arXiv:2601.00658v1 [cs.CV] 2 Jan 2026

利用双拓扑网络从星载TomoSAR点云重建建筑物高度

Zhaiyu Chen, Yuanyuan Wang, *IEEE*会员，YileiShi，*IEEE*会员，以及 Xiao Xiang Zhu，*IEEE*会士

摘要——**可靠的建筑高度估计对于各种城市应用至关重要。星载 SAR 层析成像（TomoSAR）能够提供不受天气影响的侧视观测，捕捉立面级结构，为传统光学方法提供了一种很有前景的替代方案。然而，TomoSAR点云通常存在噪声、各向异性的点分布，以及非相干表面上的数据空洞，这些问题都会阻碍精确的高度重建。为应对这些挑战，我们提出了一种基于学习的框架，将原始TomoSAR 点转换为高分辨率建筑物高度图。我们的双拓扑网络在点分支和网格分支之间交替处理：点分支对不规则散射体特征进行建模，网格分支则强化空间一致性。通过联合处理这两种表示，网络能够对输入点进行去噪，并对缺失区域进行修复，从而生成连续的高度估计值。据我们所知，这是首个直接基于TomoSAR点云开展大规模城市高度制图的概念验证。我们在慕尼黑和柏林数据上开展的大量实验验证了该方法的有效性。此外，我们还证明了该框架可以扩展以融合光学卫星影像，从而进一步提升重建质量。源代码可访问 https://github.com/zhu-xlab/tomosar2height。**

索引词—**高度估计、三维重建、SAR 层析、点云、深度学习。**

I. 引言

建成环境的大规模三维建模对于城市规划、灾害管理和环境监测等多种应用至关重要。这类建模的一个关键环节是可靠地估计建筑物高度。传统上，人们采用机载激光雷达扫描和摄影测量来获取高质量的高度数据。然而，这些技术的可扩展性有限。激光雷达测量成本高昂，而摄影测量方法则需要大量无云的高分辨率光学图像。尽管计算机视觉近年来取得了进展

已实现从单幅图像进行高度估计 [1]–[3],

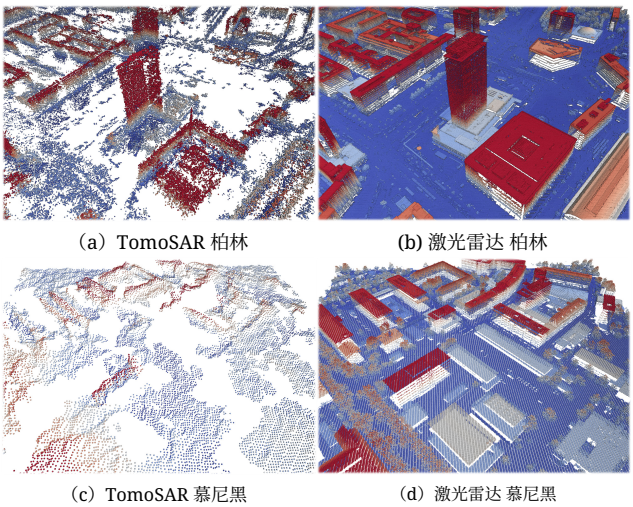
本研究由 TUM Georg Nemetschek Institute 通过 AI4TWINNING 项目、欧盟委员会通过“MultiMiner：面向矿产勘探和矿区监测的多源多尺度地球观测与新型机器学习方法”项目（该项目隶属 Horizon 2020 研究与创新计划，资助协议编号 101091374），以及慕尼黑机器学习中心共同资助。（通讯作者：*Xiao Xiang Zhu*。）

Z. Chen、Y. Wang 和 X. X. Zhu 任职于慕尼黑工业大学（TUM）地球观测数据科学讲席，地址为德国慕尼黑 80333（电子邮箱：zhaiyu.chen@tum.de；y.wang@tum.de；xiaoxiang.zhu@tum.de）。Z. Chen 和 X. X. Zhu 也任职于慕尼黑机器学习中心（MCML）。Y. Shi 任职于慕尼黑工业大学（TUM）工程与设计学院，地址为德国慕尼黑 80333（电子邮箱：yilei.shi@tum.de）。

本文已提交 IEEE，供其考虑是否发表。版权可能在未另行通知的情况下转让，届时本版本可能不再可访问。

这些单目方法仍受到晴空条件依赖性以及为消除深度歧义所需的强归纳偏置的限制。

星载合成孔径雷达（SAR）凭借其全天候成像能力以及捕获三维结构的能力，为大规模三维重建提供了互补数据源。具体而言，多基线SAR层析成像（TomoSAR）通过重建完全三维的反射率剖面，扩展了传统干涉测量技术，从而分离单个地面分辨率单元内的重叠散射体 [4], [5]。利用现代卫星获取的米级分辨率SAR影像（例如TerraSAR-X和TanDEM-X），TomoSAR能够生成城市区域一致的大规模点云 [6], [7]。这些SAR衍生点云提供了独特的几何信息，尤其能够通过侧视成像几何捕获建筑立面，而这些立面通常会被天底视角传感器遗漏。此外，TomoSAR点云还可提供较高的地理定位精度，尤其是在采用高级校准的情况下 [8]。



（a）TomoSAR 柏林 (b) 激光雷达 柏林 （c）TomoSAR 慕尼黑 （d）激光雷达 慕尼黑

图1. 柏林（a、b）和慕尼黑（c、d）相同区域上星载TomoSAR点云与机载激光雷达数据的比较。这两种点云分别由不同大小和空间分辨率的SAR影像堆栈重建而成，因此质量有所不同。TomoSAR点云具有广泛的覆盖范围，但噪声更高，点分布也更加不均匀，因此需要专门的处理技术。点按照高度进行彩色编码。

尽管具有这些优势，TomoSAR点云仍给建筑高度重建带来了重大挑战。成像过程本身以及侧视几何结构往往会导致数据稀疏且噪声较多，点密度不均并存在空隙，尤其是在低相干表面（例如

建筑屋顶）上。使用较低分辨率的合成孔径雷达影像或有限次数的数据采集时，这些挑战会更加突出。如图1所示，TomoSAR点云覆盖范围广，且能提供丰富的立面细节，但其点密度通常低于机载激光雷达，噪声则更大。这些问题会削弱空间插值或几何拟合等传统高度制图技术的效果。因此，要准确地从TomoSAR点云重建建筑物高度，需要采用专门的方法，对稀疏的各向异性采样点进行去噪，并修复数据空缺。

为应对这些挑战，我们提出了一种神经网络，可直接根据星载TomoSAR点云重建建筑物高度图。由于 TomoSAR 点的水平地理定位通常比其垂直高程更可靠，我们采用双拓扑设计，将基于点的分支与 x–y 网格分支结合起来，在保留不规则点结构细节的同时，利用网格对含噪高度进行正则化并确保空间一致性。该设计无需在推理时使用外部数字地形模型（DTM），即可实现高分辨率高度制图。此外，双重表示增强了去噪和图像修复能力，从而缓解噪声、各向异性点分布及数据空洞带来的问题。该处理流程具有良好的扩展性；我们还证明，融入光学卫星影像能够提供互补信息，进一步提升重建质量。

在慕尼黑和柏林城市区域开展的实验验证表明，我们的方法在不同数据采集条件下均有效。这些结果凸显了该方法作为概念验证方案，在实现大规模建筑高度制图方面的潜力。

我们的主要贡献总结如下：

- 我们提出了一种基于学习的框架，用于从星载TomoSAR点云大规模重建建筑高度图。
- 我们提出了一种双拓扑神经网络，在用于建模不规则散射体特征的点拓扑与用于保证空间一致性的网格拓扑之间交替，从而实现有效的去噪和图像修复。
- 我们通过整合光学影像展示了该框架的可扩展性，这进一步提升了高度重建效果，并增强了其在大规模城市制图中的应用潜力。

II. 相关工作

建筑建模通常依赖高精度三维数据，例如通过机载激光雷达和摄影测量获取的点云 [9], [10]。然而，大规模获取此类数据成本高昂，而且对许多地区而言往往难以实现。对于许多城市应用来说，仅有建筑高度数据便已足够。过去几年中，随着深度学习技术的发展以及高分辨率重复轨道星载合成孔径雷达影像的出现，基于单幅图像和SAR干涉测量的高度估计等替代方法已逐渐兴起。

A. 单幅影像高度估计

为使高度信息更易获取，研究人员开发了利用单目光学影像的方法。例如，Mou和Zhu提出了一种模型，利用全卷积网络从单幅卫星图像推断数字表面模型 [11]。后续研究探索了生成对抗网络 [12], [13], 、多任务学习 [14], [15],和混合回归 [1] ，以改进单目高度估计。这些单目光学图像方法前景广阔，但需要无遮挡且最好是高分辨率的图像。尽管这些方法在若干应用中可行 [2], [16]–[18], ，但由于图像质量不稳定且云覆盖频繁，仅凭光学影像及时生成全球高分辨率高度图仍颇具挑战。

得益于昼夜成像能力和通常不受天气影响的可用性，SAR 为建筑高度估计提供了一种颇具吸引力的替代方案。因此，许多研究致力于直接从 SAR 图像中估计城市建筑物高度。例如，Recla 和 Schmitt 提出了一种深度网络，用于学习从单幅超高分辨率 SAR 图像中预测高度图 [19], [20],；Sun 等则在已知建筑物轮廓的情况下，采用边界框回归来获取建筑物高度 [21]。结合 SAR 与光学数据的融合方法也得到了探索 [3]。尽管这些方法能够实现快速覆盖，但它们本质上缺乏明确的三维几何信息，可能导致歧义。

尽管单目光学和 SAR 单幅图像方法降低了城市高度制图的门槛，但它们对二维观测的依赖限制了其在复杂城市场景中的稳健性。相比之下，干涉合成孔径雷达（InSAR）技术利用多个观测角度来分离重叠散射体，并直接恢复三维结构。尤其是，多航过层析合成孔径雷达（TomoSAR）已成为实现精细城市重建的一种有前景的解决方案。

B. 用于城市重建的 *TomoSAR*

干涉合成孔径雷达可以生成大范围数字高程模型，但在密集城市环境中会受到叠掩影响，地面与建筑信号相互重叠，导致无法分离单个分辨单元内的多个散射体 [22]。多航过层析合成孔径雷达利用从略有差异的视角获取的一系列SAR图像，重建散射体的三维分布，从而克服这一局限。TomoSAR被广泛认为是一种强大的城市区域重建方法，能够分辨每个分辨单元中的多个反射目标，并生成精细的三维点云 [4], [23], [24]。早期TomoSAR研究引入了基于模型的反演技术，以从星载数据中恢复反射率剖面。例如，Zhu和Bamler利用TerraSAR-X数据展示了适用于城市区域的高分辨率TomoSAR [24], ，而Fornaro 等人则开发了多航过聚焦方法，用于估计单散射体和双散射体高度 [4]。尽管这些基于模型的方法能够实现精确重建，但它们往往难以分离间距较近的散射体，且通常需要数十幅图像。为解决近距离散射体分离问题，研究人员引入了压缩感知技术，用于

TomoSAR反演 [5]。特别是，Zhu和Bamler提出的L1范数正则化方法能够获得超分辨率高程估计，从而提高重建精度和散射体分离能力 [5]。尽管TomoSAR受益于SAR全天候工作的普遍优势，但应注意，作为一种多基线、多时相干涉合成孔径雷达技术，它对微小相位变化十分敏感，尤其容易受到大气延迟的影响；在热带地区，这种影响可能尤为显著。值得庆幸的是，近期研究表明，即使只有3–5个干涉图，也能实现可靠重建 [25], ，显著降低了TomoSAR对数据量的需求。近来，Qian 等人引入了基于学习的TomoSAR反演框架，通过模仿或加速压缩感知，在准确率没有明显损失的情况下提升了处理效率 [26]–[28]。得益于这些进展以及高分辨率合成孔径雷达影像日益丰富，获取全球主要城市的大规模TomoSAR点云已成为可能，能够以前所未有的规模捕获三维结构信息 [29], [30]。

给定一个城市区域的TomoSAR点云，我们可以提取散射体之外的大量建筑相关信息。例如，Yang等人 应用Pol-TomoSAR估计森林区域的高度 [31]，而Armeshi 等人 采用基于TomoSAR的正则化方法检测城市环境中的高度变化 [32]。然而，这两项研究都未生成或利用三维TomoSAR点云进行分析。相比之下，Shahzad和Zhu展示了如何从星载TomoSAR数据自动重建立面和三维建筑形状，并证实可以从此类点云推断建筑几何形态 [33], [34]。Ley 等人 [35] 提出了一种凸优化方法，用于对TomoSAR点云去噪，并填补生成的高度图中的空缺。尽管取得了这些进展，据我们所知，尚无基于学习的方法能够直接从三维TomoSAR点云生成连续的建筑物高度图。

TomoSAR点云通过沿高程轴对多次 SAR 数据采集进行相干合成获得，将垂直分辨单元内的所有回波整合为单一响应。这种数据采集几何会导致强烈的各向异性噪声：高程（z）误差通常比 x 和 y [36]方向的误差大约一个数量级。通过先进的逐点分析和大气校正，绝对平面位置精度可达到厘米级 [37]，但有效高度精度要低得多，因为高程峰值的位置取决于数据采集的数量和空间分布，以及散射体信噪比。对于 TerraSAR-X 数据栈，垂直不确定性通常在 1–20 米左右。因此，与典型的摄影测量或激光雷达点云相比，TomoSAR点云的噪声显著更大、更加稀疏且分布更不均匀，尤其是在低相干表面上。因此，简单插值并不可靠，纯粹的局部处理也不足以解决问题：模型必须捕获全局点云模式并利用空间上下文来降低噪声、填补空洞并抑制伪散射体。由于水平位置远比垂直高度可靠，因此自然适合采用双拓扑设计，在x–y 地图平面中将基于点的分支与基于网格的分支配对，使网格能够对含噪高程进行正则化并确保空间一致性。Peng 等人提出使用卷积编码器将不规则三维点融合为连续的占据网格 [38]，而 Wang 等人则提出一种交替策略，迭代细化网格和点表示 [39]。我们的模型在这些思路基础上，采用独立的点分支和网格分支，但针对 2.5D高度制图进行了调整：我们将所有散射体投影到单一天底平面，而非采用原始的三平面方案，并用细化模块取代占据解码器，直接输出连续的高度值。基于点和基于网格的特征变换采用 U-Net级联 [40],排列，使两个流在每个阶段交替演进，以降低噪声、填补空洞并确保空间一致性。

III. 方法

A. 问题定义

我们的目标是从TomoSAR点云中生成建筑高度图，并将结果表示为归一化数字表面模型（nDSM）。这需要学习一个函数，将输入点映射为基于网格的高度值，同时应对大范围的数据缺失和各向异性噪声。为此，我们设计了一种神经网络，直接预测逐像素高度。令$\mathbf{P} = \{\mathbf{p}_i\}_{i=1}^{N}$ 表示TomoSAR点的集合，其中每个点$\mathbf{p}_i = (x_i, y_i, z_i)$具有空间坐标$(x_i, y_i)$和高程 z_i。我们寻求一个映射f ，使得

$$\hat{\mathbf{H}} = f(\mathbf{P}), \tag{1}$$

其中，$\hat{\mathbf{H}} = \{\hat{h}_j\}_{j=1}^{M}$ 表示规则网格上位置(x_j, y_j)处的估计高度， M 为网格单元的数量。请注意，预测值 h_j 是在网格单元上定义的，而非在单个点上定义。

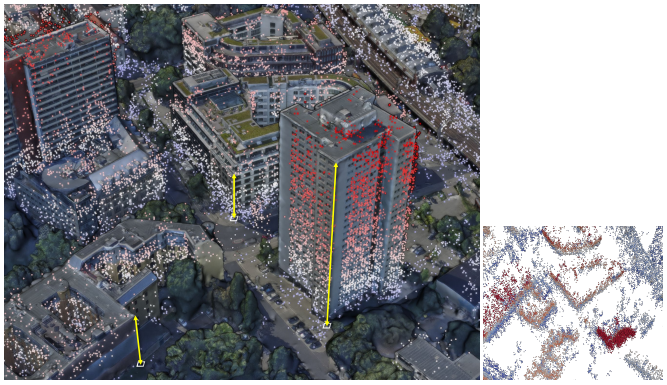


图2. 通过整体分析TomoSAR点云中的立面和邻域点模式，可以推理出建筑物高度。在柏林的这一地点，突出显示了三个逐像素高度值。右侧面板显示俯视图。为便于可视化，点按高度进行颜色编码，并相对于立面稍作偏移。

TomoSAR点云面临的主要挑战，尤其是由数据采集次数有限的条带模式生成的点云，来自较高的噪声水平和各向异性采样。噪声会降低点定位精度，因此有必要利用更广泛的空间上下文。各向异性采样进一步增加了重建的难度：有些区域完全缺少高度信息，而另一些区域则可能

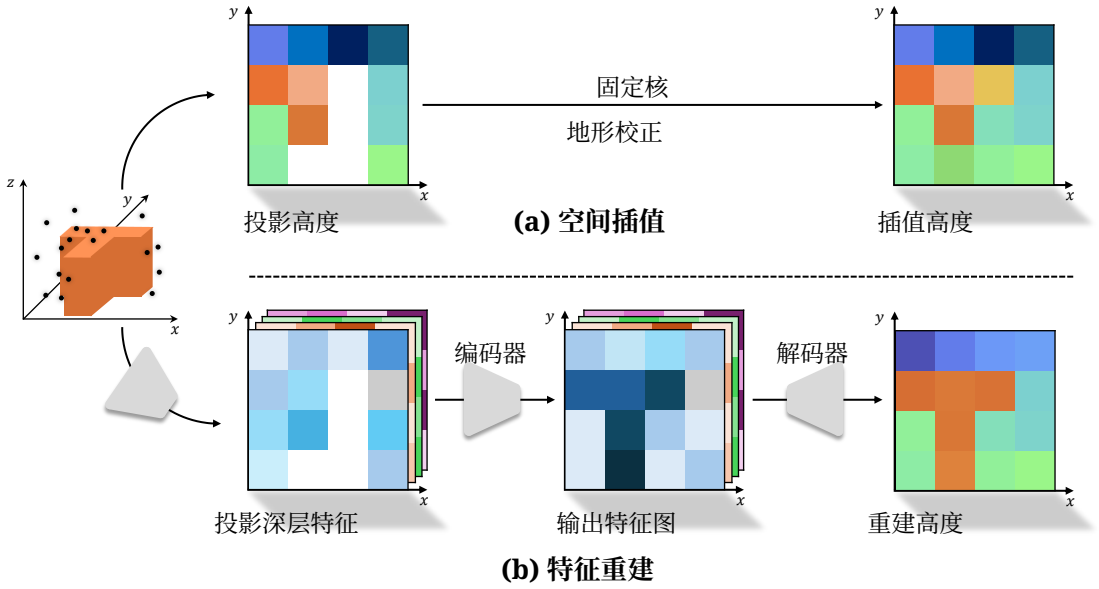


图3。传统空间插值与我们基于特征的高度重建方法的对比。空间插值（a）关注显式高度值，因此容易受到缺失或含噪点分布的影响，并且需要进行地形校正。相比之下，我们的方法（b）利用归纳偏置对从输入点编码得到的深层特征进行细化和补全，从而实现更加稳健、准确的高度重建。

由于斜距成像几何结构而包含多个候选高程，如图1所示。

直接的方法，例如通过从局部最大值中减去地形高程来估计每个（x_j，y_j）处的高度，对噪声高度敏感。在各向异性采样条件下，空间插值同样不可靠（见图3）。解决这些问题需要对点模式进行整体理解。我们认为，通过分析立面和邻域点分布，可以在推理时不依赖精确的点位置或参考数字地形模型来推断建筑物高度，如图2所示。这些挑战促使我们构建一种稳健的模型，使其能够同时对输入点进行去噪，并对缺失值进行图像修复，从而重建出可靠的建筑物高度图$\mathbf{H}$。

B. 所提方案

我们采用数据驱动的方法应对这些挑战。我们不直接在空间域内插值高度，而是将问题提升到深层特征空间，使深度神经网络能够识别全面的点模式、去除观测噪声并修复缺失区域。该方法利用了现代神经网络固有的归纳偏置。图3展示了这一方法的动机。

为了有效处理稀疏且含噪的输入点，我们的网络由三个主要组件构成：(1) 点到网格编码器，用于提取点特征并将其聚合到网格上；(2) 双拓扑细化模块，在点表示与网格表示之间交替转换，以逐步增强特征；(3) 网格特征解码器，用于生成高度图，并可选择同时生成辅助输出。图 4 概述了该架构及数据在各阶段的流动过程。

1) 点到网格编码器：我们从输入的三维点中提取特征，并根据这些点的空间坐标将特征映射到二维网格上，从而将无序点转换为结构化表示。

a) 特征提取： 为了将点 $\mathbf{P}$ 编码为能够捕捉超越单个点的结构信息的潜在特征，我们采用编码器网络 $f_e: \mathbb{R}^{N\times 3} \to \mathbb{R}^{N\times d}$，该网络利用这些点的空间关系，并生成一组$d$维特征 $\mathbf{Z} = \{\mathbf{z}_i\}_{i=1}^{N}$:

$$\mathbf{Z} = f_e(\mathbf{P}), \tag{2}$$

其中 f_e 通过一组PointNet层 [41]实现，并采用局部池化将特征汇聚到网格上，如图5所示。

b) 特征投影： 随后将点特征$\mathbf{Z}$ 投影到水平二维网格特征$\mathbf{G} = \{\mathbf{g}_j\}_{j=1}^{M}$上，其中每个$\mathbf{g}_j$ 仍是一个 d维向量。紧接着，对落在同一网格单元中的投影特征进行聚合，采用平均池化。该串联过程用 $\mathcal{P}roj$表示，具体如下:

$$\mathcal{P}roj : \{\mathbf{z}_i\}_{i=1}^{N} \mapsto \{\mathbf{g}_j\}_{j=1}^{M}, \tag{3}$$

其中，cell(j)表示水平面上网格单元j 所覆盖的空间区域。聚合后的特征向量$\mathbf{g}_j$ 为

$$\mathbf{g}_j = \frac{1}{|\mathcal{I}(j)|} \sum_{k\in\mathcal{I}(j)} \mathbf{z}_k, \tag{4}$$

其中 $\mathcal{I}(j) = \{k \mid (x_k,\ y_k) \in \text{cell}(j)\}$。图5（下）展示了这一投影过程。

2) 跨拓扑细化：点到网格编码器生成的初始网格特征较为粗糙，且由于输入点中存在噪声，部分特征不可靠。此外，各向异性采样会导致许多网格单元为空，因此需要用零填充其特征，以保持特征维度一致。为了解决这些问题，我们通过在基于点的

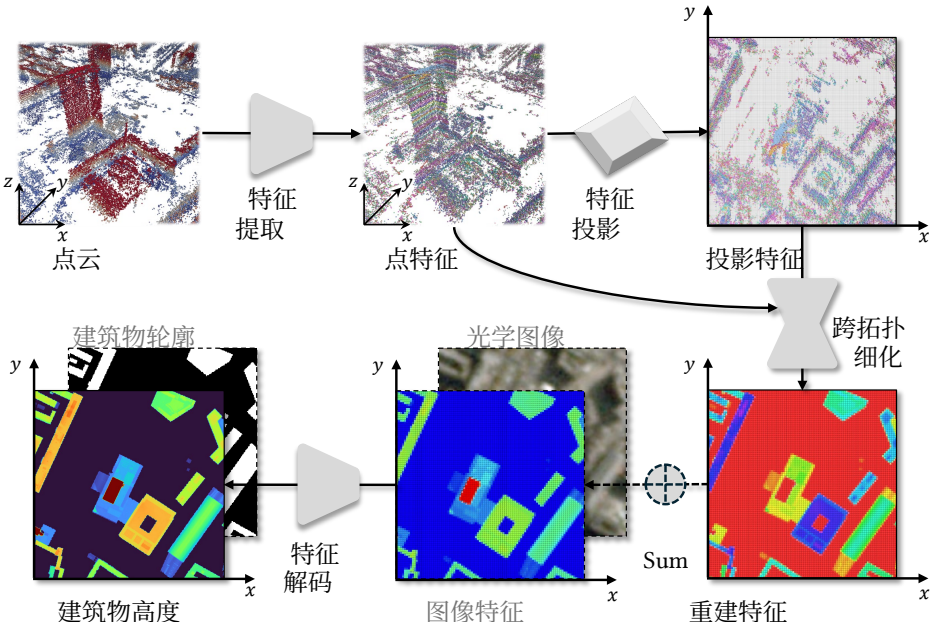


图 4. 所提出工作流程的概述。从 TomoSAR点云出发，我们首先提取逐点特征并进行局部池化。随后将这些特征投影到二维网格上，形成水平特征平面；跨拓扑细化通过在点与网格单元之间交换信息，迭代改进特征表示。最后，基于网格的解码器根据细化后的网格特征预测建筑物高度，生成连贯的高度图。可选的光学图像特征和建筑物轮廓监督以虚线轮廓表示。

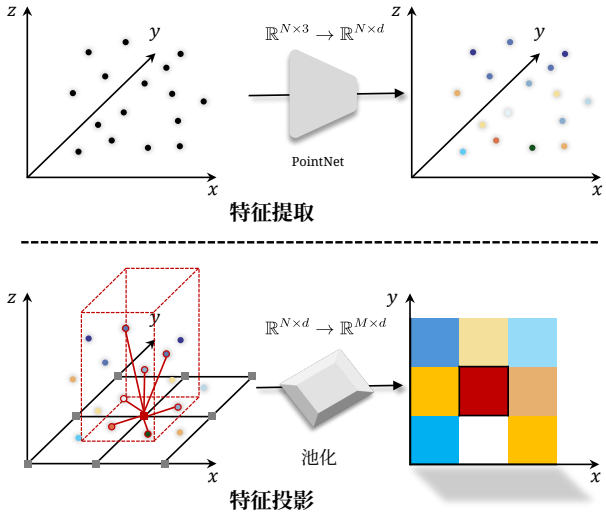


图5. 特征编码过程。上图：给定一个包含 N 个点的TomoSAR点云（x, y, z），我们首先使用 PointNet 提取逐点特征$\mathbf{Z} \in \mathbb{R}^{N\times d}$。下图：随后，我们将这些点特征投影到二维网格上，对落入每个网格单元的所有点的特征进行平均池化，得到网格特征$\mathbf{G} \in \mathbb{R}^{M\times d}$。

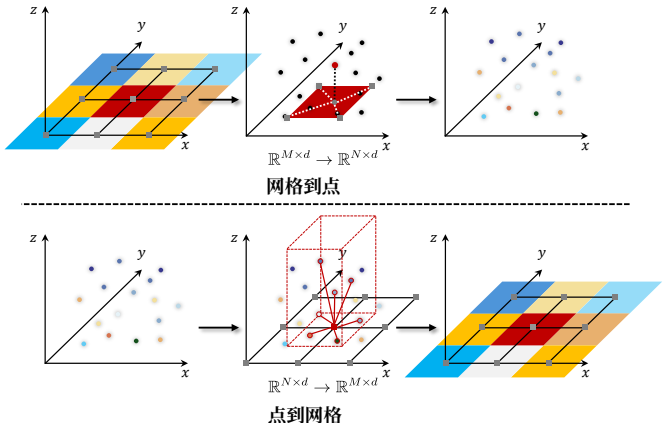


图6. 跨拓扑特征细化过程。上图：在网格到点步骤中，网格特征$\mathbf{G}^{(l)} \in \mathbb{R}^{M\times d}$ 被插值回点域$\mathbf{Z}^{(l+1)} \in \mathbb{R}^{N\times d}$。下图：在点到网格步骤中，细化后的点特征$\mathbf{Z}^{(l+1)}$ 被投影到网格上，并通过池化聚合每个网格单元内的点级信息。由此得到更新后的网格特征$\mathbf{G}^{(l+1)}$。

表示 $\mathbf{Z}^{(l)}$ 与基于网格的表示 $\mathbf{G}^{(l)}$之间交替迭代，来细化网格特征，其中 $l \in \{0, \ldots, L\}$ 表示迭代索引。图6展示了这种跨拓扑细化。这里，$\mathbf{Z}^{(0)}$ 和$\mathbf{G}^{(0)}$分别由公式2和公式4得到。

a) 网格到点变换：在第 l 次迭代中，网格特征 $\mathbf{G}^{(l)}$ 被映射回点特征 $\mathbf{Z}^{(l+1)}$，其中每个点 $\mathbf{p}_i \in \mathbf{P}$ 被投影到二维网格上，其特征 $\mathbf{z}_i^{(l+1)}$ 通过对网格特征进行双线性插值获得。设 $\mathcal{N}_i$ 表示投影位置 (x_i, y_i) 周围的四个网格单元，α_j 表示相应的插值权重。于是，

$$\mathbf{z}_i^{(l+1)} = \sum_{j\in\mathcal{N}_i} \alpha_j \cdot \mathbf{g}_j^{(l)}, \tag{5}$$

其中$\mathbf{g}_j^{(l)}$ 是第 j 个网格单元的网格特征。汇集所有点特征可得$\mathbf{Z}^{(l+1)} = \{\mathbf{z}_i^{(l+1)}\}_{i=1}^{N}$。

b) 点到网格变换：给定更新后的点特征 $\mathbf{Z}^{(l+1)}$，我们首先通过多层感知机对其进行处理，以获得更细的粒度：

$$\tilde{\mathbf{z}}_i^{(l+1)} = \mathrm{MLP}\left(\mathbf{z}_i^{(l+1)}\right). \tag{6}$$

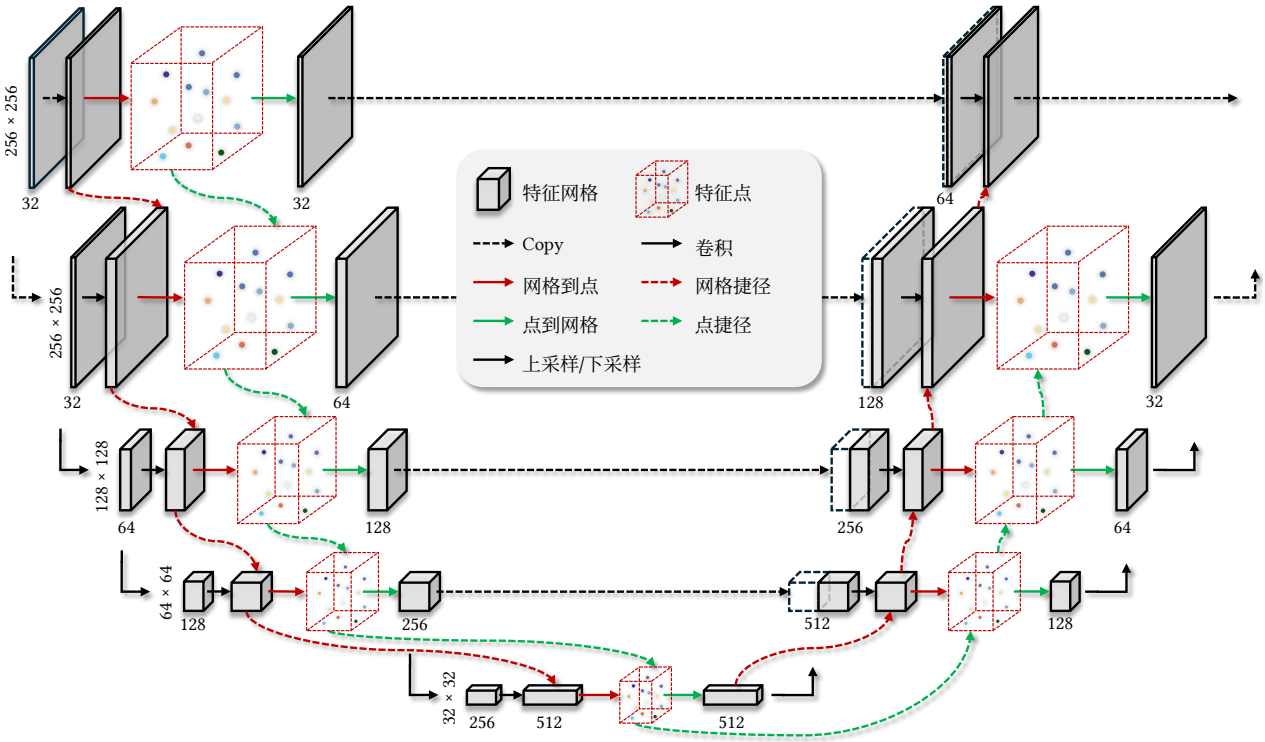


图7. 特征重建网络架构概览。该模型采用多尺度类 U-Net 设计，并融入跨拓扑细化。在每个尺度上，网格特征与点特征通过网格到点和点到网格变换交换信息（见图6）。两条路径上的跳跃连接有助于保留并传递跨尺度的细粒度细节。

然后，我们将细化后的点特征重新投影到网格上，并通过平均池化聚合特征：

$$\mathbf{g}_j^{(l+1)} = \frac{1}{|\mathcal{I}(j)|} \sum_{k \in \mathcal{I}(j)} \tilde{\mathbf{z}}_k^{(l+1)}. \tag{7}$$

由此得到更新后的网格特征 $\mathbf{G}^{(l+1)} = \{\mathbf{g}_j^{(l+1)}\}_{j=1}^{M}$。

为保留细节特征，我们在连续变换之间加入跳跃连接。遵循 Wang等人 [39], 的交替设计，我们采用 U-Net [40] 式结构来组织这些变换。随着细化过程的推进，$\mathbf{G}^{(l)}$和$\mathbf{Z}^{(l)}$ 协同演进，细化网格特征和点特征。

3）网格特征解码器：最终的细化后网格特征$\mathbf{G}^{(L)} \in \mathbb{R}^{M \times d}$ 用于重建建筑物高度图 $\hat{}\mathbf{H}$。此外，我们引入一个辅助分支，用于预测建筑物轮廓 $\hat{}\mathbf{A} = \{\hat{a}_j\}_{j=1}^{M}$，提供额外监督，从而增强模型对噪声的鲁棒性。

a) 高度图解码：随后将细化后的网格特征输入浅层卷积网络解码器 f_h ，以生成高度图$\mathbf{H}$：

$$\hat{\mathbf{H}} = f_h(\mathbf{G}^{(L)}). \tag{8}$$

b) 辅助解码：使用另一个浅层卷积网络的辅助解码器 f_a 来预测建筑物轮廓$\mathbf{A}$：

$$\hat{\mathbf{A}} = f_a(\mathbf{G}^{(L)}). \tag{9}$$

该辅助分支用于正则化神经网络，使其在处理噪声很大且稀疏的输入点时能够做出更稳健的预测。

4）优化：我们通过最小化高度估计误差来端到端训练模型，并辅以辅助轮廓损失。

a) 高度重建损失：我们使用预测高度图$\mathbf{H}$与真实高度图$\mathbf{H}$之间的平均绝对误差：

$$\mathcal{L}_h = \frac{1}{M} \sum_{j=1}^{M} \left| \hat{h}_j - h_j \right|. \tag{10}$$

b) 辅助损失：对于建筑物轮廓预测，我们在预测轮廓概率$\mathbf{A}$ 与真实值$\mathbf{A}$之间应用二元交叉熵：

$$\mathcal{L}_a = -\frac{1}{M} \sum_{j=1}^{M} [a_j \log \hat{a}_j + (1 - a_j) \log(1 - \hat{a}_j)]. \tag{11}$$

c) 总损失：总目标是高度重建损失与辅助损失的加权和：

$$\mathcal{L} = \mathcal{L}_{\mathrm{h}} + \beta \mathcal{L}_{\mathrm{a}}, \tag{12}$$

其中 β 为加权因子。

d) 后处理： 在推理过程中，我们会预测输入区域中重叠图像块的高度图。每个图像块$\hat{}$在其空间范围内提供一个局部估计$\mathbf{H}_t$。为生成连贯高度图$\mathbf{H}$并避免边缘伪影，我们使用重叠区域中的加权融合将图像块拼接起来，如图8所示。具体而言，令$\mathbf{w}_t$ 表示一种空间

融合权重，该权重朝图像块边缘线性递减。最终高度图表示为：

$$\hat{\mathbf{H}} = \max\left(0, \frac{\sum \mathbf{w}_t \cdot \hat{\mathbf{H}}_t}{\sum \mathbf{w}_t}\right), \tag{13}$$

其中，max(0, ·) 用于修正非物理预测结果，因为建筑物高度不能为负。

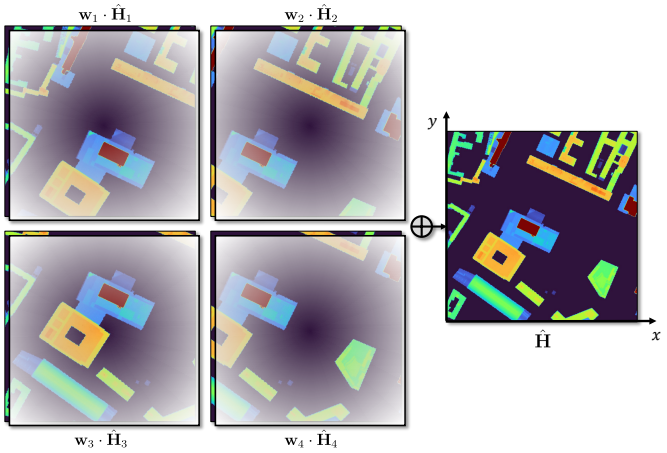


图 8. 通过图块融合生成连贯高度图。多个重叠图像块分别提供局部高度估计值 $\hat{}\mathbf{H}_t$，并按元素加权，权重 $\mathbf{w}_t$ 朝图块边缘线性递减。随后对这些加权估计值取平均，生成无缝的最终高度图 $\hat{}\mathbf{H}$。

IV. 实验与结果分析

A. TomoSAR 数据预处理

SAR数据集包括柏林地区的一组 TerraSAR-X 高分辨率聚束模式图像，空间分辨率约为 1m；以及慕尼黑地区的一组条带模式图像，空间分辨率约为 3 m。柏林数据集采集于2008年至2013年，包含108幅干涉图；慕尼黑数据集采集于2011年至2013年，仅包含5幅干涉图。表 I 总结了采集参数，图9采用相同的符号说明了TomoSAR 原理。在进行TomoSAR处理之前，对SAR图像进行了配准，并校正了大气相位。我们使用“SVD–Wiener”算法处理柏林数据 [24]，使用“NLCS–TomoSAR”算法处理慕尼黑数据 [42]。处理结果为五维点云，包括三维位置、线性形变速率和季节性运动振幅。在此分辨率下，必须考虑这两个运动分量，才能实现精确的三维重建[43]，，因为城市结构和地表可能会因热膨胀、沉降或隆升而发生位移。Wang 等人 [29] 和 Shi 等人[25] 分别提供了柏林和慕尼黑数据的更多详情。由于可获取光学卫星图像和归一化数字表面模型等补充数据源，我们选取慕尼黑和柏林的市中心区域作为研究区。图10展示了这些区域的点云。数据被划分为训练、验证和测试子集。为专门关注建筑物高度，我们应用地籍建筑物轮廓掩膜以从TomoSAR点云和归一化数字表面模型标签中移除非建筑结构。

表 I 合成孔径雷达数据采集参数。

描述	符号	慕尼黑	柏林
距中心距离	r	698 千米	624 km
波长	λ	3.1 cm	3.1 cm
入射角	θ	50.4°	36.1°
最大高程孔径	Δb	187 米	363 米
干涉图数量	N	5	108

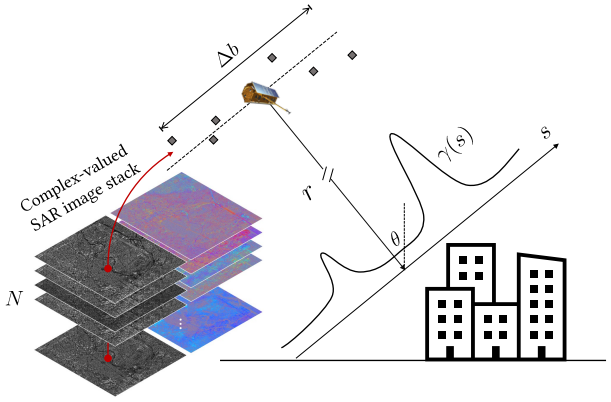


图 9. TomoSAR 数据采集与垂直反射率重建原理。系统从不同观测角度获取多个复数值 SAR 图像，其最大高程孔径为 Δb。TomoSAR 根据这些图像沿高程 s 重建反射率剖面 $\gamma(s)$。这一过程将 SAR 图像序列转换为表征建筑结构的三维点云。符号含义见表 I。

B. 实验设置

1) 实现细节：慕尼黑的建筑物高度归一化数字表面模型参考数据使用机载LiDAR数据生成[1]，，而柏林数据则来自官方摄影测量数据源[2]。两者的空间分辨率均为 1 米。我们使用带权重衰减的 Adam 优化器训练模型。公式 12 中的损失权重 β 在所有实验中固定为 10。训练期间，我们在每个区域内随机选取中心点，并剔除窗口会跨越区域边界的样本，动态采样图像块，总计执行 10,000 步。这样可确保每个图像块不会跨越两个数据划分，从而避免训练集、验证集和测试集之间的数据泄漏。为提高效率，我们将每个区域存储为若干个不重叠的磁盘分块，并在单个分块内采样每个图像块，因此额外开销可以忽略不计。我们选择验证性能最佳的检查点，并在测试集上进行评估。我们采用周期为 1,000 步的循环学习率调度策略，每个周期结束后将最大学习率减半。

2) 评估指标：为全面评估模型的重建性能，我们采用以下三种误差指标：

[1]https://geodaten.bayern.de/opengeodata/ [2] https://gdi.berlin.de/

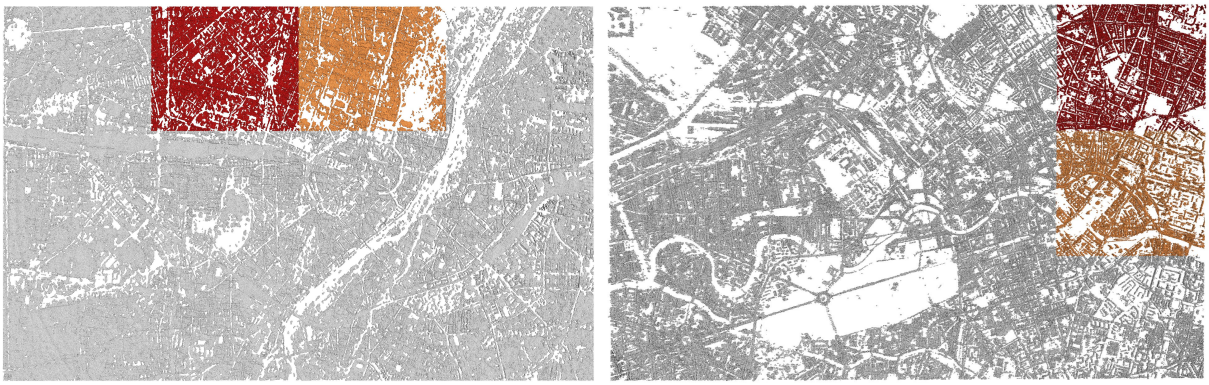


图 10. 本研究所用 TomoSAR点云的覆盖范围。橙色区域（慕尼黑：4.32 km^2；柏林：3.33 km^2）和红色区域（慕尼黑：4.32 km^2；柏林：3.25 km^2）分别用于模型设计验证和最终评估，灰色区域则用于模型训练。

表 II 柏林和慕尼黑的定量高度重建结果（米）。

City	总体面积			建筑面积			建筑实例		
	平均绝对误差	均方根误差	中位绝对误差	平均绝对误差	均方根误差	中位绝对误差	平均绝对误差	均方根误差	中位绝对误差
柏林	2.10	5.46	0.00	4.64	8.12	1.55	3.69	6.17	2.32
慕尼黑	3.27	6.44	0.04	6.38	9.04	4.26	5.06	6.87	3.31

- 平均绝对误差（MAE）：该指标用于衡量一组预测值中误差的平均大小，不考虑误差的方向。其计算方式为：

$$\mathrm{MAE} = \frac{1}{M}\sum_{j=1}^{M}\left|h_j - \hat{h}_j\right|. \tag{14}$$

- 均方根误差（RMSE）：均方根误差通过先对差值平方，再求平均并开平方，对较大的误差更为敏感。其计算公式为：

$$\mathrm{RMSE} = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\left(h_j - \hat{h}_j\right)^2}. \tag{15}$$

- 中位绝对误差（MedAE）：MedAE关注绝对误差的中位数，因此对离群值更稳健。其计算公式为：

$$\mathrm{MedAE} = \mathrm{median}_{j=1}^{M}\left|h_j - \hat{h}_j\right|. \tag{16}$$

我们针对以下区域报告这些指标，以提供更细致的性能评估：

- 总体面积：这些指标基于测试集划分中的所有像素计算。除非另有说明，否则默认报告总体指标。
- 建筑面积：这些指标基于建筑物轮廓内的像素计算。
- 建筑实例：这些指标针对每栋建筑计算，方法是先取每个轮廓内预测高度的中位数，再对所有建筑实例进行汇总。

^这里，h_j 和 h_j 分别表示像素 j处的实际高度值和预测高度值，M 表示评估像素总数。

C. 基于柏林数据的重建

表 II 报告了测试区域的误差，在柏林数据集上取得的总体平均绝对误差为 2.10m。图 11 展示了两个选定图像块的定性结果，而图 13 则概述了整个测试区域的结果。这些结果表明，该模型能够有效地修复输入点云中的缺失区域，并准确恢复大多数结构。然而，图 16 中的误差分布显示，在处理更复杂的建筑结构时，该模型的误差率较高。

D. 慕尼黑数据的重建

一种直接的方法是将同一模型应用于慕尼黑数据。然而，由于慕尼黑点的噪声更严重、稀疏性更高且分布高度各向异性，仅预测高度图的基线方法无法得出合理结果，如表 IV 所示。进一步检查发现，经过非负校正前，预测高度集中在零附近，这表明网络难以识别有意义的建筑信号。为解决这一问题，我们引入来自建筑物轮廓掩膜的辅助监督，通过鼓励模型区分建筑物轮廓内外的样本来规范训练。这些轮廓掩膜仅用于训练；推理时，模型完全依赖原始 TomoSAR 点，因此即使没有辅助数据，也能大规模部署。

如表 II 所示，慕尼黑重建的高度图保真度低于柏林，平均绝对误差为 3.27m。图 16 进一步表明，误差会随着建筑物高度增加而增大，这凸显了对较高建筑进行建模的挑战。从图 12 和图 13 可以明显看出，

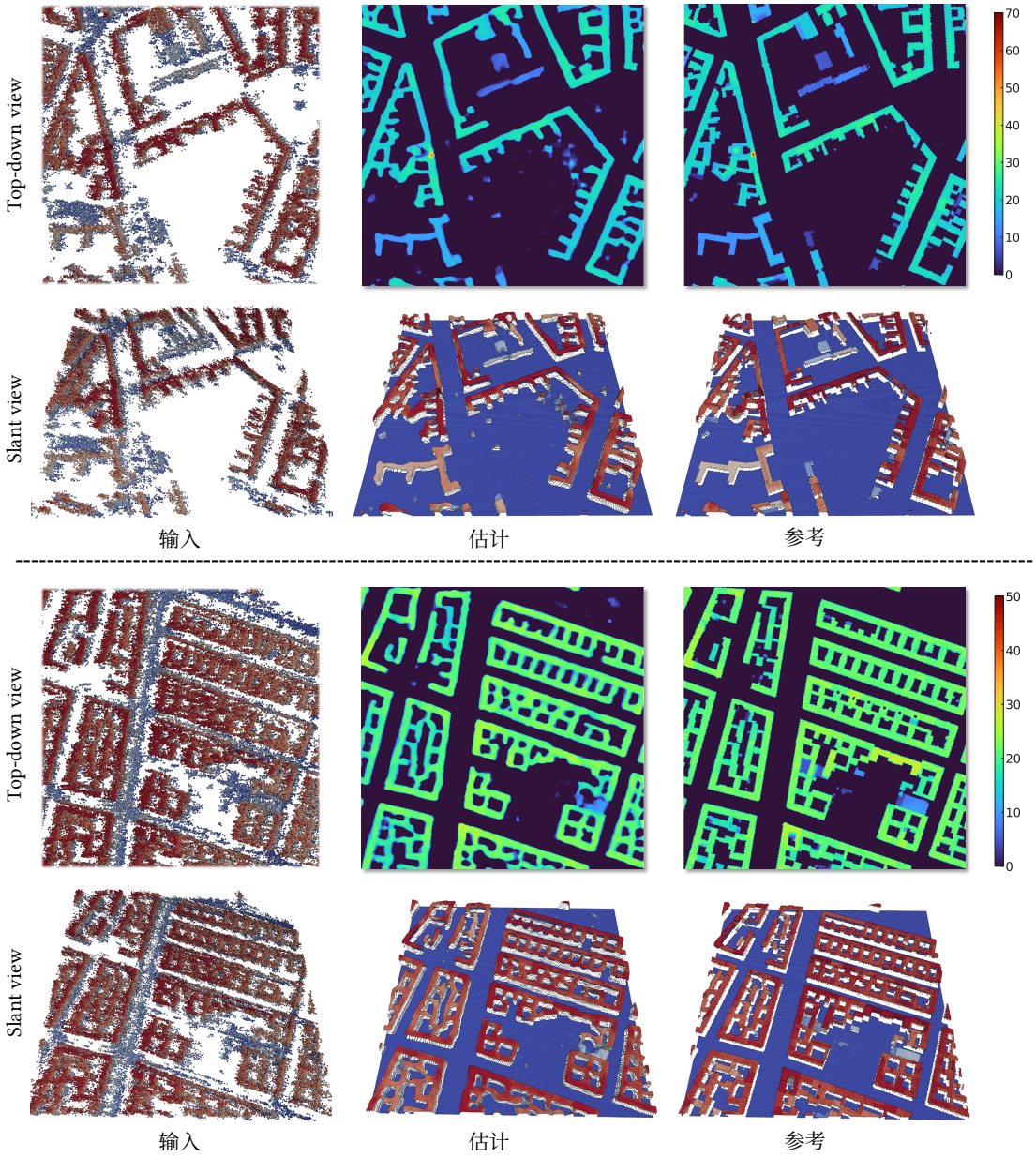


图 11. 柏林两个区域的定性高度重建结果。在斜视图中，高度图被重新投影为三维点。结果展示了该模型对含噪输入和各向异性采样的鲁棒性。

由于缺少输入点，边界的规则性较差，且部分建成区缺失。尽管如此，预测的高度图呈现出更清晰的城市结构，而这些结构在原始点云中大多被噪声掩盖。考虑到输入数据的质量，这些结果表明，该方法能够利用质量欠佳的TomoSAR数据堆栈。与柏林相比，慕尼黑的条带式SAR数据代表了一种更具成本效益的数据采集方式，有望实现更广泛的覆盖并生成大范围制图产品。

E. 消融实验

表 III 将我们的方法与插值方法进行了比较。双线性插值和反距离加权的表现都不理想，因为它们无法捕捉噪声和各向异性

表 III 插值方法与我们的方法的评估。双线性插值和反距离加权在显著噪声和各向异性数据分布下表现不佳，并且在推理阶段需要数字地形模型。我们的网络误差更低，且仅在训练数据准备期间使用数字地形模型。

方法	是否需要数字地形模型		平均绝对误差（米）	
	训练	推理	柏林	慕尼黑
我们的神经网络	✓	✗	**2.10**	**3.27**
双线性插值	n/a	✓	5.44	6.84
反距离加权插值	n/a	✓	5.53	6.85

数据的点分布。从图14可以看出，双线性插值会产生非常嘈杂的高度值。

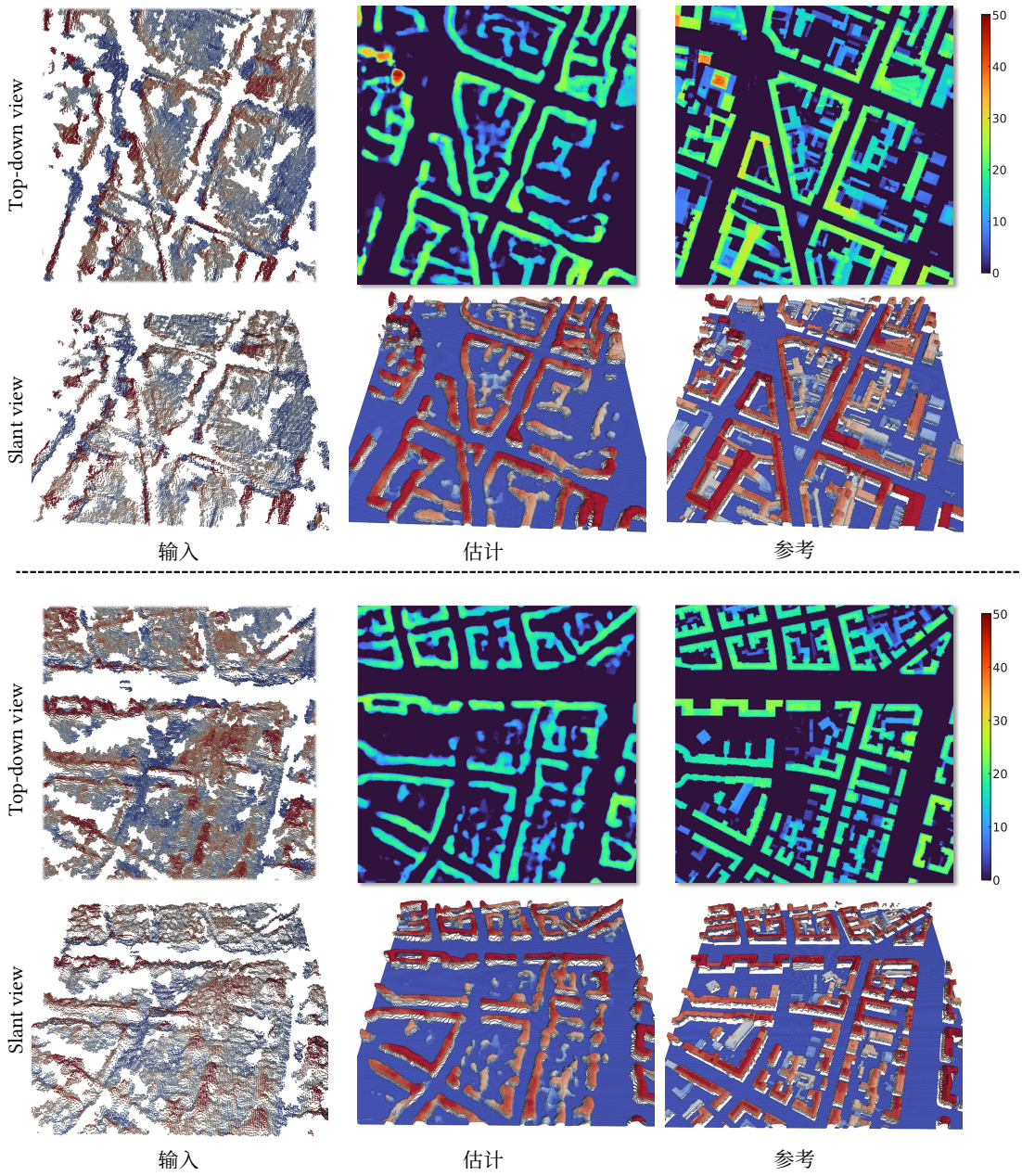


图 12. 慕尼黑的定性高度重建结果。在斜视图中，栅格化结果被重新投影为三维点。结果表明，该模型在处理具有各向异性空间分布的含噪输入时具有良好的鲁棒性。

此外，这些显式技术依赖数字地形模型来减轻地形效应。相比之下，我们提出的神经网络能够学习一种稳健的数据驱动映射，直接生成建筑物高度图，而无需将数字地形模型作为输入，因此是一种适用性更广的解决方案。

图15比较了我们的双拓扑网络与仅基于网格拓扑的原始 U-Net 的中间特征图。通过跨拓扑细化，双拓扑网络能够保留更清晰的结构线索和更鲜明的空间模式。表 IV 进一步展示了点网格变换和辅助轮廓分支的影响。在两个数据集上，融入点拓扑都能提升性能。有趣的是，掩码预测分支提供的监督仅对慕尼黑数据有效，这可能是因为只有质量较低的数据能从额外的正则化中获益。相比之下，柏林数据并未以同样的方式获益；添加辅助任务可能会使学习目标偏离高度预测。由于性能差异很小，用户可以根据数据质量决定是否采用此选项。值得注意的是，采用二维局部池化的 PointNet 编码器 [38] 优于资源消耗更高、采用三维聚合的 PointNet ++ 变体 [44] ，这表明在规则网格上利用局部结构上下文对高度估计尤为有益。

表 V 总结了网络深度（即，如方程 8 中的 L ）和特征平面分辨率的影响。对于柏林，5 层网络足以捕捉基本结构。增加深度并不能提升性能，反而会大幅增加

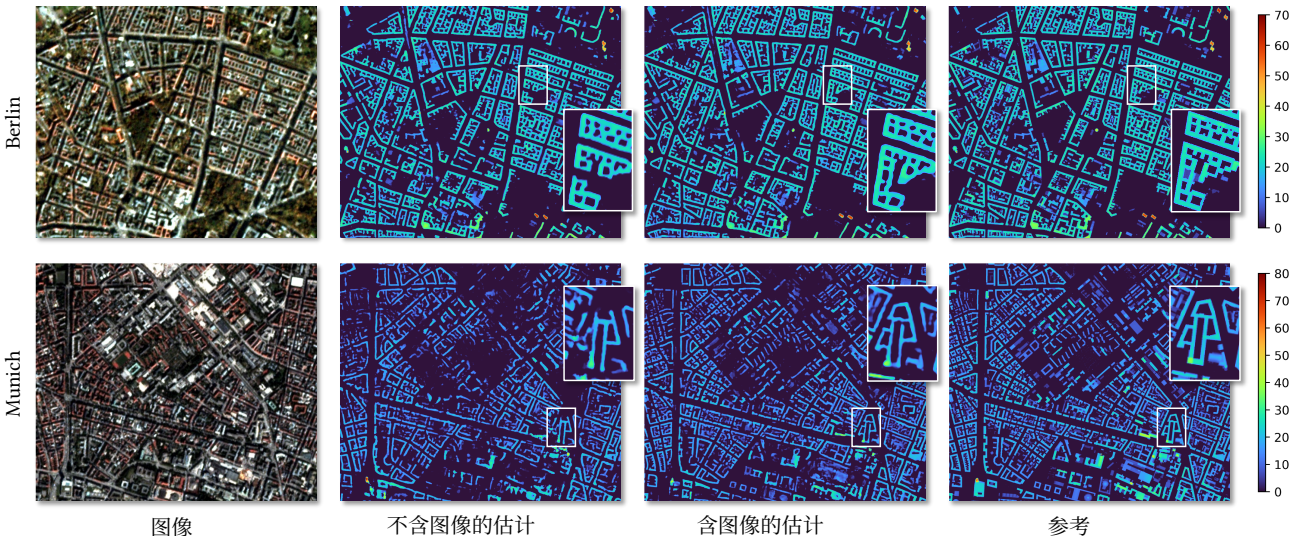


图 13. 两个数据集完整测试区域上的重建结果。加入光学图像特征可提升精度，尤其是在细节丰富以及点覆盖稀疏或不均匀的区域。定量结果见表 VI。

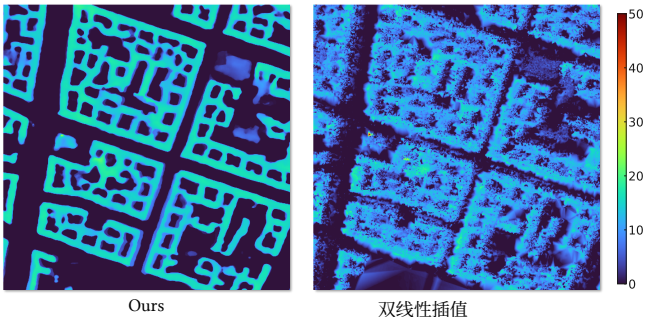


图 14. 柏林某地点采用我们的方法与双线性插值生成的高度图对比。双线性插值依赖数字地形模型，而我们在推理时使用原始点，因此能够获得更准确的重建结果。

表 IV 神经网络组件消融。浅蓝色表示我们的默认配置。“辅助”表示辅助建筑轮廓监督。

组件			平均绝对误差（米）	
Aux.	点拓扑	编码器	柏林	慕尼黑
✓	✓	PointNet 搭配局部池化	2.16	**3.27**
✗	✓	PointNet 搭配局部池化	**2.10**	4.76
✓	✗	PointNet 搭配局部池化	2.43	3.40
✗	✗	PointNet 搭配局部池化	2.38	4.76
-	✓	PointNet++	3.47	4.24

表 V 网络深度和特征平面分辨率的消融实验。浅蓝色标示我们的默认配置。

配置	数值	参数数量	平均绝对误差（米）	
			柏林	慕尼黑
深度	5	11.1 米	**2.10**	3.31
	6	43.4 米	2.13	3.27
	7	172.3 米	2.27	**3.25**
分辨率	128	43.4 米	2.28	3.34
	256	43.4 米	**2.10**	**3.27**
	512	43.4 米	2.18	3.30

参数数量和训练成本。相比之下，对于慕尼黑，质量较低的数据会受益于更深的网络，这表明需要更强的上下文建模能力。为平衡准确性与复杂度，我们将网络深度设为 6 层。此外，特征图分辨率也会影响性能：较高的分辨率能更好地保留特征细节，但也需要更强的图像修复能力来补偿缺失值；较低的分辨率则能减少神经网络补全缺失值所需的工作量，但可能会使结构变得平滑。这两个因素相互制衡，因此存在一个平衡点。对于柏林和慕尼黑的数据，256×256 的分辨率表现最佳。

F. 融入光学卫星图像

我们的框架具有可扩展性，支持将光学影像整合到处理流程中。为展示这一能力，我们融入了分辨率为 3–5 m 的 PlanetScope 光学卫星图像 [45]，这些图像在研究区域内最早可追溯至 2017 年。与TomoSAR点云类似，这些图像也能提供大范围覆盖。我们使用一个 6 层 U-Net 对图像进行编码，以生成与网格对齐的特征

其维度与TomoSAR网格特征相匹配，如图 4 所示。表 VI 报告了将影像作为额外输入时的结果。TomoSAR点云中的几何信息与影像中的语义信息相互补充，与单独使用任一信息源相比，同时使用两者可提升高度预测效果。如图 13 所示，融合两种信息源能够生成更规整的预测结果。结果还凸显了将TomoSAR点云整合到基于影像的处理流程中的益处。图 16 进一步表明，加入影像后，在整个高度范围内的重建误差均有所降低，且在较低的——

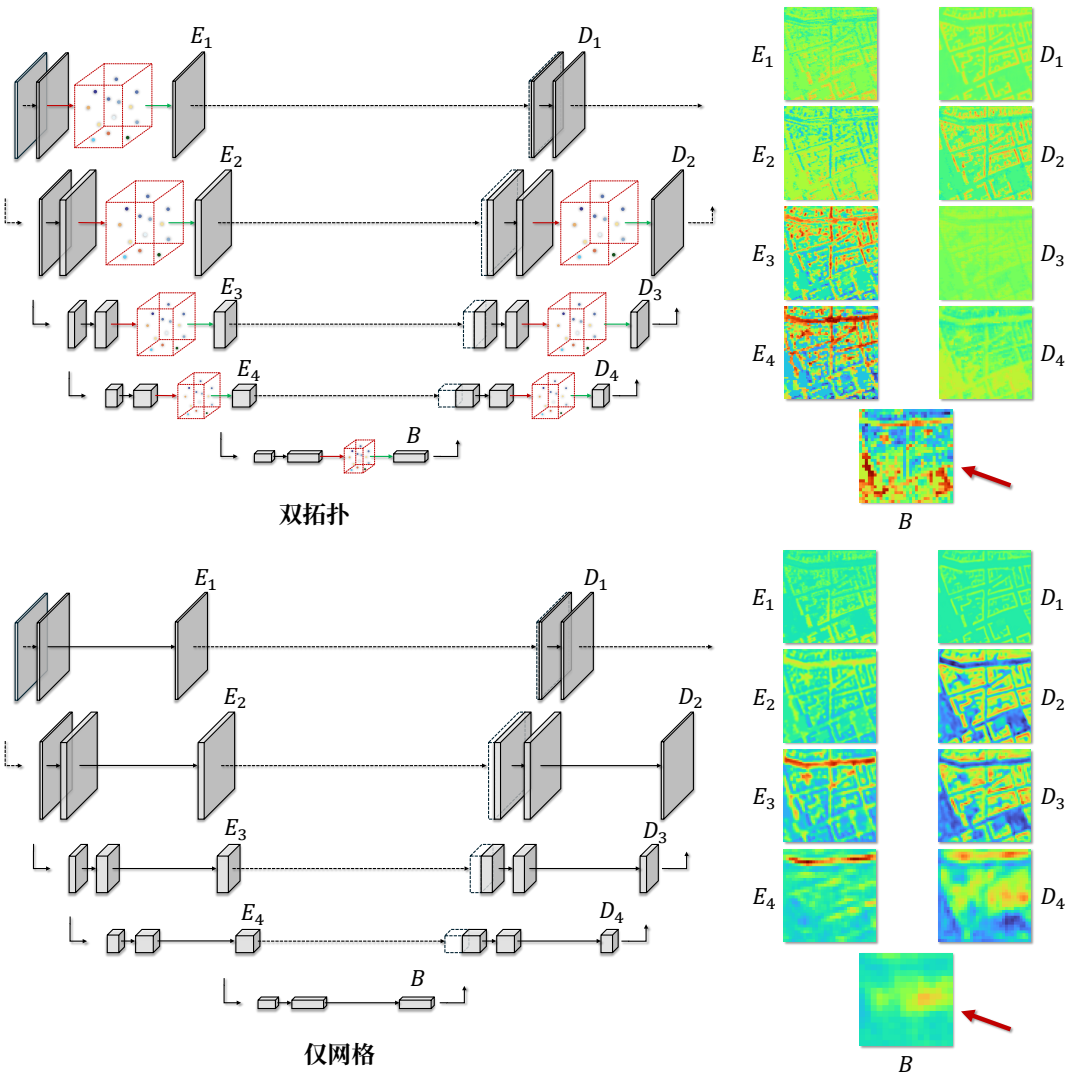


图15. 双拓扑网络与仅网格U-Net之间的中间特征图比较。每对列（E_i、 D_i）展示给定尺度下的编码器和解码器特征图，而 B 表示瓶颈特征图。借助跨拓扑细化，双拓扑网络保留了高频特征，呈现出更清晰的建筑结构和更明显的空间模式。

质量慕尼黑数据集。影像分支与骨干网络无关，且为可选项（例如，在云覆盖的情况下）；这一设计使我们能够利用所有可用数据，同时不影响大规模部署。我们的框架之所以灵活，是因为它将特征投影到天底网格上，从而使此类融合易于实现。本实验旨在展示可扩展性，而非优化性能。

表 VI 不同输入来源的平均绝对误差（米）。TomoSAR点几何（**P**）与基于图像的语义（**I**）具有互补性；将二者结合可获得最低误差。

输入	总体面积		建筑面积		建筑实例	
	柏林	慕尼黑	柏林	慕尼黑	柏林	慕尼黑
P	2.10	3.27	4.64	6.38	3.69	5.06
I	2.37	2.54	5.31	5.12	4.61	3.46
P&I	**2.00**	**2.18**	**4.46**	**4.54**	**3.54**	**3.31**

五、讨论

A. 技术依据和局限性

在本节中，我们讨论所提出方法的若干关键技术方面及其局限性。

a) 归纳偏置：我们的方法受益于卷积神经网络的归纳偏置，即使 TomoSAR 点存在噪声且采样具有各向异性，这些偏置也有助于形成空间连贯表征。尽管该设计在概念上十分直观，模型仍能学习到稳健的结构先验使其能够在具有挑战性的条件下实现可靠的高度重建，包括在严重退化的慕尼黑条带式数据上。我们还发现，该网络具备一定程度的平移不变性：即使输入点存在最多数个像素（相当于数米）的配准误差，模型仍能保持稳定。这一特性进一步表明，学习到的先验有助于网络恢复合理的

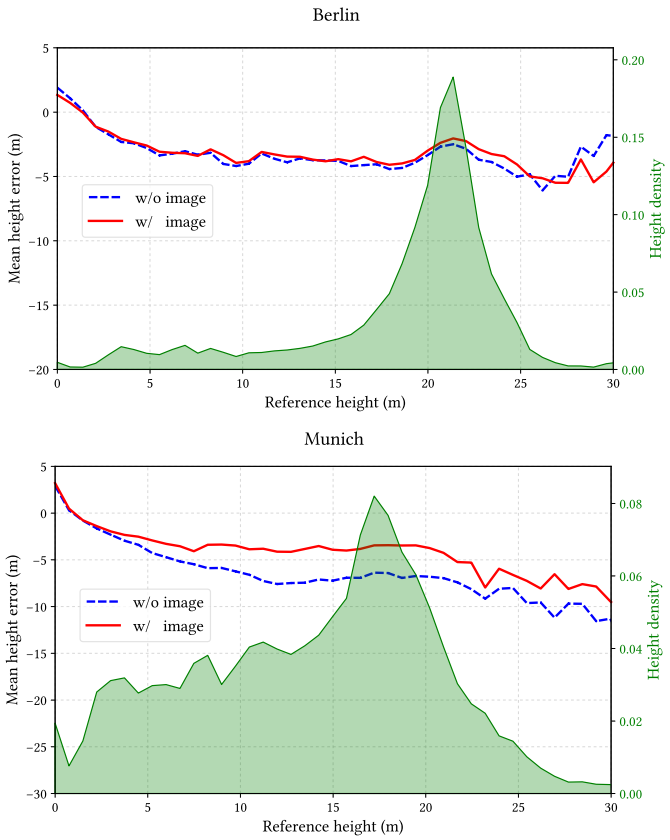


图16. 相对于真实高度值的重建误差。融合图像中的语义信息可降低误差，尤其是对于慕尼黑数据。为便于清晰展示，仅显示 0–30 m 范围内的高度值，并排除高度为零的值。

高度图，尽管存在不完美的对齐。虽然目前我们的验证仅涵盖两座城市，但我们预计该方法能够推广到具有相似 SAR 堆栈特征（例如干涉图数量、成像模式、基线分布）的其他城市区域，因为所有点云均通过基于物理的方法处理，并由相同的成像模式（聚束模式或条带模式）生成。

b) 数据融合：只要能够将多个输入数据源的特征映射到正交平面上，该网络就可以对其进行整合。在我们的实验中，加入 PlanetScope 光学影像提供了互补的语义线索并提高了精度，尤其是在质量较低的慕尼黑数据上；由于该数据中的点更加稀疏且噪声更大，因此图像引导的细化有更大的发挥空间。这种统一表示使二维和三维输入的融合变得十分直接。然而，特定模态的问题仍然存在，例如光学图像中的云覆盖；总体重建质量仍受输入数据保真度的限制。

c) 标签一致性：为尽量减少时间变化，我们将评估重点放在最稳定的城市区域。我们还从归一化数字表面模型中滤除了所有非建筑要素，包括地形以及桥梁、道路等其他基础设施。这一预处理可能会给参考高度图引入噪声；不同数据源之间的差异也会加剧这一问题，并可能导致图 16 中观察到的系统性低估。更广泛地说，我们的处理流程仍取决于参考高度图的准确性和一致性。值得注意的是，只要有相应的标签，同一处理流程也可用于纳入其他静止物体。不过，这会带来额外挑战，例如如何从参考数据中分离树木或汽车等物体。表VII 按建筑类型细分了平均绝对误差。差距较小且方向不一，这表明性能差异是由结构复杂度和局部数据质量驱动的，而非建筑类型。

表VII 按建筑类型划分的平均绝对误差（m）。

柏林		慕尼黑	
住宅类	非住宅类	住宅类	非住宅类
4.67	4.55	6.09	6.55

B. 未来机遇

受上述局限性的启发，我们概述了未来工作的两个方向。

a) 不确定性感知建模：TomoSAR点云通常通过基于模型的反演获得，在此过程中，估计值的不确定性有明确的数学表述，可根据输入数据质量（例如信噪比）进行量化。除了强烈的各向异性误差外，由于合成孔径雷达观测的动态范围很大，各点的噪声水平也存在显著差异。我们的网络设计尚未考虑这种异方差不确定性。目前，我们对所有输入点一视同仁。一种自然的扩展方式是将每个点的不确定性作为额外的输入属性，并在训练过程中据此对样本重新加权。此外，正如近期的不确定性感知学习方法所指出的，同时预测不确定性和高度或可进一步增强模型的鲁棒性，并为下游应用提供可解释的置信度度量。

b) 可扩展的多传感器集成：本研究聚焦于以高度图形式表示的2.5D建筑表示。然而，与激光雷达相比，TomoSAR能够捕获更为详细的建筑立面信息，可用于完整的三维重建，而高度估计只是其更广泛应用潜力中的一个组成部分。结合其他地球观测数据源，对建成环境进行多尺度表示值得进一步探索。值得关注的方向包括将TomoSAR点云与激光雷达或摄影测量点云融合，以实现完整的三维建模；以及将光学影像与TomoSAR结合，用于对象级重建。同时，高质量的归一化数字表面模型数据并非总能用于监督。在这种情况下，弱监督或自监督策略或有助于维持性能。前者可以利用质量较低的数据或替代性高程数据源，后者则可以利用几何一致性等内在约束。最后，建筑轮廓监督只是用于规范训练的一种辅助线索，其他线索仍有待探索。综合而言，这些方向有望推动标签稀缺区域的学习，并提升方法在精细制图区域之外的可扩展性。

VI. 结论

我们提出了一种框架，采用双拓扑网络设计，在点和网格表示上从星载TomoSAR点云重建建筑物高度图。对两个质量各异的TomoSAR数据集进行的实验表明，我们的方法能够有效地对输入点进行去噪，并修补缺失数值，从而生成高保真度的高度图。此外，该框架易于扩展，可融入卫星光学影像，利用其提供的互补线索进一步提升重建质量。作为概念验证，我们的方法展现出推进大规模建筑高度制图的巨大潜力。

参考文献

[1] S. Chen, Y. Shi, Z. Xiong, and X. X. Zhu, “HTC-DC Net：基于单幅遥感影像的单目高度估计，” *IEEE*地球科学与遥感汇刊，第61卷，第1–18页，2023年。[2] Y. Cao and Q. Weng，“一种基于深度学习的超分辨率方法，用于在北半球以2.5米空间分辨率进行建筑高度估计，”《环境遥感》，第310卷，第114241页，2024年。[3] R. Yadav, A. Nascetti, and Y. Ban，“我们有多高？利用Sentinel-1合成孔径雷达和Sentinel-2多光谱成像仪时间序列，以10米分辨率进行大规模建筑高度估计，”《环境遥感》，第318卷，114556，2025年，第。[4] G. Fornaro, D. Reale, and F. Serafino，“用于单个和双重散射体高度估计与监测的四维SAR成像，”*IEEE*地球科学与遥感汇刊，第47卷，第1期，第224–237页，2009年1月。[5] X. X. Zhu and R. Bamler，“通过L1范数正则化进行层析SAR反演——压缩感知方法，”*IEEE*地球科学与遥感汇刊，第48卷，3846，2010年10月，第3839–页。[6] Y. Shi, Y. Wang, R. Bamler, and X. X. Zhu，“基于TanDEM-X数据构建高分辨率全球城市三维模型，”发表于*EARSeL*第五届联合研讨会“城市遥感——挑战与解决方案”，德国波鸿，2018年9月。[7] X. X. Zhu, Y. Sun, Y. Shi, Y. Wang, and N. Ge，“利用TanDEM-X数据实现全球三维/四维城市建模，”发表于*EUSAR 2018*，德国亚琛，2018年。[8] X. X. Zhu, S. Montazeri, C. Gisinger, R. F. Hanssen, and R. Bamler，“大地测量SAR层析成像，”*IEEE*地球科学与遥感汇刊，第54卷，第1期，第18–35页，2016年。[9] Z. Chen, Y. Shi, L. Nan, Z. Xiong, and X. X. Zhu，“PolyGNN：用于从点云进行三维建筑重建的基于多面体的图神经网络，”*ISPRS*摄影测量与遥感学报，第218卷，第693–706页，2024年。[10] C. Stucker, B. Ke, Y. Yue, S. Huang, I. Armeni, and K. Schindler，“ImpliCity：利用深度隐式占据场从卫星影像进行城市建模，”*ISPRS*摄影测量、遥感与空间信息科学年鉴，第V-2-2022卷，第193–201页，2022年。[11] L. Mou and X. X. Zhu，“IM2HEIGHT：通过全残差卷积-反卷积网络从单幅单目影像进行高度估计，”*arXiv:1802.10249 [cs]*, 2018年2月。[12] P. Ghamisi and N. Yokoya，“IMG2DSM：使用条件生成对抗网络从单幅影像进行高度模拟，”*IEEE*地球科学与遥感快报，第15卷，第5期，第794–798页，2018年5月。[13] M. E. Paoletti, J. M. Haut, P. Ghamisi, N. Yokoya, J. Plaza, and A. Plaza，“U-IMG2DSM：利用生成对抗网络进行数字表面模型的非配对模拟，”*IEEE*地球科学与遥感快报，第18卷，第7期，第1288–1292页，2020年。[14] S. Srivastava, M. Volpi, and D. Tuia，“利用卷积神经网络联合估计单目航拍影像的高度并进行语义标注，”发表于*2017*年*IEEE*国际地球科学与遥感研讨会（*IGARSS*）。IEEE，2017年，第5173–5176页。[15] Z. Xiong, S. Chen, Y. Shi, and X. X. Zhu，“用于可解释单目高度估计的解耦潜在变换器，”*arXiv*预印本 *arXiv:2201.06357*，2022年。[16] X. Li, Y. Zhou, P. Gong, K. C. Seto, and N. Clinton，“开发一种利用Sentinel-1数据估计建筑物高度的方法，”《环境遥感》，第240卷，第111705页，2020年。[17] H. Huang、P. Chen、X. Xu、C. Liu、J. Wang、C. Liu、N. Clinton 和 P. Gong，《基于 ALOS AW3D30 估算中国建筑物高度》，《ISPRS摄影测量与遥感学报》，第185卷，第146–157页，2022年。[18] W.-B. Wu、J. Ma、E. Banzhaf、M. E. Meadows、Z.-W. Yu、F.-X. Guo、D. Sengupta、X.-X. Cai 和 B. Zhao，《利用多源地球观测数据和机器学习首次估算中国10米分辨率建筑物高度（CNBH-10 m）》，《环境遥感》，第291卷，文章编号113578，2023年。[19] M. Recla 和 M. Schmitt，《SAR2Height框架：从单幅SAR强度图像重建城市高度图》，《ISPRS摄影测量与遥感学报》，第211卷，第104–120页，2024年。[20] ——，《基于深度学习的超高分辨率SAR强度数据单幅图像高度重建》，《ISPRS摄影测量与遥感学报》，第183卷，第496–509页，2022年。[21] Y. Sun、L. Mou、Y. Wang、S. Montazeri 和 X. X. Zhu，《基于边界框回归网络的单幅SAR影像大范围建筑物高度提取》，《ISPRS摄影测量与遥感学报》，第184卷，第79–95页，2022年。[22] G. Krieger、A. Moreira、H. Fiedler、I. Hajnsek、M. Werner、M. Younis 和 M. Zink，《TanDEM-X：用于高分辨率SAR干涉测量的卫星编队》，《IEEE地球科学与遥感汇刊》，第45卷，第11期，第3317–3341页，2007年。[23] G. Fornaro、F. Lombardini 和 F. Serafino，《三维多航过SAR聚焦：长期星载数据实验》，《IEEE地球科学与遥感汇刊》，第43卷，第4期，第702–714页，2005年。[24] X. X. Zhu 和 R. Bamler，《城市环境中的超高分辨率星载SAR层析成像》，《IEEE地球科学与遥感汇刊》，第48卷，第12期，第4296–4308页，2010年。[25] Y. Shi、R. Bamler、Y. Wang 和 X. X. Zhu，《逼近极限的SAR层析成像：仅使用3–5幅TanDEM-X双基地干涉图重建建筑物高度》，《IEEE地球科学与遥感汇刊》，第58卷，第8026–8037页，2020年。[26] K. Qian、Y. Wang、Y. Shi 和 X. X. Zhu，《γ-Net：通过深度学习实现SAR层析反演超分辨率》，《IEEE地球科学与遥感汇刊》，第60卷，第1–16页，2022年。[27] K. Qian、Y. Wang、P. Jung、Y. Shi 和 X. X. Zhu，《应用于SAR层析超分辨率的循环神经网络基追踪去噪》，《IEEE地球科学与遥感汇刊》，第1–1页，2022年。[28] ——，《HyperLISTA-ABT：一种用于精确多分量差分SAR层析反演的超轻量展开网络》，《IEEE地球科学与遥感汇刊》，第62卷，第1–15页，2024年。[29] Y. Wang、X. X. Zhu、B. Zeisl 和 M. Pollefeys，《融合米级分辨率四维干涉合成孔径雷达点云与光学图像，实现城市基础设施语义监测》，《IEEE地球科学与遥感汇刊》，第55卷，第1期，第14–26页，2017年1月。[30] Z. Jiao、C. Ding、X. Qiu、L. Zhou、L. Chen、D. Han 和 J. Guo，《利用机载TomoSAR进行城市三维成像：一种基于上下文信息的统计方法》，《ISPRS摄影测量与遥感学报》，第170卷，第127–141页，2020年。[31] W. Yang、S. Vitale、H. Aghababaei、G. Ferraioli、V. Pascazio 和 G. Schirinzi，《基于Pol-TomoSAR数据的森林区域高度估计深度学习方案》，《IEEE地球科学与遥感汇刊》，第61卷，第1–14页，2023年。[32] H. Armeshi、M. R. Sahebi 和 H. Aghababaei，《一种基于TomoSAR正则化的城市区域高度变化检测方法》，《国际应用地球观测与地理信息学期刊》，第129卷，文章编号103852，2024年。[33] M. Shahzad 和 X. X. Zhu，《利用星载TomoSAR点云大范围稳健重建建筑立面》，《IEEE地球科学与遥感汇刊》，第53卷，第2期，第752–769页，2014年。[34] ——，《从星载TomoSAR点云自动检测并重建二维/三维建筑形状》，《IEEE地球科学与遥感汇刊》，第54卷，第3期，第1292–1310页，2015年。[35] A. Ley、O. D’Hondt 和 O. Hellwich，《投影高度图域中TomoSAR点云的正则化与补全》，《IEEE地球观测与遥感精选主题期刊》，第11卷，第6期，第2104–2114页，2018年。[36] X. X. Zhu 和 R. Bamler，《压缩感知在谱估计中的超分辨率能力与稳健性及其在星载层析SAR中的应用》，《IEEE地球科学与遥感汇刊》，第50卷，第1期，第247–258页，2011年。[37] M. Eineder、C. Minet、P. Steigenberger、X. Cong 和 T. Fritz，《成像大地测量学——迈向TerraSAR的厘米级测距精度》，

X》，*IEEE*地球科学与遥感汇刊，第49卷，第2期，第661–671页，2011年。[38] S. Peng、M. Niemeyer、L. Mescheder、M. Pollefeys 和 A. Geiger，《卷积占用网络》，载于欧洲计算机视觉会议（*ECCV*），2020年。[39] Z. Wang、S. Zhou、J. J. Park、D. Paschalidou、S. You、G. Wetzstein、L. Guibas 和 A. Kadambi，《ALTO：用于隐式三维重建的交替潜在拓扑》，载于*IEEE/CVF*计算机视觉与模式识别会议论文集，2023年，第259–270页。[40] O. Ronneberger、P. Fischer 和 T. Brox，《U-Net：用于生物医学图像分割的卷积网络》，载于医学图像计算与计算机辅助干预国际会议。Springer，2015年，第234–241页。[41] C. R. Qi、H. Su、K. Mo 和 L. J. Guibas，《PointNet：用于三维分类与分割的点集深度学习》，载于*IEEE*计算机视觉与模式识别会议论文集，2017年，第652–660页。[42] Y. Shi、X. X. Zhu 和 R. Bamler，《基于非局部压缩感知的SAR层析成像》，*IEEE*地球科学与遥感汇刊，第57卷，第5期，第3015–3024页，2019年。[43] X. X. Zhu 和 R. Bamler，《让我们扭曲时间：差分SAR层析成像中的多分量非线性运动估计》，*IEEE*地球科学与遥感快报，第8卷，第4期，第735–739页，2011年。[44] C. R. Qi、L. Yi、H. Su 和 L. J. Guibas，《PointNet++：度量空间中点集的深层次层级特征学习》，神经信息处理系统进展，第30卷，2017年。[45] Planet团队，《Planet应用程序接口：在太空中服务地球生命》，*Planet*，2017年。