250	500	1000	2500	5000
PointDSC	32.24±0.81	78.38±0.89	240.46±2.18	1401.97±12.24	5504.11±10.32
TEASER++	6.40±1.88	6.68±0.66	16.74±1.21	104.24±0.53	484.93±1.87
SC ² -PCR	19.34±0.63	63.23±0.55	215.98±1.24	1282.73±4.05	5210.17±8.30
MAC	7.32±0.55	23.32±0.38	56.45±1.41	282.67±7.83	3259.38±12.66

表 10 平均内存消耗对比

# Corr.	250	500	1000	2500	5000
PointDSC	3531.46	3538.26	3582.57	3634.22	3736.10
TEASER++	1631.92	1634.77	2029.22	2266.84	2484.83
SC ² -PCR	448.01	453.18	508.40	621.27	690.22
MAC	15.59	17.43	23.49	52.79	150.86

表 9 和表 10 分别显示了几种执行良好的方法的时间效率和内存消耗结果。关于时间效率实验，所有方法都进行了十轮测试，并报告了平均值和标准差结果。

所有方法都仅使用 CPU 进行测试。结果表明，当输入匹配数小于 2.5k 时，MAC 是非常轻量和高效率的。

3.8. 可视化结果

在图 5 和图 6 中我们展示了更多可视化结果。图 5 表明在低内点率数据上，MAC 能够进行高精度配准；图 6 表明 MAC 能够处理其他方法配准失败的低重叠率数据。

四、总结

我们提出了 MAC 来解决点云配准问题，通过使用极大团约束从匹配中生成精确的姿态假设。我们的方法在所有测试的数据集上都实现了最先进的性能，并且可以适应深度学习方法来进一步提高其性能。

责任编辑 金鑫 王金甲

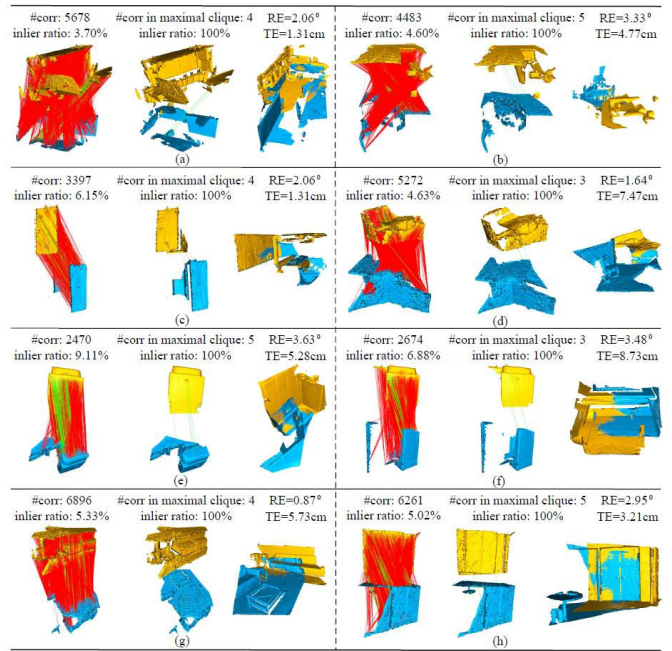


图 5 MAC 在 3DMatch 数据集上的配准过程可视化。红色和绿色的线分别代表外点和内点。

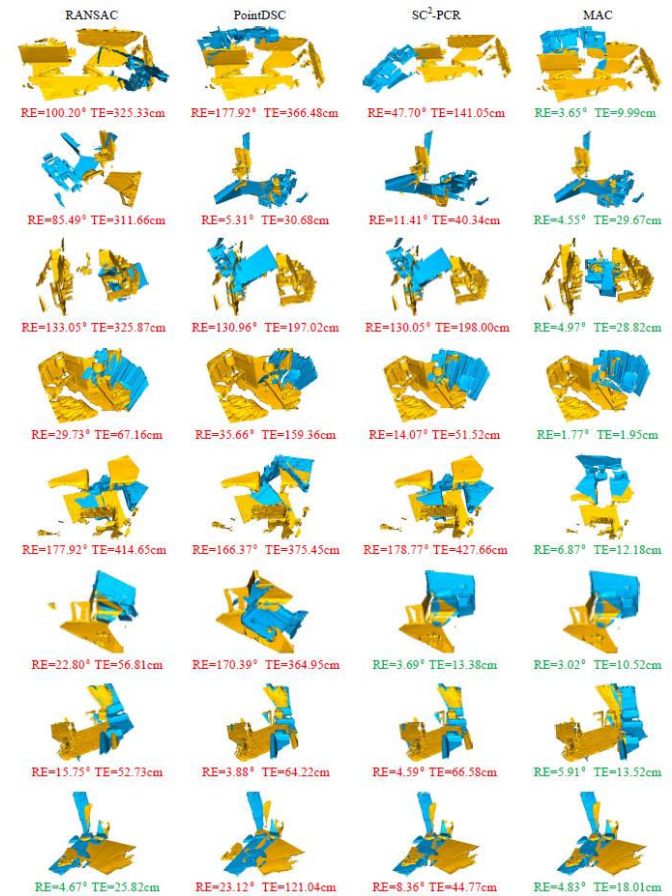


图 6 不同方法在 3DLoMatch 数据集上的配准结果对比。红色和绿色分别代表配准失败与配准成功。

参考文献

- [1] Xuyang Bai, Zixin Luo, Lei Zhou, Hongkai Chen, Lei Li, Zeyu Hu, Hongbo Fu, and Chiew-Lan Tai. Pointdsc: Robust point cloud registration using deep spatial consistency. In CVPR 2021.
- [2] Christopher Choy, Wei Dong, and Vladlen Koltun. Deep global registration. In CVPR 2020.
- [3] Junha Lee, Seungwook Kim, Minsu Cho, and Jaesik Park. Deep hough voting for robust global registration. In CVPR 2021.
- [4] G Dias Pais, Srikumar Ramalingam, VenuMadhav Govindu, Jacinto C Nascimento, Rama Chellappa, and Pedro Miraldo. 3Dregnet: A deep neural network for 3d point registration. In CVPR 2020.
- [5] Sheng Ao, Qingyong Hu, Bo Yang, Andrew Markham, and Yulan Guo. Spinnet: Learning a general surface descriptor for 3d point cloud registration. In CVPR 2021.
- [6] Christopher Choy, Jaesik Park, and Vladlen Koltun. Fully convolutional geometric features. In CVPR 2019.
- [7] Shengyu Huang, Zan Gojcic, Mikhail Usvyatsov, Andreas Wieser, and Konrad Schindler. Predator: Registration of 3D point clouds with low overlap. In CVPR 2021.
- [8] Zheng Qin, Hao Yu, Changjian Wang, Yulan Guo, Yuxing Peng, and Kai Xu. Geometric transformer for fast and robust point cloud registration. In CVPR 2022.
- [9] Hao Yu, Fu Li, Mahdi Saleh, Benjamin Busam, and Slobodan Ilic. Cofinet: Reliable coarse-to-fine correspondences for robust point cloud registration. In CVPR 2021.
- [10] XAlvaro Parra Bustos and Tat-Jun Chin. Guaranteed outlier removal for point cloud registration with correspondences. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(12):2868–2882, 2017.
- [11] Zhi Chen, Kun Sun, Fan Yang, and Wenbing Tao. Sc²-pcr: A second order spatial compatibility for efficient and robust point cloud registration. In CVPR 2022.
- [12] Marius Leordeanu and Martial Hebert. A spectral technique for correspondence problems using pairwise constraints. 2005.
- [13] Heng Yang, Jingnan Shi, and Luca Carlone. Teaser: Fast and certifiable point cloud registration. IEEE Transactions on Robotics, 37(2):314–333, 2020.
- [14] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Fast global registration. In ECCV 2016.
- [15] Jiaqi Yang, Zhiqiang Huang, Siwen Quan, Qian Zhang, Yanning Zhang, and Zhiguo Cao. Toward efficient and robust metrics for ransac hypotheses and 3D rigid registration. IEEE Transactions on Circuits and Systems for Video Technology. 893–906, 2021.
- [16] Alvaro Parra Bustos and Tat-Jun Chin. Guaranteed outlier removal for point cloud registration with correspondences. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2868–2882, 2017.
- [17] Xuyang Bai, Zixin Luo, Lei Zhou, Hongbo Fu, Long Quan, and Chiew-Lan Tai. D3feat: Joint learning of dense detection and description of 3d local features. In CVPR 2020.
- [18] Shengyu Huang, Zan Gojcic, Mikhail Usvyatsov, Andreas Wieser, and Konrad Schindler. Predator: Registration of 3d point clouds with low overlap. In CVPR 2021.
- [19] François Pomerleau, Ming Liu, Francis Colas, and Roland Siegwart. Challenging data sets for point cloud registration algorithms. The International Journal of Robotics Research, 31(14):1705–1711, 2012.
- [20] Radu Bogdan Rusu, Nico Blodow, and Michael Beetz. Fast point feature histograms (fpfh) for 3D registration. In IEEE International Conference on Robotics and Automation, 3212–3217, 2009.



张曦予

西北工业大学计算机学院 2022 级硕士研究生，导师为杨佳琪副教授，主要研究方向为三维特征匹配与点云配准。

Email: 2426988253@mail.nwpu.edu.cn



杨佳琪

西北工业大学计算机学院的副教授，研究方向包括局部几何描述、三维配准、三维特征匹配和三维物体识别。他分别于 2014 年和 2019 年在华中科技大学获得学士学位和博士学位。在中国国家留学基金委员会的资助下，他于 2017 年至 2018 年期间访问了宾夕法尼亚大学的 GRASP 实验室。

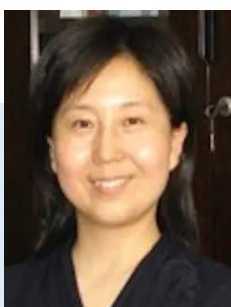
Email: jqyang@nwpu.edu.cn



张世坤

西北工业大学计算机学院 2021 级硕士研究生，导师为杨佳琪副教授，主要研究方向为多视图点云配准。

Email: zhangshikun@mail.nwpu.edu.cn



张艳宁

西北工业大学副校长，教授，国家级人才，国防 973 项目首席。长期致力于图像处理、模式识别、计算机视觉与智能信息处理等的研究，并与航天、航空等方面的国家重大需求相结合。获国家教学成果二等奖 1 项、省部级科技进步奖 3 项，曾获全国三八红旗手称号和总装 863 科技攻关先进个人。先后承担国防 973 项目、国家自然科学基金重点项目、国家/国防 863、总装预研等国家级项目 40 余项。在 IEEE TPAMI、IEEE TIP、PR、IEEE TSMC-B、Info. Fusion、CVPR、ICCV 等国内外本领域权威期刊和重要国际会议上发表论文百余篇。研究成果被多个国家级重大工程攻关项目采用，成功应用于航天、航空、能源、水利等行业的 20 余家单位。获国家/国防授权发明专利 50 余项，陕西省科技进步一等奖一项，国防技术发明一等奖一项，出版专著 3 部。

Email: ynzhang@nwpu.edu.cn

专题综述

基于超像素交互的图像超分辨率

中山大学 张爱平 任文琦 操晓春

本文是中山大学团队解读其在ICCV2023录用的工作SPIN^[1]。论文研究的问题是**基于超像素交互的图像超分辨率 (Super-Resolution, SR)**。现有方法中, 基于Transformer的方法在单幅图像SR任务上取得了高质量结果, 将低分辨率输入图像划分为矩形窗口并行处理, 然后融合以生成高分辨率图像。然而, 这种划分过于粗糙且缺乏可解释性, 影响相似结构间的信息交互从而降低超分性能, 并容易出现伪影。为了应对这些挑战, 我们提出了一种新的超像素信息交互网络 (SPIN), 利用超像素对局部相似像素进行聚类, 形成可解释的连通区域, 并利用超像素内的注意力实现局部信息交互。它是可解释的, 因为只有相似的区域相互补充, 而不同的区域被排除在外。此外, 我们设计了一个超像素交叉注意模块, 利用超像素作为代理来促进全局信息的传播。实验表明, 所提出的SPIN模型在准确性和轻量级方面优于最先进的SR方法。

一、研究背景

单幅图像超分辨率 (Single Image Super-Resolution, SISR) 是底层计算机视觉中的一项关键任务, 旨在提高低分辨率 (LR) 图像的分辨率和视觉质量, 生成高分辨率 (HR) 图像。

自从 SRCNN^[2]的开创性工作以来, 已经有许多神经网络来应对从低分辨率输入重建高质量图像的挑战。近年来, 注意力机制^[3], 尤其是 Transformer, 已被证明对高层视觉任务和底层视觉领域都有显著效果, 包括超分辨率 (SR) 任务。注意力机制有很强的上下文特征提取能力, 允许网络选择性地关注输入的相关区域,

这可以提高 SR 输出的质量。然而, 注意力机制具有较高的计算复杂度和内存消耗, 因此通常需要将大图像分割成小窗口, 分别在窗口内做注意力操作, 如 SwinIR^[4]和 ESRT^[5]和 ELAN^[6]。然而, 基于固定矩形形状划分窗口会导致本身连续的相似结构的分割在不同的窗口, 影响信息直接交互。同时, 每个窗口中应用的局部注意力机制涉及语义不相关区域, 从而导致不理想的模型预测。为了解决这些问题, 我们提出了一种新的方法, 将局部和全局注意力机制与精细的超像素划分相结合。

二、基于超像素交互的超分辨率模型

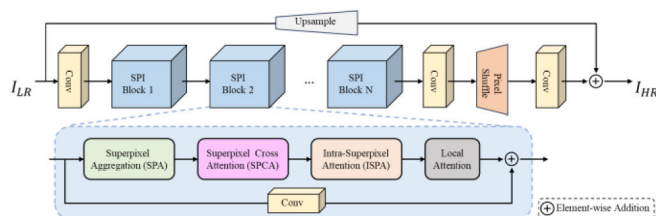


图 1 基于超像素交互的图像超分辨率模型架构

我们提出的模型的架构如图 1 所示。首先, 我们使用编码器, 即 3×3 卷积, 将低分辨率图像 I_{LR} 嵌入到高维特征空间中

$$x_{emb} = f_{encoder}(I_{LR})$$

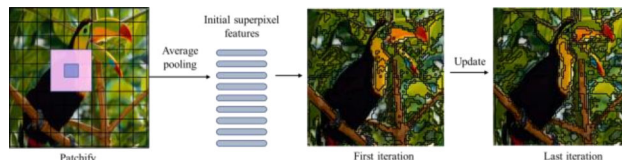


图 2 超像素迭代更新示意图

然后, 通过多个超像素交互 (SPI) 模块进行深层特征提取。该模块由四个组件组成: 超像素聚合 (SPA)、超像素交叉注意力 (SPCA)、超像素内注意力 (ISPA)

和局部注意力。如图 2 所示, SPI 模块利用 SSN^[7]的基于 soft k-means 的可微超像素算法, 迭代更新超像素特征和像素到超像素的关联映射图 A 。SPA 模块负责聚合来自输入图像中的超像素的信息, SPCA 模块通过超像素的替代来捕捉像素之间的交互, 而 ISPA 模块捕捉每个超像素内像素之间的相互作用, 局部注意力模块用来加强局部内部的信息交互。对于第 i 个 SPI 模块, 我们有

$$x_i = x_{i-1} + f_{local}(f_{ISPA}(f_{SPCA}(x_{i-1}, s_i)))$$

其中, s_i 为超像素特征。最后, 我们利用 3×3 卷积和 pixel-shuffle [34] 来获得全局残差信息, 该信息被添加到 I_{LR} 的上采样图像中, 用于获取高分辨率图像 I_{SR} 。

2.1. 超像素交叉注意力

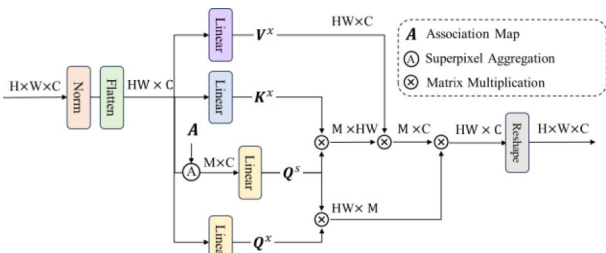


图 3 超像素交叉注意力

如图 3 所示, 超像素交叉注意力(SPCA)旨在利用超像素进行长距离的信息传递。由于超像素有着单一的语义信息, 且与像素特征的相似性高, 很适合作为代理(proxy)特征来进行信息传递。具体来说, 我们使用注意力机制^[3]首先将像素信息传播到超像素。我们使用线性映射来计算 query, key 和 value:

$$Q^s = sW_q^s, \quad K^x = xW_k^x, \quad V^x = xW_v^x$$

其中, $W_q^s \in \mathbf{R}^{C \times D}$, $W_k^x \in \mathbf{R}^{C \times D}$, $W_v^x \in \mathbf{R}^{C \times C}$ 分别为 query, key 和 value 的权重矩阵, x 代表像素, s 代表超像素。首先计算 query 和 key 之间的相似性, 并将其用作权重来聚合 key 来获得输出

$$s_u = \text{softmax}(Q^s(K^x)^T / \sqrt{D})V^x$$

s_u 相当于新的超像素特征。这个过程不考虑邻居的限制, 确保了长距离信息的传播。

由于信息从像素传播到了超像素, 有必要将聚合的信息分发回像素, 以实现像素之间的信息传播。在这里,

我们进一步使用注意力机制。类似地, 我们利用另一个权重矩阵 W_q^s 来从像素特征获得 query。为了减少参数的数量, 我们直接使用 Q^s 作为 key, 并使用更新的超像素特征 s_u 作为 value, 并利用交叉注意力将更新的超像素特征映射回像素级别。与 Transformer^[3]类似, 在注意力机制后使用 FFN, 我们的 FFN 包含一个层归一化^[8]层, 之后我们使用特征门控^[9]来调制输入特征, 并使用通道注意力^[10]来提取全局信息。之后, 使用两个完全连接层和 GELU^[11]激活函数。

2.2. 超像素内部注意力

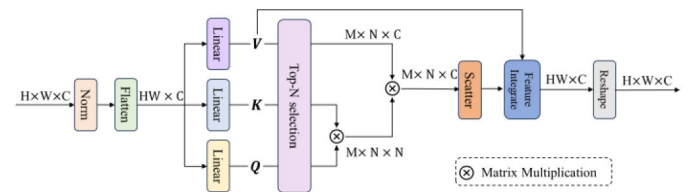


图 4 超像素内部注意力

如图 4 所示, 超像素内部注意力(ISPA)旨在利用超像素内部特征的相似性进行信息互补。提高超分辨率图像质量的直观方法是利用同一超像素内相似像素的互补性。为了实现这一目的, 我们需要获得每个超像素的相应像素。然而, 不同的超像素可能包含不同数量的像素, 这使得难以进行并行处理, 也会导致额外的内存消耗, 因为总有一些超像素包含大量像素。为了解决这个问题, 如图 5 所示, 我们求助于关联图 A , 并选择与每个超像素最相似的前 N 个像素。假设一个超像素的隶属像素为

$$f = \{x(i)\}_N \in \mathbf{R}^{N \times C}$$

其中 N 表示所选像素的数量。我们遵循标准的自注意机制^[3], 进行超像素内注意, 其中包括用于 query、key 和 value 投影的权重矩阵 W_q^f , W_k^f , W_v^f 。在超像素内交互之后, 我们利用前 N 个选择过程中生成的索引, 将细化的像素特征分散回图像中它们各自的位置。前 N 个选择可能导致一些“被忽略”的像素, 即这些像素不包括在任何超像素中。对于那些“被忽略”的像素, 我们利用值投影 W_v^f 来投影它们以获得更新的特征, 然后将其与通过超像素内交互更新的像素进行集成。与 SPCA 模块类似, 我们在 ISPA 模块之后采用了相同的 FFN。

表 1 超分辨率实验结果

Methods	Scale	Params	Set5	Set14	BSDS100	Urban100	Manga109
			PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
CARN	×2	1592K	37.76/0.9590	33.52/0.9166	32.09/0.8978	31.92/0.9256	38.36/0.9765
IMDN		694K	38.00/0.9605	33.63/0.9177	32.19/0.8996	32.17/0.9283	38.88/0.9774
LatticeNet+		756K	38.15/0.9610	33.78/0.9193	32.25/0.9004	32.29/0.9291	-
SwinIR-light		878K	38.14/0.9611	33.86/0.9206	32.31/0.9012	32.76/0.9340	39.12/0.9783
ESRT		677K	38.03/0.9600	33.75/0.9184	32.25/0.9001	32.58/0.9318	39.12/0.9774
ELAN-light		582K	38.17/0.9611	33.94/0.9207	32.30/0.9012	32.76/0.9340	39.11/0.9782
SPIN (Ours)		497K	38.20/0.9615	33.90/0.9215	32.31/0.9015	32.79/0.9340	39.18/0.9784
CARN	×3	1592K	34.29/0.9255	30.29/0.8407	29.06/0.8034	28.06/0.8493	33.43/0.9427
IMDN		703K	34.36/0.9270	30.32/0.8417	29.09/0.8046	28.17/0.8519	33.61/0.9445
LatticeNet+		765K	34.53/0.9281	30.39/0.8424	29.15/0.8059	28.33/0.8538	-
SwinIR-light		886K	34.62/0.9289	30.54/0.8463	29.20/0.8082	28.66/0.8624	33.98/0.9478
ESRT		770K	34.42/0.9268	30.43/0.8433	29.15/0.8063	28.46/0.8574	33.95/0.9455
ELAN-light		590K	34.64/0.9288	30.55/0.8463	29.21/0.8081	28.69/0.8624	34.00/0.9478
SPIN (Ours)		569K	34.65/0.9293	30.57/0.8464	29.23/0.8089	28.71/0.8627	34.24/0.9489
CARN	×4	1592K	32.13/0.8937	28.60/0.7806	27.58/0.7349	26.07/0.7837	30.42/0.9070
IMDN		715K	32.21/0.8948	28.58/0.7811	27.56/0.7353	26.04/0.7838	30.45/0.9075
LatticeNet+		777K	32.30/0.8962	28.68/0.7830	27.62/0.7367	26.25/0.7873	-
SwinIR-light		897K	32.44/0.8976	28.77/0.7858	27.69/0.7405	26.47/0.7980	30.92/0.9151
ESRT		751K	32.19/0.8947	28.69/0.7833	27.69/0.7379	26.39/0.7962	30.75/0.9100
ELAN-light		601K	32.43/0.8975	28.78/0.7858	27.69/0.7406	26.54/0.7982	30.92/0.9150
SPIN (Ours)		555K	32.48/0.8983	28.80/0.7862	27.70/0.7415	26.55/0.7998	30.98/0.9156

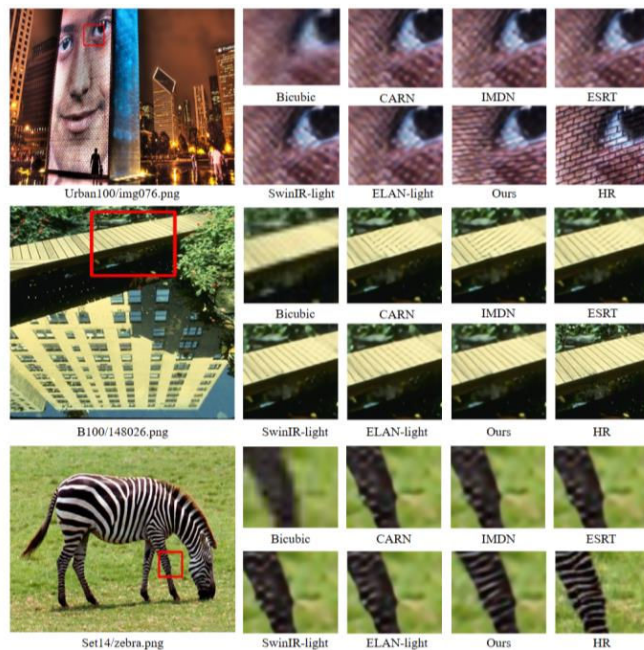


图 5 视觉定性结果

三、实验结果

3.1. 定量结果

将我们的模型与最先进的轻量级 SR 模型进行了比较，包括基于 CNN 的 CARN^[12]、IMDN^[13]、LatticeNet^[14]等模型，以及基于 Transformer 的 ESRT^[5]、SwinIR^[4]和 ELAN^[6]模型。如表 1 所示，通过利用图像块之间的长距离相似性，基于 Transformer 的模型^[4,5,6]在 PSNR 和 SSIM 方面始终优于基于 CNN 的方法^[12,13,15,16,17,18,19]。然而，他们总是将图像划分为规则的窗口，这可能会破坏输入图像中的对象、边界等。相反，我们的方法利用超像素为 Transformer 进行可解释和连续的区域划分。我们在所有五个基准数据集和所有三个量表上获得了最好或第二好的 PSNR/SSIM 分数。此外，参数的数量小于现有的基于 Transformer 的方法。

3.2. 定性结果

图 5 显示了 Urban100、BSDS100 和 Set14 数据集上比例×4 的可视化比较。结果表明，只要 LR 图像中存在相应的非局部信息，所提出的 SPIN 就可以有效地恢复被严重破坏的纹理。而缺乏非局部注意力的模型无法准确重建严重低质的结构，例如，比较图像“B100/148026”的重建结果时，很明显我们的模型产生的结果非常接近 HR，而其他没有非局部关注的 SISR

模型，如 CARN^[12]和 IMDN^[13]，不适合恢复这种严重受损的区域此外。与其他基于注意力的深度 SISR 方法（如 ESRT^[5]、SwinIR-light^[4]和 ELAN-light^[6]）相比，我们的 SPIN 模型仍然保持着卓越的重建质量。此外，对于图像“Urban100/img020”，即使没有太多的纹理信息，我们的方法也可以准确地恢复低质图像。

四、总结

我们提出了一种基于超像素交互网络（SPIN）的图像超分辨率新方法，该方法利用超像素将局部相似像素分割为可解释的连通区域。我们的方法采用超像素内注意力来促进不规则局部超像素区域内的局部信息交互，而超像素交叉注意力模块通过超像素代理来促进长距离信息交互。实验表明，SPIN在精度和轻量级方面优于最先进的超分辨率方法。此外，所提出的方法为利用可解释的区域划分处理整个图像的挑战提供了一个新的思路。

责任编辑 崔海楠

参考文献

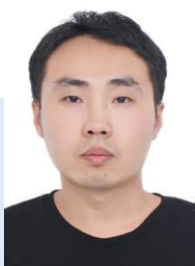
- [1] Aiping Zhang, Wenqi Ren, Yi Liu, Xiaochun Cao. Lightweight Image Super-Resolution with Superpixel Token Interaction. In ICCV 2023.
- [2] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In ECCV 2014.
- [3] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In Advances in Neural Information Processing Systems (NeurIPS), 2017.
- [4] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In ICCV 2021.
- [5] Zhisheng Lu, Juncheng Li, Hong Liu, Chaoyan Huang, Lili Zhang, and Tiejiong Zeng. Transformer for single image super-resolution. In CVPR 2022.
- [6] Xindong Zhang, Hui Zeng, Shi Guo, and Lei Zhang. Efficient long-range attention network for image super-resolution. In ECCV 2022.
- [7] Varun Jampani, Deqing Sun, Ming-Yu Liu, Ming-Hsuan Yang, and Jan Kautz. Superpixel sampling networks. In ECCV 2018.
- [8] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. arXiv preprint arXiv:1607.06450, 2016.
- [9] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. In AAAI 2020.
- [10] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In CVPR 2018.
- [11] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units. arXiv preprint arXiv:1606.08415, 2016.
- [12] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network. In ECCV 2018.
- [13] Zheng Hui, Xinbo Gao, Yunchu Yang, and Xiumei Wang. Lightweight image super-resolution with information multi-distillation network. In Proceedings of the ACM International Conference on Multimedia (ACM MM), 2019.
- [14] Xiaotong Luo, Yuan Xie, Yulun Zhang, Yanyun Qu, Cuihua Li, and Yun Fu. Latticenet: Towards lightweight image super-resolution with lattice block. In ECCV 2020.
- [15] Chaofeng Wang, Zheng Li, and Jun Shi. Lightweight image super-resolution with adaptive weighted learning network. arXiv preprint arXiv:1904.02358, 2019.
- [16] Rushi Lan, Long Sun, Zhenbing Liu, Huimin Lu, Cheng Pang, and Xiaonan Luo. Madnet: a fast and lightweight network for single-image super resolution. IEEE Transactions on Cybernetics, 51(3):1443–1453, 2020.
- [17] Wenbo Li, Kun Zhou, Lu Qi, Nianjuan Jiang, Jiangbo Lu, and Jiaya Jia. Lapar: Linearly-assembled pixel-adaptive regression network for single image super-resolution and beyond. In Advances in Neural Information Processing Systems (NeurIPS), 33:20343–20355, 2020.
- [18] Jie Liu, Jie Tang, and Gangshan Wu. Residual feature distillation network for lightweight image super-resolution. In ECCV 2020.
- [19] Xinyan Zhang, Peng Gao, Sunxiangyu Liu, Kongya Zhao, Guitao Li, Liuguo Yin, and Chang Wen Chen. Accurate and efficient image super-resolution via global-local adjusting dense network. IEEE Transactions on Multimedia, 23:1924–1937, 2021.



张爱平

中山大学网络空间安全学院 2022 级博士研究生，主要研究方向为计算机视觉。

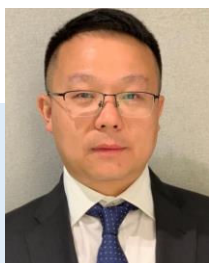
Email: zhangaip7@mail2.sysu.edu.cn



任文琦

副教授，硕士生导师，中山大学“百人计划”引进人才。国家自然科学基金优秀青年基金获得者。天津大学与美国加州大学 Merced 分校联合培养博士，从事计算机视觉与多媒体内容安全领域的研究。近年在 CCF-A 类期刊和会议长文发表学术论文 60 余篇，Google 学术引用 9000 余次，入选爱思唯尔中国高被引学者、百度全球高潜力 AI 华人青年榜。担任 AI 和 CV 领域多个顶级学术会议的领域主席和程序委员会委员。获中国计算机学会优博奖、吴文俊人工智能优秀青年奖。

Email: renwq3@mail.sysu.edu.cn



操晓春

中山大学信息学部副主任、网络空间安全学院院长，国家杰出青年/优秀青年基金获得者。主要从事人工智能基础研究和网络空间内容安全应用研究；发表 ACM/IEEE 汇刊 100 余篇，CCF-A 类期刊及会议长文文章 160 余篇；Google 引用 18000 余次，H-index 64；获得省部级一等奖和二等奖各 1 项。现兼任 TPAMI 的 Associate Editor、TIP 的 Senior Area Editor、电子学报的编委，曾兼任 TMM 和 TCSVT 的 Associate Editor，10 余次兼任 NeurIPS/ICCV/CVPR/IJCAI/ACMMM 的 Area/Track Chairs。指导博士生获得中国电子学会优博、CCF 优博、中科院优博论文；指导的研究生有 3 人入选国家级人才计划。

Email: caoxiaochun@mail.sysu.edu.cn