

# CoANet: 用于从卫星图像中提取道路的连通性注意力网络

Jie Mei, Rou-Jing Li, Wang Gao, Ming-Ming Cheng

**摘要**—从卫星图像中提取道路是一种有效、及时地更新道路网络动态变化的有前途的方法。然而，由于其他物体造成的遮挡和复杂的交通环境，基于像素的方法往往会产生碎片化的道路并且无法预测拓扑的正确性。在本文中，受图网络中道路形状和连接的启发，我们提出了一个连通性注意力网络 (CoANet) 来共同学习分割和成对依赖关系。由于条形卷积更符合道路大跨度、狭窄且连续分布的形状，我们开发了一个带状卷积模块 (SCM)，它利用四个带状卷积从不同方向捕获远程上下文信息并避免来自不相关区域的干扰。此外，考虑到由建筑物和树木引起的道路区域遮挡，提出了连通性注意力模块 (CoA) 来探索相邻像素之间的关系。CoA 模块结合了图信息，可以更好地保留道路的连通性。在流行的基准测试 (SpaceNet 和 DeepGlobe 数据集) 上进行的大量实验表明，我们提出的 CoANet 建立了新的最先进的结果。源代码将在以下位置公开提供：<https://mmcheng.net/coanet/>。

**Index Terms**—道路提取、卫星图像、连通性注意力、条带卷积、拓扑连接。

## I. 介绍

在自动驾驶、城市规划、车辆导航和地理信息更新等众多应用领域中，创建路线图是一个基本且必不可少的步骤。一些地图公司采用的现有地图采集方法通常很耗时，例如从 LIDAR 点云中提取、GPS 轨迹聚合或手动道路标记。这些方法不适用于大面积区域，不足以在快速变化的环境中更新路网的动态变化 [1], [2]。卫星影像不仅可以代表道路的几何特征，还可以提供多时期甚至实时的影像。为了加快更新过程，从卫星图像中提取道路网络 [3]–[6] 已成为一种很有前景的方法。

传统研究侧重于利用手工设计的特征并定义某些标准来从卫星图像中提取道路的算法 [7]–[11]。在处理大区域的

J. Mei and M.-M. Cheng are with the TKLNDST, College of Computer Science, Nankai University, Tianjin, 300350, China. (E-mail: meijie@mail.nankai.edu.cn, cmm@nankai.edu.cn)

R.-J. Li is with the State Key Laboratory of Earth Surface Processes and Resource Ecology, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China. (E-mail: lirj@mail.bnu.edu.cn)

W. Gao is with Science and Technology on Complex System Control and Intelligent Agent Cooperation Laboratory, 100191, Beijing, China

M.-M. Cheng is the corresponding author.

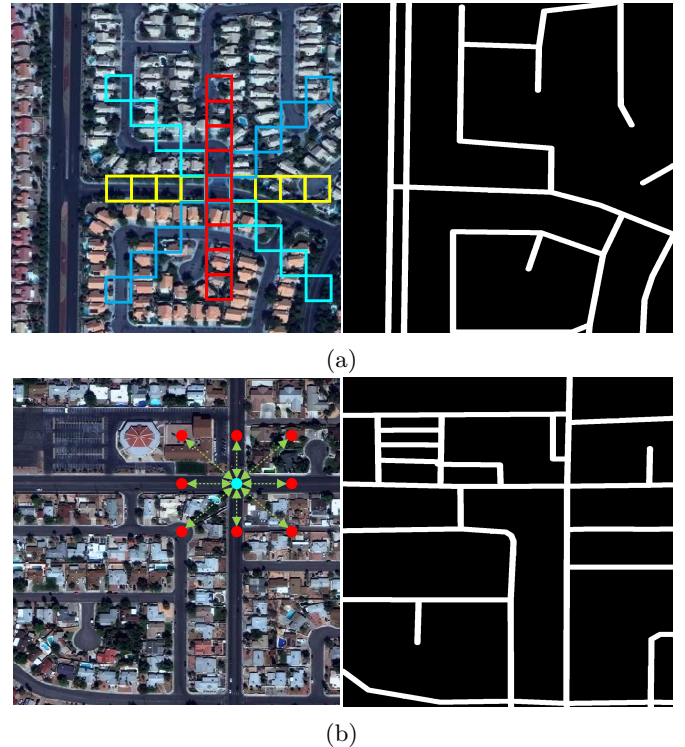


图 1: 我们提出条带卷积模块和连通性注意力模块的动机。从左到右: 卫星图像、地面真值。(a) 用于捕获道路的线性特征的具有四种形状的带状卷积。(b) 预测一个像素与相邻像素的连通性以捕获局部成对依赖关系并确保道路拓扑的正确性。

卫星图像时，这些方法通常效率低下。随着深度学习的发展，卷积神经网络 (CNN)，特别是具有全卷积网络 (FCN) [12] 架构的网络被提出并被证明在图像语义分割方面是有效的 [13]–[18]。有几项工作将具有编码器-解码器架构的 CNN 应用于道路分割任务 [19]–[22]，通常会获得良好的分割结果。然而，从卫星图像中提取道路仍然具有挑战性，因为：(a) 建筑物和树木的遮挡，(b) 复杂的城市环境和交通，(c) 道路与其他物体之间的相似性。这些困难导致道路分割的碎片化，上述方法不能保证道路的连通性。最近，[1], [23]–[25] 介绍了使用分割和后处理步骤来细化缺失连接的方法，其中最短路径算法通常用作后处理，它不能应用于

复杂的道路环境。为了直接获得连通性较好的道路，[5]A 采用迭代搜索过程自动提取道路网络，[26] 利用 U-Net 结合多种损失函数对道路轮廓进行迭代细化。此外，Liu 等人 [27] 集成了包括路面、边缘和中心线在内的多级特征，以改进道路预测。然而，这些方法非常耗时，而且通常需要复杂的步骤来训练。

在本文中，我们提出了一种用于从卫星图像中提取道路的连通性注意力网络 (CoANet)。我们首先引入编码器-解码器架构网络来学习道路特征，其中采用空洞空间金字塔池化模块 (ASPP) 来增加特征点的感受野并捕获多尺度特征。由于道路跨度大、狭窄且连续分布，因此带状卷积更符合道路形状。我们利用它并开发了一个带状卷积模块 (SCM)，它被放置在解码器网络中。如图 Fig. 1 (a) 所示，SCM 利用水平、垂直、左对角线和右对角线的四个带状卷积从四个不同方向捕获远程上下文信息。此外，它可以防止不相关的区域干扰特征学习。为了减轻由建筑物和树木引起的道路区域的遮挡，我们提出了一个连通性注意力模块 (CoA) 来探索相邻像素之间的关系。如图 Fig. 1 (b) 所示，预测给定像素与八个相邻像素的连通性，从而实现道路的拓扑正确性。在基于像素和基于图的指标方面对流行基准进行的大量实验证明了我们的 CoANet 与几种最先进的方法相比的优越性。

我们的贡献总结如下。

- 我们提出了一个连通性注意力网络，它联合学习相邻像素之间的分割和关系以提高道路的连通性，在广泛使用的道路数据集上取得了比其他方法的显著改进。
- 我们开发了一个带状卷积模块，它利用四个不同方向的带状卷积来捕获远程上下文信息并避免来自不相关区域的干扰。
- 我们设计了一个连通性注意力模块，它通过利用成对相邻像素之间的依赖关系并结合图形信息来提高道路连通性。

本文的其余部分组织如下。Sec. II 总结了道路提取和多任务学习的相关工作。在 Sec. III 中，我们介绍了我们提出的 CoANet 的细节。在 Sec. IV 中，提供了数据集、评估指标和实现细节，进行了大量实验以评估我们从卫星图像中提取道路的方法的性能。结论和讨论在 Sec. V 中。

## II. 相关工作

### A. 道路分割

许多研究已经尝试从卫星图像中提取道路网络。传统的道路提取方法通常利用手工设计的特征并定义某些标准来匹配 [28]–[30]。Hei 等人 [11] 提出了一种基于颜色的道路检测算法，它结合了灰度图像的边缘估计和彩色图像的

道路区域提取结果。Zhang 等人 [31] 引入了许多角度纹理的描述符，并使用模糊逻辑分类器识别路段。Laptev 等人 [8] 基于多尺度道路检测进行道路提取，结合使用蛇形几何约束进行边缘提取。[10] 使用 Gibbs 点处理框架从遥感图像中提取道路。并且 [9] 开发了使用连接点过程以恢复航空和视网膜图像中的线网络。Wegner 等人 [32] 提出了一种用于道路网络提取的高阶条件随机场 (CRF) 模型。然而，这些方法通常效率低下，不适合大面积区域。

随着深度学习的发展，具有编码器-解码器架构的 CNN [12], [14], [16], [33]–[37] 被提出并被证明在语义分割方面是有效的。一些研究 [20], [22], [38]–[40] 使用基于 CNN 的模型将道路提取制定为分割问题。Mnih 等人 [19] 通过使用在图形处理上实现的神经网络来检测道路。Cheng 等人 [20] 提出了一个级联的端到端 CNN (CasNet) 来同时处理来自超高分辨率 (VHR) 遥感图像的道路分割和中心线提取任务。Panboonyuen 等人 [21] 提出了道路分割的 DCNN 框架，然后提出了景观度量来减少错误分类的道路像素，并采用 CRF 来锐化提取的道路。U-Net [13] 和 LinkNet [15] 是备受推崇的用于语义分割的编码器-解码器结构，它们的变体也被提出用于学习细长的道路特征。Zhang 等人 [22] 提出一种语义分割神经网络 ResUnet，结合残差学习和 U-Net 进行道路区域提取。在 CVPR DeepGlobe 2018 道路提取挑战赛中，Zhou 等人 [39] 提出了一个 D-LinkNet，它采用 LinkNet 结构构建，并在中心部分添加了扩张卷积层。这些方法通常都能获得较好的道路分割结果，但不能保证道路的连通性。

### B. 道路连通性

道路连通性是最重要的道路特征之一，是车辆导航、自动驾驶和路线规划所必需的。最近，研究人员越来越关注这一特性，并提出了几种方法。Wegner 等人 [23] 首先将航拍图像分割成超像素，并使用最短路径算法连接道路似然性高的候选路径。Máttyus 等人 [1] 使用具有编码器-解码器结构的模型获得航拍图像的初始分割。由于分割结果无法预测连接的道路，因此他们通过使用最短路径算法推理缺失的连接来引入后处理步骤。在这些方法中，道路连通性是通过后处理实现的，不适用于有遮挡、道路外观模糊和道路密度高的区域等复杂环境。

为了直接获得连通性更好的道路提取结果，Mosinska 等人 [26] 利用 U-Net 结合像素级损失和拓扑感知损失对道路轮廓进行迭代细化。Bastani 等人 [5] 提出了 Road-Tracer，它使用由基于 CNN 的决策函数引导的迭代搜索过程，从航拍图像中自动提取道路网络。Batra 等人 [6] 提出了一个堆叠的多分支模块，以有效利用分割和方向学习任务之间的互信息。他们还开发了一种连通性细化方法，

以迭代地细化预测道路网络的拓扑结构。然而，迭代步骤很耗时，而且这些方法通常需要很长时间来训练。

有一些研究将多级道路特征或其他地理数据进行整合以获得连通道路提取结果。RoadNet 由 Liu 等人 [27] 提出，用于同时预测道路表面、边缘和中心线，其中集成了多级特征以处理各种场景中的道路并改进道路预测。Li 等人 [41] 开发了一种新颖的框架来有效地整合道路形状特征，包括点、边缘和区域特征。然后引入方向感知注意力模块以进一步提高道路连通性和道路识别精度。光探测和测距 (LiDAR) 数据 [42]–[44] 和 GPS 轨迹 [45]–[49] 也已用于推断路线图。[50] 将众包 GPS 数据与航拍图像相结合以提取道路，实验表明其结果优于单独使用 GPS 数据或图像的模式。LiDAR 和 GPS 数据可能有助于改善道路连通性，尤其是在有遮挡的区域。然而，收集足够覆盖大区域的 LiDAR 和 GPS 数据具有挑战性，并且这些数据的预处理通常很复杂。

### C. 成对依赖

在语义分割的任务中，一些研究设计了聚合上下文信息以提高性能。Shen 等人 [51] 提出了一个联合目标来整合分割特征、高阶上下文和边界引导，其中采用引导 CRF 来进一步提高分割性能。Dai 等人 [52] 引入了可变形卷积和可变形 RoI 池化来增强 CNN 的可变换建模能力，这是一种对密集空间变换进行建模的简单有效的方法。为了捕获远程依赖关系，[53] 提出了基于不同位置之间的关系计算上下文响应的非本地块，但它需要很高的计算成本。Criss-cross network [54] 为了获取交叉路径上所有像素的上下文信息而开发，效率更高。Wang 等人 [55] 提出了一种用于语义分割的逐像素对比方法，该方法通过利用不同图像中像素之间的全局上下文来学习结构良好的像素语义嵌入空间。

我们的连通性注意力模块的想法也与在语义分割中学习像素亲和度的方法有关。Bertasius 等人 [56] 引入了卷积随机游走网络来集成语义分割和像素级亲和力，解决了边界定位差和空间碎片化的问题。Cheng 等人 [57] 设计了一个局部敏感的 DeconvNet，其中采用亲和矩阵来学习相邻像素之间的关系。AffinityNet 在 [58] 中被提出来预测相邻图像坐标对之间的高级语义亲和力。Fan 等人 [59] 开发了一个跨图像亲和模块来探索像素级的跨图像关系。上述方法旨在恢复对象形状或细化语义分割模型的输出。我们的目标是通过利用相邻像素之间的成对亲和力来改善道路的连通性。

在卫星图像中标注道路时，人类需要识别像素是否属于道路并考虑正确道路拓扑的重要性连接道路像素。受语义分割中的上下文聚合和像素亲和力学习的启发，我们提

出了一连通性注意力模块来捕获相邻像素之间的成对依赖关系。该模块能够改善道路的连通性，这在我们基于像素和基于图形的指标的实验中得到了展示。

## III. 方法

道路连通性是一个重要的道路特征，而基于分割的方法通常会产生支离破碎的道路。为了缓解这个问题，我们开发了一个连通性注意力网络 (CoANet)，用于从卫星图像中提取道路，如图 Fig. 2 所示。在 CoANet 中，提出了一个带状卷积模块 (SCM) 来对齐道路的形状并提取其线性特征。我们进一步提出了一个连通性注意力模块 (CoA) 来预测相邻像素之间的道路连接。

### A. 网络结构

在 CoANet 中，我们使用在 ImageNet [60] 上预训练的 ResNet-101 [61] 作为编码器，因为它在特征学习方面表现出色。由于空洞卷积是控制过滤器视野和调整特征图分辨率的强大工具，如 [16]，我们将扩张率  $r = 2$  和  $r = 4$  的空洞卷积应用于 ResNet-101 中的最后两个卷积块用于更密集的特征提取。

为了在多个尺度上有效地学习特征，采用了 [18] 中的空洞空间金字塔池化模块 (ASPP)。由于道路狭窄、复杂、跨度大，ASPP 的使用将增加特征点的感受野，提高道路的连通性。解码器模块包含四个带状卷积块，用于将特征图上采样到适当的大小并提取道路的线性特征。每个条带卷积块包含四个不同方向的条带卷积，用于捕获长距离上下文信息，包括水平、垂直、左对角线和右对角线。此外，条带卷积块的每个输出特征图都经过  $1 \times 1$  卷积调整，然后在编码器中与卷积块的相应特征图连接。

**损失函数。** 解码器模块之后有两个分支：分割分支和连通性分支。连通性分支对应于我们开发的连通性注意力模块，其中预测给定像素与八个相邻像素的连通性以结合图形信息并保证道路的拓扑正确性。对于分割分支，它包括一个  $3 \times 3$  的卷积和一个用于将通道数量减少到一个的  $1 \times 1$  卷积。分割分支的损失函数定义为：

$$L_{seg} = L_{BCE} + \alpha(1 - L_{Dice}), \quad (1)$$

其中  $L_{BCE}$  是二元交叉熵， $L_{Dice}$  是骰子系数，定义为：

$$L_{BCE} = -\frac{1}{N} \sum_{i=1}^N [y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i)], \quad (2)$$

$$L_{Dice} = \frac{2 \sum_{i=1}^N (y_i \hat{y}_i)}{\sum_{i=1}^N y_i^2 + \sum_{i=1}^N \hat{y}_i^2}, \quad (3)$$

其中  $\alpha$  是一个常数。 $N$  表示  $H \times W$  切片中元素的数量， $y_i$  是表示位置  $i$  中给定像素的道路或背景的地面真值， $\hat{y}_i$  是对应分割分支的预测概率。

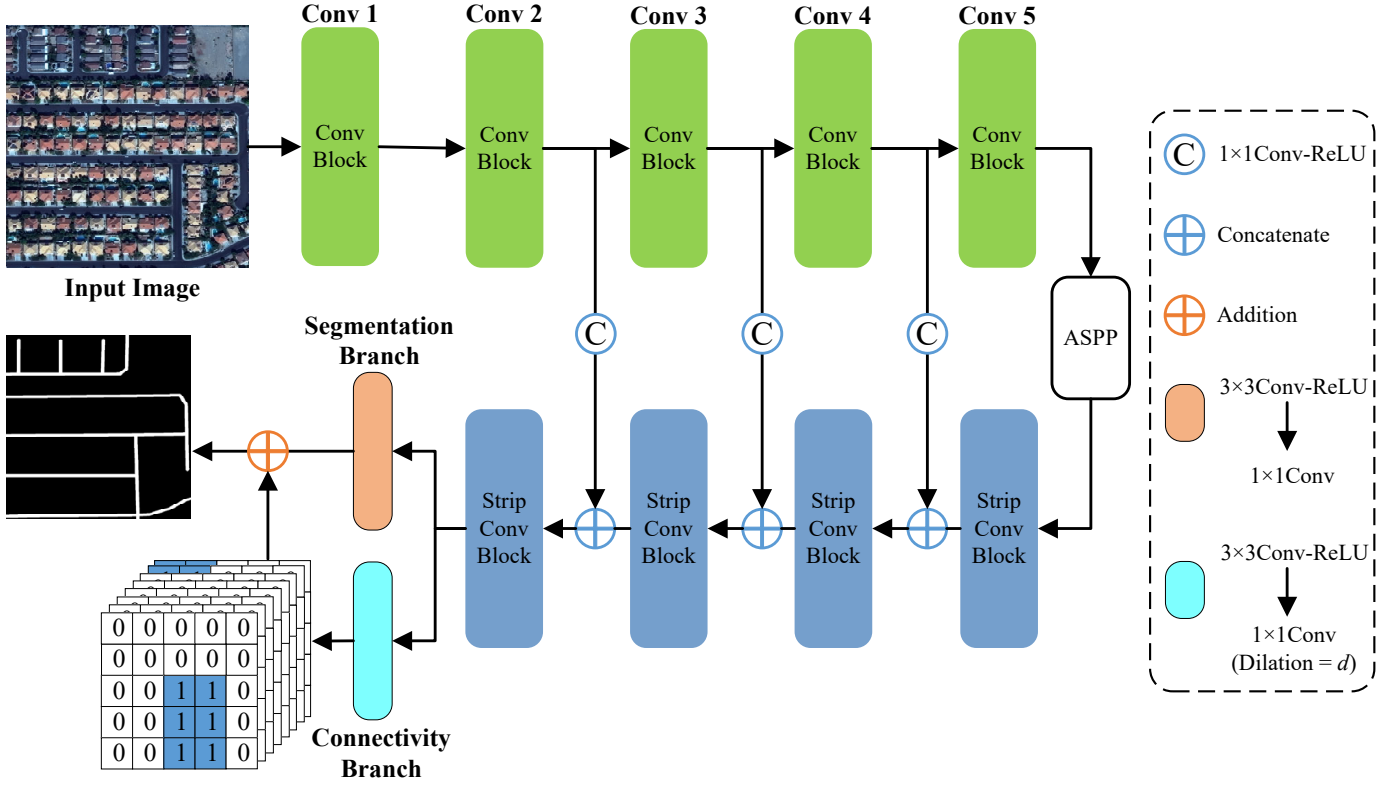


图 2: 所提出的连接注意网络 (CoANet) 的整体架构。编码器模块包含五个卷积块。ASPP 是空洞空间金字塔池化模块, 它通过应用多尺度的空洞卷积来学习多尺度特征。并且解码器模块包括四个条状卷积块。\$d\$ 表示给定像素与其相邻像素的间隔。

### B. 条带卷积模块

大多数 CNN 架构中的卷积通常具有方形内核并在方形窗口内学习特征图, 这适用于大多数具有块状形状的自然物体。然而, 道路跨度大、狭窄, 且连续分布。利用平方卷积不能很好地捕捉道路的线性特征, 并且不可避免地会包含来自相邻像素的无关信息。带状卷积更符合道路的形状, 它利用沿一个空间方向的长核形状来捕获道路区域中的远程依赖性。此外, 它沿另一个空间方向捕获局部上下文, 并防止不相关的区域干扰特征学习。

受上述事实的启发, 我们提出了一种新颖的条状卷积模块 (SCM)。如图 Fig. 3 所示, SCM 利用水平、垂直、左对角线和右对角线的四个带状卷积从四个不同方向捕获远程上下文信息。记  $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$  为 SCM 的输入张量, 其中  $H$ ,  $W$ , 和  $C$  表示高, 宽和通道的数量。在条状卷积模块中,  $\mathbf{X}$  在经过一个  $1 \times 1$  卷积后被送入四个平行路径, 每个都包含一个形状的带状卷积。然后将四个带状卷积的输出特征图连接起来, 然后进行上采样操作和一个  $1 \times 1$  卷积获得带状卷积块的输出。

令  $\mathbf{w} \in \mathbb{R}^{2k+1}$  为大小为  $2k+1$  的带状卷积滤波器,  $\mathbf{D} = (D_h, D_w)$  是滤波器  $\mathbf{w}$  的方向,  $\mathbf{Z}_D \in \mathbb{R}^{H \times W \times C'}$  表

示条带卷积的结果。带状卷积可以定义为:

$$\begin{aligned} \mathbf{Z}_D[i, j] &= (\mathbf{X} * \mathbf{w})_D[i, j] \\ &= \sum_{l=-k}^k x[i + D_h l, j + D_w l] \cdot w[k - l], \end{aligned} \quad (4)$$

其中  $\mathbf{X} * \mathbf{w}$  表示卷积操作。 $\mathbf{D}$  是条状卷积的方向向量, 取值  $(0, 1)$ ,  $(1, 0)$ ,  $(1, 1)$ , 和  $(-1, 1)$ , 为分别用于水平、垂直、左对角线和右对角线的卷积。对于滤波器  $\mathbf{w}$ , 我们设置  $k = 4$  使每个条状卷积有 9 个参数, 与  $3 \times 3$  卷积核相同。

在上面的 SCM 中, 允许输出特征图中的每个位置与输入特征图中四个方向的多个位置建立关系。我们选择的四个方向与卫星图像中大部分道路的分布是一致的, 并且相对容易实现。

### C. 连通性注意力模块

由于建筑物和树木造成的遮挡会干扰道路连通性, 因此从卫星图像中提取道路具有挑战性。为了缓解这个问题, 我们开发了一个连通性注意力模块 (CoA) 来有效预测相邻像素之间的道路连通性并摆脱背景区域。CoA 能够探索像素对之间的关系, 这与特征学习过程无缝结合。该模块

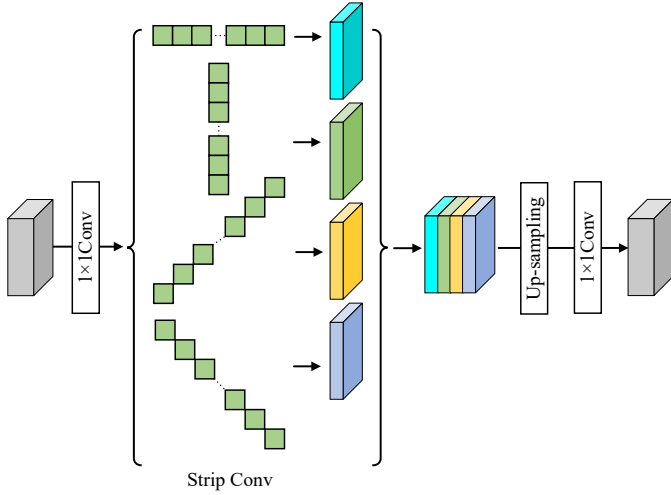


图 3: 条状卷积块。条状卷积包含四种不同的形状: 水平、垂直、左对角线和右对角线。

使我们的模型能够整合通常在图形模型中学到的信息, 并导致更好的道路连通性。

利用二元地面实况掩码, 我们首先生成地面实况连通性立方体  $O \in \mathbb{R}^{H \times W \times C_o}$ , 其中  $C_o$  表示给定像素的采样相邻像素的数量, 我们设置  $C_o = 8$ 。

在连通性立方体中,  $O_{i,j,c}$  表示一个像素与特定位置的相邻像素的连通性, 其中  $i, j$  表示像素的空间位置,  $c$  表示其相邻像素的位置。  $O_{i,j,c} = 1$  如果两个像素相连, 则表示它们都是道路像素。并且不连接背景像素以减少不相关的噪声。对于相邻像素, 我们从给定像素中选择间隔为  $d = 1$  的像素。如图 Fig. 4 (a) 所示,  $C1 - C8$  位置上的像素被选为相邻像素。通过检查每个像素与其在特定位置的相邻像素是否连接并连接位置  $C1 - C8$  的连接掩码, 我们可以获得连通性立方体  $O$  的地面真值。

在我们的 CoA 模块中, 如图 Fig. 4 (b) 所示, 输入张量喂进一个  $3 \times 3$  卷积, 其次是一个扩张率  $r = d$  的  $3 \times 3$  空洞卷积。空洞卷积用于增加感受野并学习相邻像素之间的关系。然后采用 [62] 中的挤压激励 (SE) 块来充分利用连通性, 通过使用通道注意力机制重新校准预测的连通性立方体。SE 块包含两个全连接层和一个 sigmoid 函数。输入特征图在全局平均池化后被送入这个块, 我们可以获得一个范围为  $(0, 1)$  的向量, 其中每个因子乘以输入特征图中的相应通道。CoA 模块最终输出是一个  $H \times W \times C_o$  的连通性立方体, 用于预测相邻像素之间的连通性。

我们提出的 CoANet 中的连接分支由两个连接注意模块组成, 其中一个在上述  $d = 1$  的模块, 另一个是  $d = 3$  的模块。对于  $d = 3$  的 CoA 模块, 给定像素与其相邻像素的间隔设置为 3, CoA 模块中的  $3 \times 3$  空洞卷积率设置

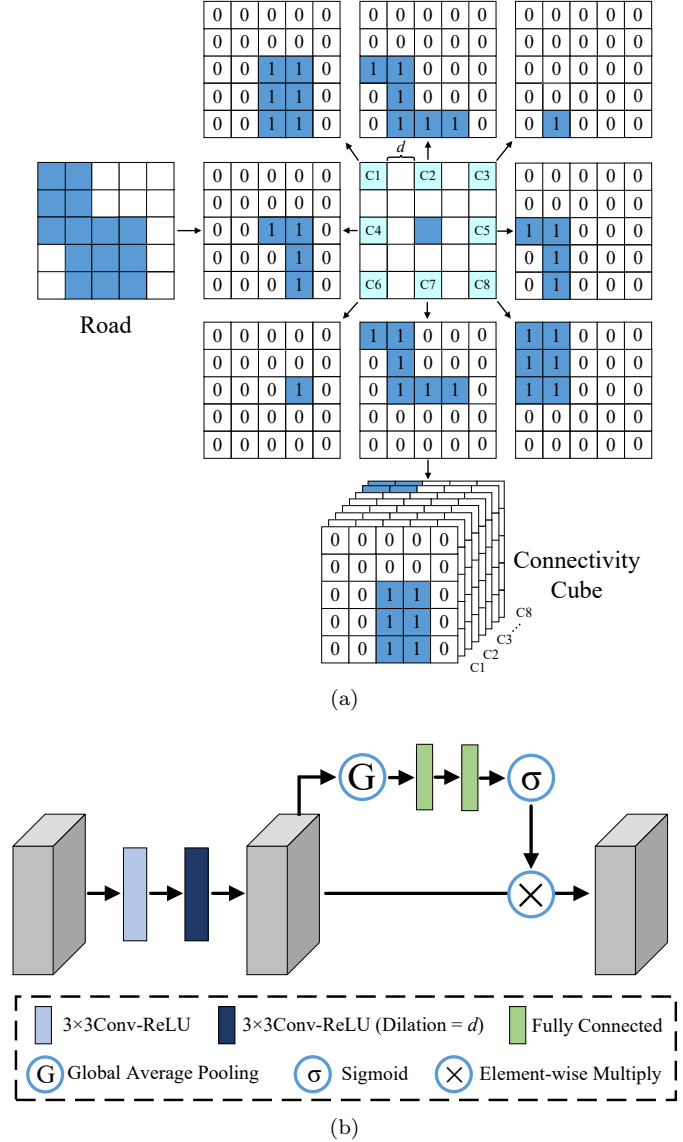


图 4: (a) 连通性立方体是如何生成的图示, 采样像素之间的间隔  $d = 1$ 。在道路图像中, 白色像素表示背景, 蓝色像素表示道路。在连通性立方体中, 所有像素都有二进制值, 其中 1 表示像素在一个方向上与相邻像素相连, 0 表示未连接的像素。(b) 连通性注意力模块 (CoA)。

为  $r = 3$ 。采用不同设置的两个 CoA 模块来捕获多尺度连通性信息, 提高预测道路的连通性。我们将在实验中对我们的方法在 CoA 模块的不同配置下的性能提供更多分析。

连通性分支的损失函数定义为:

$$L_{con} = L_{d1} + \beta L_{d3}, \quad (5)$$

$$L_{d1} = -\frac{1}{C_o \times N} \sum_{c=1}^{C_o} \sum_{i=1}^N [y_i^c \cdot \log(\hat{y}_i^c) + (1 - y_i^c) \cdot \log(1 - \hat{y}_i^c)], \quad (6)$$

其中  $\beta$  是个常数。  $C_o$  表示给定像素的采样相邻像素的数量,  $N$  是  $H \times W$  切片中元素的数量。  $y_i^c$  是表示位置  $i$  中的给定像素与其在位置  $c$  中的相邻像素的连通性或非连通性的真值,  $\hat{y}_i^c$  是连通分支对应的预测连通性。 损失函数  $L_{d3}$  与  $L_{d1}$  相同。

整体损失函数可以定义为:

$$L_{CoANet} = L_{seg} + \lambda L_{con}, \quad (7)$$

其中  $\lambda$  是常数。

#### IV. 实验

在本节中, 我们将介绍实验中使用的两个数据集和评估指标。 此外, 还提供了定量和定性的详细评估结果。

##### A. 数据集

我们的实验中使用了两个数据集来评估所提出方法的性能。

**SpaceNet [63]:** 该数据集提供了来自巴黎、拉斯维加斯、上海和喀土穆四个不同城市的 30 厘米/像素分辨率为  $1300 \times 1300$  的图像。 道路的注释以线串的形式提供, 表示道路的中心线。 数据集由 2,780 张图像组成, 分为 2,213 张用于训练的图像和 567 张用于测试的图像, 遵循 [64]。 我们通过创建  $650 \times 650$  的裁剪图像来扩充训练数据集。

**DeepGlobe [65]:** 该数据集包括 50 厘米/像素的图像, 像素分辨率为  $1024 \times 1024$ 。 这些图像来自三个不同的地区: 泰国、印度尼西亚和印度。 它提供像素级注释, 包括道路和背景类。 该数据集包含 6,226 张图像, 遵循 [64], 我们将其分成 4,696 张训练集和 1530 张测试集。 通过创建  $512 \times 512$  的裁剪图像来扩充训练数据集。

##### B. 评估指标

**基于像素的指标。** 为了评估我们的道路分割方法的性能, 我们使用 F1 得分和联合交集 ( $IoU$ ) 度量。 由于 SpaceNet 数据集的注释是以线串的形式提供的, 我们对等宽线串进行光栅化来获得道路分割的真实情况。 在我们的实验中, 道路中心线的缓冲区设置为 3 米 (10 个像素)。

**基于图形的指标。** 我们的实验中使用平均路径长度相似性 (APLS) [63] 来评估道路的拓扑正确性和连通性。 ALPS 度量测量真实图  $G$  和建议图  $\hat{G}$  中所有节点对之间光路长度的差异, 其定义为:

$$APLS = 1 - \frac{1}{n} \sum \min \left\{ 1, \frac{|L(a,b) - L(\hat{a},\hat{b})|}{L(a,b)} \right\} \quad (8)$$

表 I: 我们提出的 CoANet 与 SpaceNet 数据集上一些最先进的道路提取方法的定量比较 (%)。

	F1	IoU	APLS
DeepRoadMapper [1] <i>ICCV17</i>	71.47	55.61	46.76
Topology Loss [26] <i>CVPR18</i>	58.44	41.29	39.08
LinkNet34 [15] <i>VCI17</i>	73.96	58.68	63.12
D-LinkNet [39] <i>CVPRW18</i>	69.77	53.57	50.20
RoadCNN [5] <i>CVPR18</i>	73.74	58.40	59.39
ImprovedConnectivity [6] <i>CVPR19</i>	75.91	61.17	62.81
VecRoad [68] <i>CVPR20</i>	63.63	46.65	61.64
CoANet	<b>76.91</b>	<b>62.48</b>	<b>65.53</b>

其中  $\hat{a}, \hat{b}$  分别是预测图  $\hat{G}$  中最接近真实图  $G$  中节点  $a, b$  位置的节点。  $L(\hat{a}, \hat{b})$  and  $L(a, b)$  分别表示图  $\hat{G}$  和  $G$  中对应节点之间的路径长度。  $n$  是唯一路径的数量。

##### C. 实现细节

在我们的 CoANet 中, 使用了批量大小为 16 的随机梯度下降 (SGD) 优化器。 动量和权重衰减系数分别设置为 0.9 和  $5 \times 10^{-4}$ 。 学习率最初设置为 0.01, 我们采用 “poly” 策略逐渐降低学习率, 其中学习率乘以  $(1 - \frac{iter}{maxiter})^{power}$ ,  $power = 3$ 。 我们的方法是使用机器学习框架 PyTorch [66] 执行的, 实验是在 4 个 NVIDIA RTX TITAN GPU 上实现的, 内存为 24GB。 PyTorch [66] 和 Jittor [67] 版本的源代码都将公开发布。 在训练过程中, 应用随机旋转、水平翻转、重新缩放和高斯模糊等数据增强来提高模型的泛化能力。 最后, 将两个数据集的图像裁剪为固定大小  $512 \times 512$ 。

在推理阶段, 我们利用连通性分支的预测来增强道路提取的结果。 连接注意力模块的输出中有八个通道, 其中每个通道的预测可以看作是分割任务的一个子问题。 例如, 给定预测的连通性立方体  $O$ , 如果  $\sigma(O_{i,j,c}) > t$ ,  $(i, j)$  位置处的像素于位置  $c$  处的邻居像素是相连的, 并且它们两个都是道路像素。  $\sigma()$  是 sigmoid 非线性函数,  $t$  是阈值。 我们估计  $O$  的每个通道, 并沿通道维度求和, 就可以得到一维的道路掩码。 将其加入到分割分支的输出中, 得到最终的道路提取结果。

##### D. 与最先进方法的比较

在本节中, 我们将 CoANet 的性能与 SpaceNet 和 DeepGlobe 数据集上的几种最先进的道路提取方法进行比较, 包括 DeepRoadMapper [1], Topology Loss [26], LinkNet34 [15], D-LinkNet [39], RoadCNN [5], ImprovedConnectivity [6], 和 VecRoad [68]。 在这些方法中, Deep-

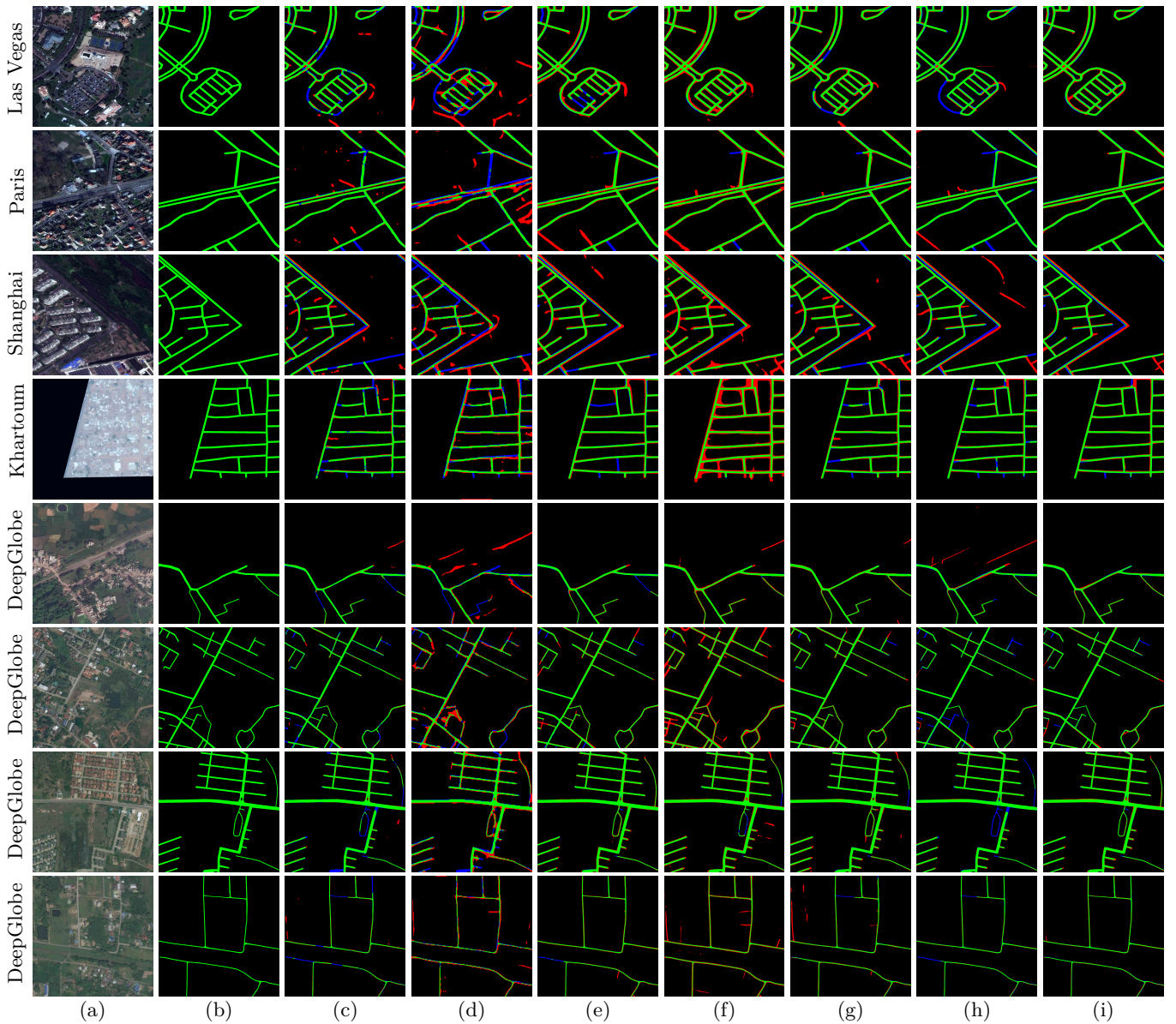


图 5: 我们的 CoANet 与其他最先进方法的定性比较。绿色: 真阳性, 红色: 假阳性, 蓝色: 假阴性。第一行到第四行分别是拉斯维加斯、巴黎、上海、喀土穆等 SpaceNet [63] 中不同城市的对比结果。The fifth to the eighth rows show the comparison results of DeepGlobe [65]. 第五行到第八行是 DeepGlobe [65] 的对比结果。(a) 卫星图像。(b) 真值图。(c)-(i) DeepRoadMapper 道路提取结果 [1], 拓扑损失 [26], LinkNet34 [15], D-LinkNet [39], RoadCNN [5], ImprovedConnectivity [6], 和我们的 CoANet.

RoadMapper [1] 和 ImprovedConnectivity [6] 利用后处理步骤提高道路的连通性, 而其他方法直接生成提取结果。

**SpaceNet 数据集上的实验。** SpaceNet 数据集上的定量实验结果如表 1 所示。值得注意的是, 我们的 CoANet 在基于像素和基于图的指标方面优于其他方法。例如, CoANet 取得了 76.91% 的 F1 得分和 62.48% 的 IoU 得分, 分别比 ImprovedConnectivity 提高了 1.00% 和 1.31%。对于用于评估道路拓扑正确性的 APLS, 我们的方法将次优方法 LinkNet34 [15] 提高了 2.4%。由于我们

提出的 CoANet 将相邻像素之间的分割和关系集成到一个框架中, 因此它可以从卫星图像中更准确地提取道路。更重要的是, 使用连通性注意力模块可以进一步提高道路的拓扑连通性。我们还比较了一种基于图的方法 VecRoad, 它引入了具有灵活步骤的迭代图探索模型。我们的方法在 IoU 分数上获得了 15.83% 的改进, 在 APLS 分数上获得了 3.89% 的改进。基于图的方法通常保证提取道路的连通性, 但可能存在大量缺失的道路。因此, 我们的 CoANet 也比基于图的方法具有优势。

表 II: 我们提出的 CoANet 与 DeepGlobe 数据集上一些最先进的道路提取方法的定量比较 (%)。

	F1	IoU	APLS
DeepRoadMapper [1] <i>ICCV17</i>	78.04	63.98	58.85
Topology Loss [26] <i>CVPR18</i>	56.07	38.95	46.99
LinkNet34 [15] <i>VCIP17</i>	79.65	66.18	72.93
D-LinkNet [39] <i>CVPRW18</i>	77.49	63.26	71.81
RoadCNN [5] <i>CVPR18</i>	79.08	65.40	71.15
ImprovedConnectivity [6] <i>CVPR19</i>	79.93	66.58	71.69
CoANet	<b>81.22</b>	<b>68.37</b>	<b>73.48</b>

SpaceNet [63] 是一个遥感影像主要采集于市区的数据集，包含了城市中的多种道路类型，如高速公路、住宅区和高速公路等。还有建筑物、建筑物的阴影和树木的遮挡。我们的方法在基于像素和基于图的指标上取得了最好的结果，这表明 CoANet 能够处理复杂的城市交通环境并提取具有更好连通性的道路。

**DeepGlobe 数据集上的实验。**表 Table II 显示了我们提出的 CoANet 与之前在 DeepGlobe 数据集上的最先进方法相比的定量结果。由于 VecRoad [68] 需要道路线串的 ground truth，而 DeepGlobe 只有像素级的标注，所以不展示 VecRoad 的实验结果。我们的 CoANet 实现了 68.37% 的 IoU 分数，优于其他方法，并将第二好的方法 ImprovedConnectivity [6] 提高了 1.79%。此外，CoANet 还获得了最佳 APLS 分数，其性能优于 LinkNet34 0.55% 和 D-LinkNet 1.67%。

DeepGlobe [65] 中的遥感图像主要来自农村地区。它包含大量的乡村道路，道路的宽度在不断变化，这意味着一条道路可能有不同的宽度。并且有树木和树影造成的严重遮挡。实验结果表明，我们的 CoANet 对农村地区的道路也有效。结合在 SpaceNet [63] 数据集上的结果，表明我们提出的方法对不同地区的不同道路类型具有鲁棒性。

**定性比较。**我们的 CoANet 和其他方法的定性比较结果如图 Fig. 5所示，其中展示来自 SpaceNet [63] 不同城市的四个示例和 DeepGlobe [65] 的四个示例 需要注意的是，我们的方法提取的道路与地面实况的道路一致，并且存在很少的假阳性像素。在一些存在遮挡的区域，其他方法提取的道路可能会断开，而我们的 CoANet 很好地保持了连通性。例如，在拉斯维加斯的结果(图 Fig. 5中的第一行)中，图像的左下角有一个停车场，那里停着很多车辆并种了几棵树。通过 DeepRoadMapper [1]，拓扑损失 [26]，LinkNet34 [15] 和 ImprovedConnectivity [6] 提取的道路有很多缺陷，未能保持连接，但是通过 CoANet 得到的结果与地面真值相一致。

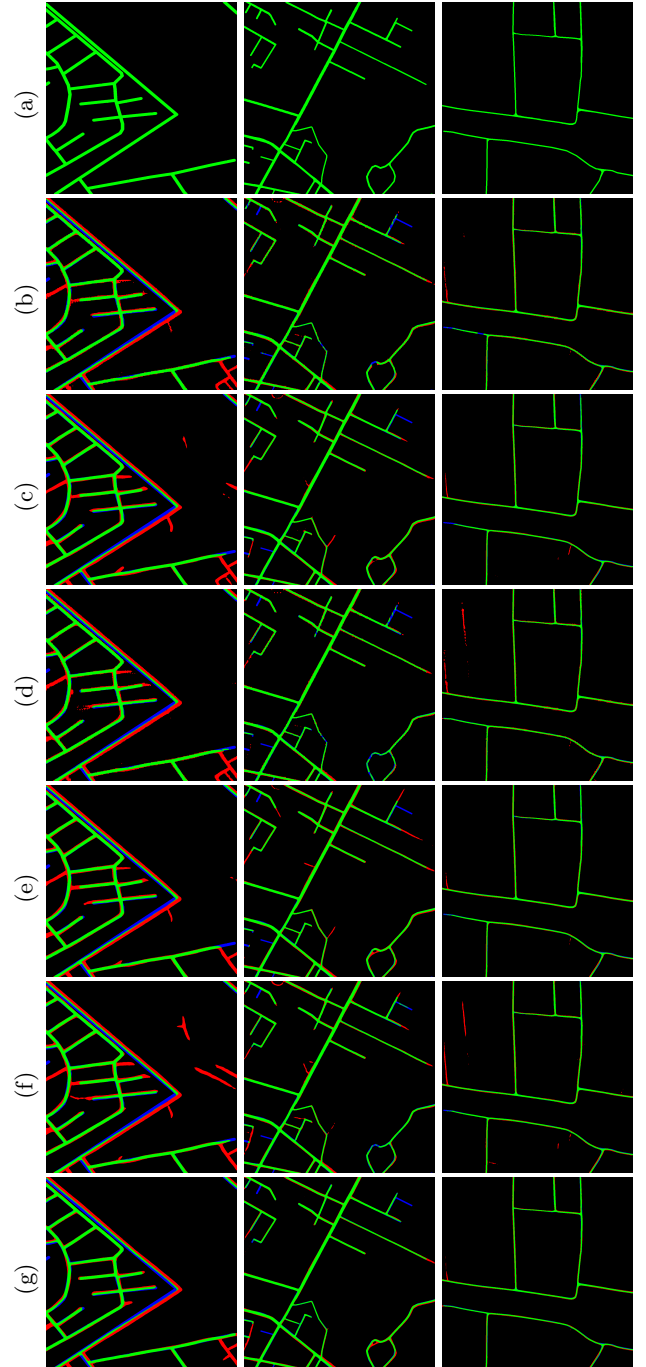


图 6: 所提方法在不同模型配置下的可视化结果。绿色: 真阳性, 红色: 假阳性, 蓝色: 假阴性。第一列到第三列显示来自 SpaceNet [63] 和 DeepGlobe [65] 的三个样本。(a) 真值图。(b) 基线。(c) 基线 + SCM。(d) 基线 + CoA。(e) CoANet 只包含  $d = 1$  的 CoA 模块。(f) 包含三个 CoA 模块的 CoANet。(g) CoANet。

在 DeepGlobe(图 Fig. 5中第八行)的结果中，道路来自农村地区，存在树木造成的严重遮挡。与我们的方法相比，其他方法提取的道路连通性较差。这些可视化结果验

表 III: 提出的条带卷积模块 (SCM) 和连接注意模块 (CoA) 的消融研究 (%). 基线是基于 ResNet-101 (No. 1) 的分割模型。我们添加了 SCM 模块和 CoA 模块来展示它们的有效性 (No. 2 和 No. 3)。第 4 号是我们提出的 CoANet 的完整版本。

No.	SCM	CoA	SpaceNet		DeepGlobe	
			IoU	APLS	IoU	APLS
1			59.57	58.69	63.09	69.13
2	✓		61.84	63.93	63.89	70.09
3		✓	61.10	61.27	64.32	70.45
4	✓	✓	<b>62.48</b>	<b>65.53</b>	<b>68.37</b>	<b>73.48</b>

表 IV: 具有不同配置的提出的连通性注意力模块 (CoA) 的消融研究。d1, d3, 和 d5 分别表示  $d = 1$ ,  $d = 3$ , and  $d = 5$  的 CoA 模块。No.1 是提出的只包含一个  $d = 1$  CoA 模块的 CoANet。No.2 代表我们方法的完整版本。No. 3 表示由三个 CoA 模块组成的连通性分支。

No.	d1	d3	d5	SpaceNet		DeepGlobe	
				IoU	APLS	IoU	APLS
1	✓			62.04	64.87	64.39	70.50
2	✓	✓		<b>62.48</b>	<b>65.53</b>	<b>68.37</b>	<b>73.48</b>
3	✓	✓	✓	62.09	65.13	64.89	70.83

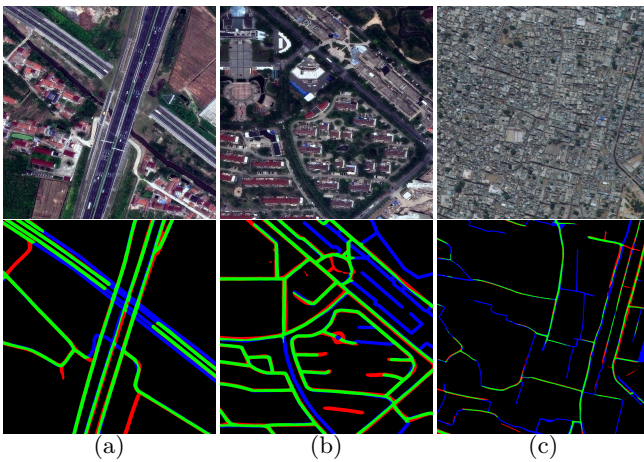


图 7: 某些情况下, 我们的 CoANet 失败的可视化结果。绿色: 真阳性, 红色: 假阳性, 蓝色: 假阴性。第一行是卫星图像, 第二行是我们 CoANet 的道路提取结果。(a)-(b) 来自 SpaceNet [63] 和 DeepGlobe [65] 的三个样本。

证了我们的 CoANet 在从卫星图像中提取道路任务中的优越性。

表 V: 具有不同配置的拟议条带卷积模块 (SCM) 的消融研究 (%). ‘H’、‘V’、‘L’ 和 ‘R’ 分别表示具有不同形状四个条带卷积: 水平、垂直、左对角线, 和右对角线。特别地, No. 1 是包含两个条状卷积的 SCM: 水平和垂直。4 号代表我们方法的完整版本。No.5 表示 SCM 每个形状包含两个卷积, 总共有 8 个卷积。

No.	SCM	SpaceNet		DeepGlobe	
		IoU	APLS	IoU	APLS
1	H&V	61.85	64.58	64.57	70.65
2	H&V&L	62.01	65.01	64.69	70.92
3	H&V&R	61.93	64.85	64.97	70.74
4	H&V&L&R	<b>62.48</b>	<b>65.53</b>	<b>68.37</b>	<b>73.48</b>
5	2×H&V&L&R	61.98	65.41	65.19	71.03

表 VI: 我们提出的 CoANet 和一些最先进的道路提取方法在相同条件下的运行时间。我们列出了相同批量大小的图像的训练时间和一张图像 (s) 的推理时间。

	Training	Inference
DeepRoadMapper [1] <i>ICCV17</i>	0.579	0.121
Topology Loss [26] <i>CVPR18</i>	0.168	0.049
LinkNet34 [15] <i>VCIP17</i>	0.201	0.066
D-LinkNet [39] <i>CVPRW18</i>	0.218	0.071
RoadCNN [5] <i>CVPR18</i>	0.156	0.028
ImprovedConnectivity [6] <i>CVPR19</i>	0.361	0.074
CoANet	0.146	0.022

## E. 消融研究

**我们提议的模块的有效性。**在 CoANet 模型中, 提出了 SCM 和 CoA 模块来分别捕捉道路区域的长程关系并探索相邻像素之间的依赖关系。为了验证这两个模块的有效性, 我们进行了不同配置实验, 如图 Table III 所示。基线 (No. 1) 是基于 ResNet-101 的 FCN 模型, 当我们把我们提出的 SCM 和 CoA 加入 baseline, 在 SpaceNet 数据集上, IoU 得分从 59.57% 提高到了 61.84% 和 61.10%。至于 SpaceNet 上的 APLS 分数, 添加 SCM 和 CoA 将基线提高了 5.24% 和 2.58%。在结合 SCM 和 CoA 之后, 我们在 IoU 分数上实现了 2.91% 的改进, 在 APLS 分数上实现了 6.84% 的改进。通过添加两个模块, DeepGlobe 数据集上的实验性能也得到了提升。这些结果验证了我们提出的两个模块对于道路提取是有效的。

我们的方法在不同设置下的定性结果如图 Fig. 6 所示。基线结果中存在一些断路段, 尤其是被树木遮挡的区域。添加 CoA 模块后, 基线中的大部分断路段是连接的, 而

道路的边缘是粗糙的，并且有一些离散点被识别为道路。SCM 模块有助于改善连通性，提取的道路更顺畅。但是，由于 SCM 是为捕获远程依赖关系而开发的，因此其他类别的某些区域可能会被识别为道路，这会使道路比实际更长。需要注意的是，如果我们单独使用这两个模块，会有一些缺点。SCM 模块连接 CoA 模块中相邻像素无法到达的断路段，CoA 模块可以防止来自背景的无关噪声。因此，我们的 CoANet 通过结合两个模块获得了最好的道路提取结果。

**不同配置对 CoA 的影响。**如 Sec. III-C 中描述的，我们提出的 CoANet 中的连接分支包含两个 CoA 模块，其中一个  $d = 1$ ，另一个  $d = 3$ 。这里我们分析了连接分支中 CoA 模块不同配置的效果，结果列于表 Table IV 中。我们定义 CoANet- $d1$  表示我们提出的 CoANet 只包含一个  $d = 1$  的 CoA 模块，CoANet- $d5$  表示我们的 CoANet 由  $d = 1, d = 3$ , and  $d = 5$  的三个 CoA 模块组成。CoANet- $d1$  在 SpaceNet [63] 上获得 62.04% 的 IoU 分数和 64.87% 的 APLS 分数。添加了  $d = 3$  的 CoA 模块之后，性能在 IoU 分数上提高到 62.48%，在 APLS 分数上提高到 65.53%。然而，如果 CoANet 由三个 CoA 模块组成 (Fig. IV 中的 No.3)，其性能高于 CoANet- $d1$  (No. 1)，但是低于有两个 CoA 模块的 CoANet (No. 2)。从可视化结果可以分析原因。如图 Fig. 6 所示，CoANet- $d1$  的结果中还有一些断路段此外，由于给定像素与其相邻像素的间隔更大，CoANet 的道路提取结果包含更多的背景区域并且更长。因此我们选择具有两个 CoA 模块的配置，它获得的结果更符合地面真值。

**SCM 不同配置的影响。**如表 Table V 所示，我们分析了解码器中不同配置对 SCM 的影响。当 SCM 包含水平和垂直条带卷积时，它在 SpaceNet 数据集的 IoU 得分上获得 61.85%，在 APLS 得分上获得 64.58%。可以通过添加一种带状卷积来提高性能，如表 Table V 中的 No. 2 和 No. 3。添加左对角线和右对角线条带卷积后，我们的 CoANet 在 IoU 得分上达到了 62.48%，在 APLS 得分上达到了 65.53%。以上实验结果符合我们的预期。使用不同形状的 4 个带状卷积，我们的方法可以捕获与多个空间方向沿本地上下文。需要注意的是，这四个方向与卫星图像中的大部分道路是一致的，并且相对容易实现。我们还报告了 SCM 每个形状包含两个卷积并且总共有八个卷积的配置结果。如 No. 5 所列，由于额外的条带卷积，它损害了性能的提高并增加了计算成本。因此，我们在 SCM 中应用了四个带状卷积。

## F. 讨论

**失败案例分析。**正如上面的实验中提到的，我们提出的 CoANet 在两个公共数据集 SpaceNet [63] 和 DeepGlobe [65] 上实现了新的最先进的性能但是，我们的方法仍然存在一些失败案例。正如图 Fig. 7 中展示的，我们展示了来自两个数据集的三个样本。对于 Fig. 7，卫星图像中间有一个隧道和一个立交桥。此外，由于树木和建筑物导致道路区域中的服务器阻塞，如图 Fig. 7 (b) 和 (c) 所示。由于遮挡区域非常大，并且这些区域的道路在卫星图像中可能不可见，我们的 CoANet 无法在这些区域生成道路并保持道路连通性。未来，我们将考虑添加其他信息来提取这些具有挑战性的区域的道路，例如行人和汽车的 GPS 轨迹。更重要的是，卫星图像中的道路网络可以被视为具有边和节点的图。我们可以利用图卷积网络来提取道路，这可能对遮挡区域的道路有效。

**运行时间分析。**我们在 SpaceNet [63] 数据集上分析了 CoANet 和几种最先进的道路提取方法的运行时间。

所有方法的对比实验均在具有 4 个 NVIDIA RTX TITAN GPU 的工作站上执行。为了公平比较，我们报告了相同批量大小的图像的训练时间和大小为  $512 \times 512$  的图像的推理时间值得注意的是，我们的 CoANet 实现了最快的训练时间和推理时间。此外，利用后处理步骤的方法，如 DeepRoadMapper [1] 和 ImprovedConnectivity [6]，需要更多的时间进行训练和推理。凭借更好的性能和更快的执行速度，我们提出的 CoANet 更适合从卫星图像中提取道路。

## V. 结论

在本文中，我们提出了一种用于从卫星图像中提取道路的连通性注意力网络 (CoANet)，它共同学习分割和成对依赖关系。我们首先介绍了一个编码器-解码器架构网络来学习道路的特征。受道路形状启发，这些道路是大跨度、狭窄且连续分布的。我们提出了一个带状卷积模块 (SCM)，因为它更符合道路的形状。SCM 利用四个带状卷积从四个不同方向捕获远程上下文信息，从而防止不相关区域干扰特征学习。更重要的是，为了减轻由建筑物和树木引起的道路区域的遮挡，开发了连通性注意力模块 (CoA) 来探索相邻像素之间的关系。对给定像素的八个相邻像素的连通进行预测，这其中包含了图形信息，使道路的连接性更好地保留。对流行基准 (SpaceNet 和 DeepGlobe 数据集) 的大量实验证明了我们提出的 CoANet 与几种最先进的方法相比的优越性。我们还进行了消融实验以展示 SCM 和 CoA 模块的有效性，并提供对 CoA 模块不同配置选择的见解。未来，我们的目标是开发更多可以使用连接注意模

块的任务，例如显著分割和语义分割，因为连接立方体的预测可以被视为分割任务的一系列子问题。此外，我们可以将卫星图像中的道路网络视为图，并利用图形卷积网络来提取道路。

### 致谢

本研究得到了 2018AAA0100400 号资助的新一代人工智能重大项目，国家自然科学基金 (61922046) 和天津市自然科学基金 (17JCJQJC43700) 的支持。

### 参考文献

- [1] G. Mátyus, W. Luo, and R. Urtasun, "Deeproadmapper: Extracting road topology from aerial images," in *IEEE Int. Conf. Comput. Vis.*, 2017, pp. 3438–3446.
- [2] A. Van Etten, "City-scale road extraction from satellite imagery v2: Road speeds and travel times," in *IEEE Winter Conf. Applications Comput. Vis.* IEEE, 2020, pp. 1775–1784.
- [3] R. Alshehhi and P. R. Marpu, "Hierarchical graph-based segmentation for extracting road networks from high-resolution satellite images," *ISPRS J. Photogramm. Remote Sens.*, vol. 126, pp. 245–260, 2017.
- [4] B. Liu, H. Wu, Y. Wang, and W. Liu, "Main road extraction from zy-3 grayscale imagery based on directional mathematical morphology and vgi prior knowledge in urban areas," *PLoS one*, vol. 10, no. 9, p. e0138071, 2015.
- [5] F. Bastani, S. He, S. Abbar, M. Alizadeh, H. Balakrishnan, S. Chawla, S. Madden, and D. DeWitt, "Roadtracer: Automatic extraction of road networks from aerial images," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2018, pp. 4720–4728.
- [6] A. Batra, S. Singh, G. Pang, S. Basu, C. Jawahar, and M. Paluri, "Improved road connectivity by joint learning of orientation and segmentation," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2019, pp. 10385–10393.
- [7] M. Barzohar and D. B. Cooper, "Automatic finding of main roads in aerial images by using geometric-stochastic models and estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 7, pp. 707–721, 1996.
- [8] I. Laptev, H. Mayer, T. Lindeberg, W. Eckstein, C. Steger, and A. Baumgartner, "Automatic extraction of roads from aerial images based on scale space and snakes," *Machine Vision and Applications*, vol. 12, no. 1, pp. 23–31, 2000.
- [9] D. Chai, W. Forstner, and F. Lafarge, "Recovering line-networks in images by junction-point processes," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2013, pp. 1894–1901.
- [10] R. Stoica, X. Descombes, and J. Zerubia, "A gibbs point process for road extraction from remotely sensed images," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 121–136, 2004.
- [11] Y. He, H. Wang, and B. Zhang, "Color-based road detection in urban traffic scenes," *IEEE Trans. Intell. Transport. Syst.*, vol. 5, no. 4, pp. 309–318, 2004.
- [12] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2015, pp. 3431–3440.
- [13] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Med. Image. Comput. Comput. Assist. Interv.* Springer, 2015, pp. 234–241.
- [14] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, 2017.
- [15] A. Chaurasia and E. Culurciello, "Linknet: Exploiting encoder representations for efficient semantic segmentation," in *IEEE Visual Communications and Image Processing*. IEEE, 2017, pp. 1–4.
- [16] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Eur. Conf. Comput. Vis.*, 2018, pp. 801–818.
- [17] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deepplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, 2017.
- [18] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.
- [19] V. Mnih and G. E. Hinton, "Learning to detect roads in high-resolution aerial images," in *Eur. Conf. Comput. Vis.* Springer, 2010, pp. 210–223.
- [20] G. Cheng, Y. Wang, S. Xu, H. Wang, S. Xiang, and C. Pan, "Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3322–3337, 2017.
- [21] T. Panboonyuen, K. Jitkajornwanich, S. Lawawirojwong, P. Srestasathien, and P. Vateekul, "Road segmentation of remotely-sensed images using deep convolutional neural networks with landscape metrics and conditional random fields," *Remote Sens.*, vol. 9, no. 7, p. 680, 2017.
- [22] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual u-net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, 2018.
- [23] J. D. Wegner, J. A. Montoya-Zegarra, and K. Schindler, "Road networks as collections of minimum cost paths," *ISPRS J. Photogramm. Remote Sens.*, vol. 108, pp. 128–137, 2015.
- [24] V. Mnih and G. E. Hinton, "Learning to label aerial images from noisy data," in *Int. Conf. Mach. Learn.*, 2012, pp. 567–574.
- [25] D. Costea and M. Leordeanu, "Aerial image geolocalization from recognition and matching of roads and intersections," *arXiv preprint arXiv:1605.08323*, 2016.
- [26] A. Mosinska, P. Marquez-Neila, M. Koziński, and P. Fua, "Beyond the pixel-wise loss for topology-aware delineation," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2018, pp. 3136–3145.
- [27] Y. Liu, J. Yao, X. Lu, M. Xia, X. Wang, and Y. Liu, "Roadnet: Learning to comprehensively analyze road networks in complex urban scenes from high-resolution remotely sensed images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2043–2056, 2018.
- [28] H. Kong, J.-Y. Audibert, and J. Ponce, "General road detection from a single image," *IEEE Trans. Image Process.*, vol. 19, no. 8, pp. 2211–2220, 2010.
- [29] W. Song, J. M. Keller, T. L. Haithcoat, and C. H. Davis, "Automated geospatial conflation of vector road maps to high resolution imagery," *IEEE Trans. Image Process.*, vol. 18, no. 2, pp. 388–400, 2008.

- [30] M. Amo, F. Martínez, and M. Torre, “Road extraction from aerial images using a region competition algorithm,” *IEEE Trans. Image Process.*, vol. 15, no. 5, pp. 1192–1201, 2006.
- [31] Q. Zhang and I. Couloigner, “Benefit of the angular texture signature for the separation of parking lots and roads on high resolution multi-spectral imagery,” *Pattern Recogn. Let.*, vol. 27, no. 9, pp. 937–946, 2006.
- [32] J. D. Wegner, J. A. Montoya-Zegarra, and K. Schindler, “A higher-order crf model for road network extraction,” in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2013, pp. 1698–1705.
- [33] 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 Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 2, pp. 652–662, 2021.
- [34] G. Lin, A. Milan, C. Shen, and I. Reid, “Refinenet: Multi-path refinement networks for high-resolution semantic segmentation,” in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2017, pp. 1925–1934.
- [35] S. Jégou, M. Drozdal, D. Vazquez, A. Romero, and Y. Bengio, “The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation,” in *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*, 2017, pp. 11–19.
- [36] H. Wu, J. Zhang, K. Huang, K. Liang, and Y. Yu, “Fastfcn: Rethinking dilated convolution in the backbone for semantic segmentation,” *arXiv preprint arXiv:1903.11816*, 2019.
- [37] T. Takikawa, D. Acuna, V. Jampani, and S. Fidler, “Gated-scnn: Gated shape cnns for semantic segmentation,” in *IEEE Int. Conf. Comput. Vis.*, 2019, pp. 5229–5238.
- [38] R. Alshehhi, P. R. Marpu, W. L. Woon, and M. Dalla Mura, “Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks,” *ISPRS J. Photogramm. Remote Sens.*, vol. 130, pp. 139–149, 2017.
- [39] L. Zhou, C. Zhang, and M. Wu, “D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction,” in *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*, 2018, pp. 182–186.
- [40] G. Mátyus and R. Urtasun, “Matching adversarial networks,” in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2018, pp. 8024–8032.
- [41] X. Li, Y. Wang, L. Zhang, S. Liu, J. Mei, and Y. Li, “Topology-enhanced urban road extraction via a geographic feature-enhanced network,” *IEEE Trans. Geosci. Remote Sens.*, 2020.
- [42] Y.-W. Choi, Y.-W. Jang, H.-J. Lee, and G.-S. Cho, “Three-dimensional lidar data classifying to extract road point in urban area,” *IEEE Geosci. Remote Sens. Let.*, vol. 5, no. 4, pp. 725–729, 2008.
- [43] M. Yadav, A. K. Singh, and B. Lohani, “Extraction of road surface from mobile lidar data of complex road environment,” *Int. J. Remote Sens.*, vol. 38, no. 16, pp. 4655–4682, 2017.
- [44] J. Liang, N. Homayounfar, W.-C. Ma, S. Wang, and R. Urtasun, “Convolutional recurrent network for road boundary extraction,” in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2019, pp. 9512–9521.
- [45] J. Hu, A. Razdan, J. C. Femiani, M. Cui, and P. Wonka, “Road network extraction and intersection detection from aerial images by tracking road footprints,” *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 4144–4157, 2007.
- [46] J. Biagioni and J. Eriksson, “Map inference in the face of noise and disparity,” in *Int. Conference on Advances in Geographic Information Systems*, 2012, pp. 79–88.
- [47] Z. Shan, H. Wu, W. Sun, and B. Zheng, “Cobweb: a robust map update system using gps trajectories,” in *Int. Joint Conference on Pervasive and Ubiquitous Computing*, 2015, pp. 927–937.
- [48] R. Stanojevic, S. Abbar, S. Thirumuruganathan, S. Chawla, F. Filali, and A. Aleimat, “Kharita: Robust map inference using graph spanners,” *arXiv preprint arXiv:1702.06025*, 2017.
- [49] J. Yuan and A. M. Cheriyyadat, “Image feature based gps trace filtering for road network generation and road segmentation,” *Machine Vision and Applications*, vol. 27, no. 1, pp. 1–12, 2016.
- [50] T. Sun, Z. Di, P. Che, C. Liu, and Y. Wang, “Leveraging crowdsourced gps data for road extraction from aerial imagery,” in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2019, pp. 7509–7518.
- [51] F. Shen, R. Gan, S. Yan, and G. Zeng, “Semantic segmentation via structured patch prediction, context crf and guidance crf,” in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2017, pp. 1953–1961.
- [52] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, “Deformable convolutional networks,” in *IEEE Int. Conf. Comput. Vis.*, 2017, pp. 764–773.
- [53] X. Wang, R. Girshick, A. Gupta, and K. He, “Non-local neural networks,” in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2018, pp. 7794–7803.
- [54] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, “Cenet: Criss-cross attention for semantic segmentation,” in *IEEE Int. Conf. Comput. Vis.*, 2019, pp. 603–612.
- [55] W. Wang, T. Zhou, F. Yu, J. Dai, E. Konukoglu, and L. Van Gool, “Exploring cross-image pixel contrast for semantic segmentation,” *arXiv preprint arXiv:2101.11939*, 2021.
- [56] G. Bertasius, L. Torresani, S. X. Yu, and J. Shi, “Convolutional random walk networks for semantic image segmentation,” in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2017, pp. 858–866.
- [57] Y. Cheng, R. Cai, Z. Li, X. Zhao, and K. Huang, “Locality-sensitive deconvolution networks with gated fusion for rgb-d indoor semantic segmentation,” in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2017, pp. 3029–3037.
- [58] J. Ahn and S. Kwak, “Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation,” in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2018, pp. 4981–4990.
- [59] Y. Hou, Z. Ma, C. Liu, T.-W. Hui, and C. C. Loy, “Inter-region affinity distillation for road marking segmentation,” in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2020, pp. 12486–12495.
- [60] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *IEEE Conf. Comput. Vis. Pattern Recog.* Ieee, 2009, pp. 248–255.
- [61] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2016, pp. 770–778.
- [62] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2018, pp. 7132–7141.
- [63] A. Van Etten, D. Lindenbaum, and T. M. Bacastow, “Spacenet: A remote sensing dataset and challenge series,” *arXiv preprint arXiv:1807.01232*, 2018.
- [64] S. Singh, A. Batra, G. Pang, L. Torresani, S. Basu, M. Paluri, and C. Jawahar, “Self-supervised feature learning for semantic

segmentation of overhead imagery.” in *Brit. Mach. Vis. Conf.*, vol. 1, no. 2, 2018, p. 4.

- [65] I. Demir, K. Koperski, D. Lindenbaum, G. Pang, J. Huang, S. Basu, F. Hughes, D. Tuia, and R. Raska, “Deepglobe 2018: A challenge to parse the earth through satellite images,” in *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.* IEEE, 2018, pp. 172–17209.
- [66] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, “Pytorch: An imperative style, high-performance deep learning library,” in *Adv. Neural Inform. Process. Syst.*, 2019, pp. 8026–8037.
- [67] S.-M. Hu, D. Liang, G.-Y. Yang, G.-W. Yang, and W.-Y. Zhou, “Jittor: a novel deep learning framework with meta-operators and unified graph execution,” *Science China Information Sciences*, vol. 63, no. 12, pp. 1–21, 2020.
- [68] Y.-Q. Tan, S.-H. Gao, X.-Y. Li, M.-M. Cheng, and B. Ren, “Veroad: Point-based iterative graph exploration for road graphs extraction,” in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2020, pp. 8910–8918.



**Ming-Ming Cheng** received his Ph.D. degree from Tsinghua University in 2012. Then he did two years research fellow with Prof. Philip Torr in Oxford. He is now a professor at Nankai University, leading the Media Computing Lab. His research interests include computer graphics, computer vision, and image processing. He received research awards, including ACM China Rising Star Award, IBM Global SUR Award, and CCF-Intel Young Faculty Researcher Program. He is on the editorial boards of IEEE TIP.



**Jie Mei** is a Ph.D. student in College of Computer Science, Nankai University. He is supervised via Prof. Ming-Ming Cheng. His research interests include computer vision, machine learning, and remote sensing image processing.



**Rou-Jing Li** is currently a master at Faculty of Geographical Science, Beijing Normal University, Beijing, China. Her research interests are remote sensing image processing, machine learning, and spatial analysis.



**Wang Gao** received his master degree from the Third Research Institute of China Aerospace Science and Industry Corporation in 2017. He is currently with the Science and Technology on Complex System Control and Intelligent Agent Cooperation Laboratory, Beijing, China. His research interests include computer vision, scene matching, and

visual navigation.