

# VecRoad: 用于道路图提取的基于点的迭代图探索方法

Yong-Qiang Tan    Shang-Hua Gao    Xuan-Yi Li    Ming-Ming Cheng    Bo Ren✉

TKLNDST, College of CS, Nankai University

<https://mmcheng.net/vecroad/>

## Abstract

与直接从野外采集相比, 自动从航拍图像中提取路线图是一种更有效且成本更低方法。这种方法可以通过对 CNN 预测得到的道路分割进行矢量化的后处理步骤来实现, 但是不完善的预测会导致道路图的连通性较低。另一方面, 可以通过迭代式的下一步移动探索的方式来构建具有更好道路连通性的道路图, 但是在该过程只关注了局部信息, 通常无法实现与真实道路的精确对齐。为了增强道路连通性, 同时保持图与道路之间的精确对齐, 我们提出了一种基于点的迭代图探索方案, 该方案具有分割提示引导和灵活的步骤。在我们的方法中, 我们将下一个移动的位置表示为一个“点”, 该点统一了多个约束的表示形式, 例如每个移动步骤中的方向和步长。诸如路段分割和路口交叉口的信息提示将被联合检测, 并用于指导下一步行动和实现更好的道路对齐。我们证明, 相对于最新的道路图提取方法, 我们提出的方法在常见数据集上的 F 度量和道路连通性指标方面有很大的改进。

## 1. 介绍

道路图是道路地图的矢量化表示形式, 它可以用于诸如最短路径搜索导航等现实应用。通常, 可靠的路线图是通过昂贵且费时的现场采集和手动标记生成的。近年来, 采用卷积神经网络 (CNN) [19, 1, 25] 从航空图像中自动构建高精度, 覆盖范围广的道路图, 减少了很多工作量。最常见是使用后

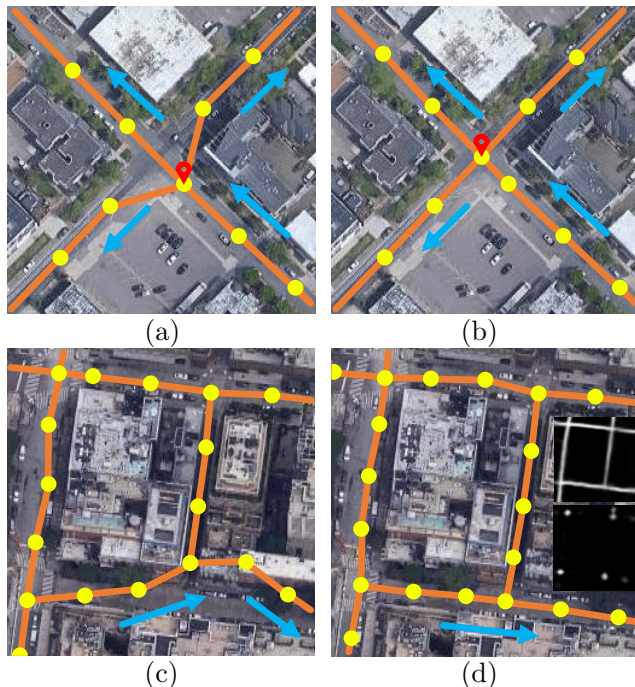


图 1. (a) 由于固定的移动步长大小, 导致路口平移。(b) 以我们提出的灵活步长移动。(c) 在没有分割指导的情况下, 图形和道路之间的路线未对齐。(d) 分割线索指导的下一动作预测生成具有精确对齐的图形。

处理方法 [16, 2], 例如形态学运算 [30] 和硬编码规则 [16, 7], 从 CNN 预测的骨架分割中提取道路图。然而, 使用这种方式获得的图受分割质量的影响很大, 在这种情况下, 间歇性分割通常会导致图的连接性较低。为了加强道路连通性, 最近提出了一些通过迭代的下一步移动探索来构造道路图的方法 [1, 25]。在他们的方法中, 通过预测局部补丁中的下一个动作并将其连接到当前路线图, 可以通过迭代方法生

成完整的路线图。但是，迭代下一步探索在估计下一步动作时仅侧重于局部信息，这可能会导致图 and 实际道路之间的位置不一致。如 Fig. 1(c) 所示，即使在大致正确的方向上探索该图，但由于其局部的下一步移动查找策略，预测图的一部分会位于实际路线图之外。另外，如 Fig. 1(a) 所示，现有方法中的固定步长的方式很容易导致图中的路口偏移。

为了在保持图与道路之间精确对齐的同时增强道路连通性，我们提出了一种基于点的迭代图探索方案，该方案具有分割提示指导和灵活的步骤。我们首先将下一个动作的位置表示为“点”，以统一多个约束条件的表示方法，例如每个运动步骤中的方向和步长。我们设计的网络会在每个下一个移动推理步骤中通过学习输出多个估计点位置的高斯概率分布。与以前使用移动角度作为下一步指示的方法相比 (Fig. 2(a))，我们的新的表示形式能够轻松推导方向和步长，同时避免了训练过程中复杂的多重监督。我们提出的下一步“点”表示的示例如图 2(b) 和 (c) 所示。通过监督点坐标，我们提出的方法学会了在非平凡点（路口，路口和连接点）处以现有道路图的灵活步长预测下一招的正确位置，如图 1(b) 和 Fig. 4 所示。与推理阶段相同，在训练阶段，我们还可以利用分割提示可以提供的全局信息，从而对道路进行概述。因此，我们使用道路分割和路口线索作为隐式指导来预测具有精确对齐的道路图，如图 1(d) 所示。对于端到端设计，我们在具有共享主干的统一网络中共同提取道路分割和路口线索以及下一步行动预测。我们的主要贡献如下所示：

- 基于点的迭代式下步移动探索方法，具有灵活的步长检测技术，可以在下一步移动探索中精确定位非平凡的点。
- 从分割线索进行探索的指导，生成具有良好连通性和对齐精度的路线图。

## 2. 相关工作

### 2.1. 道路分割

从航空影像中提取道路成二进制像素是遥感领域中一项经过充分研究的任务。传统方法通过各种

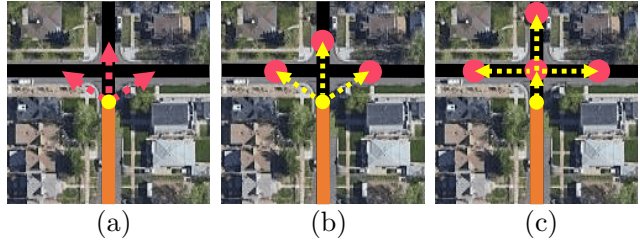


图 2. 基于探索的方法的下一步的表示（用红色指示符显示）。(a) 固定步长的角度；(b) 固定步长的点；(c) 可变步长的点。灵活的步长可以更好地表示道路的几何形状并确定路口的位置。

技术来构造路线图，例如利用附近的建筑物和车辆 [11]，形状因子 [21]，模拟退火技术 [22] 以及独特的光谱对比度和局部线性轨迹 [6]。在构建道路图时也用了最小生成树 [23]，高阶条件随机场 [26, 27] 和连接过程 [3] 等方法。

最近的工作将深度学习应用于生成性能更高的路线图。在 [17] 中，受限的玻尔兹曼机被应用于道路检测，而预处理则用于减少输入数据的维数。

后处理则用于进一步去除不连续的点并填充道路上的孔洞。Saito 等人 [19] 使用 CNN 直接从原始遥感影像生成道路分割，无需进行预处理。

Cheng 等人 [5] 使用级联神经网络提取了道路中心线。Zhang 等人 [31] 将残差连接 [10] 应用于 U-Net [18]，从而学习得到道路分割的更多精细特征。D-linknet [33] 结合了空洞卷积 [29] 和 Linknet [4]，以扩大从高分辨率卫星图像中提取的道路的接收场。

### 2.2. 道路图构建

要生成精细的路线图，即路线图的矢量化表示，应同时考虑连通性和校准。有两个主流框架可以获取路线图。一种利用分割和后期处理，另一种从航空图像直接转换为图。

**道路分割后处理。** 后处理方法采用阈值对道路分割进行二值化处理。然后应用形态学细化技术获得一个一像素宽的道路骨架。为了消除图的冗余，以前的方法采用了 Ramer-Douglas-Peucker 算法 [7]。Máttyus 等人 [16] 在第一个过程中使用具有 soft-IOU 损失的轻量 CNN 来生成分割输出。图转换完成后，他们使用  $A^*$  算法（最短路径问题）消除短

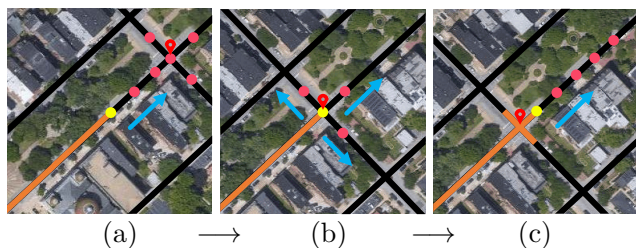


图 3. 从 (a) 到 (c) 的连续的基于点的迭代探索, 显示了下一移动轨迹, 而且没有连接的歧义。黑线是真实的道路图注释, 而橙色线是步行路径。黄点是当前顶点的坐标。粉色点表示连接点, 红色点是高斯分布表示的目标。

边和缺少连接的部分。Batra 等人 [2] 引入了定向学习和删除细化学习。定向学习使神经网络能够处理像素之间的连接。此外, 删除细化学习会学习道路连接的模式, 并从第一步开始优化道路的分割输出。所获得的路线图在 APLS 指标 [24] 上具有更好的连通性。

**迭代式路线图探索** Bastani 等人 [1] 采用迭代探索算法直接生成路线图。首先, 在训练和推断时分别从图的 ground-truth 和额外的道路分割网络生成起点。然后, 他们将以起始点为中心的航拍图像迭代地将裁剪后的补丁喂进神经网络中。网络输出之一是确定算法是否需要继续搜索或停止的概率。另一个输出是角度矢量, 它表示从当前顶点到下一个移动的角度。他们使用地图匹配算法来使当前道路保持搜索路径, 而非进入其他道路。Li 等人 [14] 利用多边形来适应道路和建筑物的形状。基于多边形的方法使用 CNN-RNN 体系结构, 以反复提取道路几何形状的关键点。然后应用右手定则勾勒出道路轮廓。

### 3. 方法

迭代探索框架 [1] 通过不断预测下一步行动并将其合并到现有的路线图中来构造路线图。我们采用此框架, 并提出了几种方案来改善路线图构建的性能。我们利用基于点的下一移动表示形式, 它是移动角度和距离的统一组合。得益于点表示, 无需复杂的监督即可轻松应用多个约束。我们提出了一种灵活的步长检测技术, 该技术旨在在训练阶段与路

口动态对齐。我们利用道路和路口分割线索来指导探索并实现更好的道路对齐。我们将这些方案整合到我们提出的框架路点网络 (RP-Net) 中, 以生成具有高度连接性和精确对齐的路标。在本节中, 我们将首先回顾迭代探索框架, 然后描述我们方法的细节。

#### 3.1. 迭代探索概述

道路图  $G$  是道路地图的一种向量化表示, 其中包括一个顶点集  $V = \{v_1, v_2, \dots, v_n\}$  和一个边集  $E = \{e_1, e_2, \dots, e_m\}$ 。边  $e$  是两个顶点之间的直线, 表示这两个顶点之间的道路。

道路图的构建是通过迭代探索沿道路的新顶点并将其添加到现有道路图  $G$  中的方式进行的, 并且在两个顶点之间有一条边。具体而言, 迭代探索从指示探索起点的起始顶点集  $S$  开始。通常,  $S$  是从道路分割 [1] 或路口分割的峰值点获得的。 $G$  中的节点集  $V$  初始化为  $S$  的拷贝。在每次探索中, 都从  $S$  弹出一个顶点  $v$  作为起点。

神经网络将以该顶点为中心的航拍图像补丁作为输入, 并预测下一个顶点集  $V'$ 。如果预测得到的  $v \in V'$  在  $V$  的相同区域中具有匹配的顶点, 则将使用匹配的顶点作为新获得的顶点。然后, 将新获得的顶点以及现有和新顶点之间的道路分别添加到  $V$  和  $E$ , 以形成新的  $G$ 。 $S$  由  $S \cup V'$  更新。从  $S$  获得新的起始顶点以开始新的探索。当  $S$  为空时探索停止。在 Fig. 3 中, 我们展示了探索过程中的动态过程。随着探索的进行, 该图不断被迭代构造。

#### 3.2. 基于点的迭代式探索

**基于点的下步预测。** 在这项工作中, 我们将下一个移动的位置表示为“点”, 统一了移动角度和距离的表示, 如 Fig. 2(b) 所示。在训练阶段, 下一步的监视始终设置在道路的中心线, 因此可以确保输出迭代地跟踪实际道路。在高斯分布的监督下, 以下一个动作的位置为中心, 对探测探测器进行了训练。将基于点的探索作为像素任务, 神经网络可以精确地预测“行进中”的下一步行动。在推论过程中, 可以从预测分布的峰值中获取下一步的位置。在训练阶

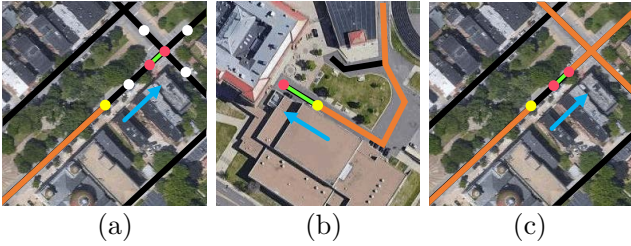


图 4. 当下一步遇到 (a) 道路交叉点, (b) 道路末端和 (c) 链接点时, 采用灵活的步长 (用绿线表示)。橙色线是人行道, 是连接链接点的重要先验。

段很容易在点表示上应用多个约束 (例如, 方向和步长), 而无需复杂的多种形式的监督。在下面, 我们将详细讨论我们的方法。

**灵活的步长方案.** 在 [1] 中, 使用了具有固定步长的角度分类器来检测多达 64 个顶点, 如 Fig. 2(a) 所示。如 Fig. 4 所示, 道路上有多种非平凡点, 例如路口, 道路尽头和连接点当前位置与附近非平凡下步动作之间的道路长度几乎无法匹配固定步长的整数倍。如 Fig. 1(a) 所示, 步长固定的检测器在下一步遇到路口时, 可能会生成与实际道路不对齐的图形。为了确保道路和交叉路口的精确对齐, 我们设计了一种灵活的步长方案。在训练阶段, 我们在 ground-truth 的监督下对空的图进行探索。在每个探索步骤中, 我们都会动态遵循 ground-truth, 以生成下一步行动监管。我们将固定步长标记为  $s$ , 可变步长大小被设计为  $0.5 \times s$  和  $1.5 \times s$  之间的可调整大小。当在距当前顶点  $s$  的 1.5 个标度内有一个平凡点时, 我们将在检测到的点上精确地生成监督高斯分布。通过灵活的步长, 可以轻松处理非平凡点 (例如路口), 因此图形将与真实道路对齐。此方案还有助于增强图的连接性。Fig. 4(c) 显示了一种特定情况, 使用灵活步长, 可以轻松匹配并连接先前中断的探索的折断端点。对于下一个探索区域中没有非平凡点的情况, 在 Fig. 3(c) 中, 我们使用从起点开始的固定步长沿 ground-truth 生成新的监督高斯分布。总的来说, 我们在道路中间使用固定步长监控, 并在非平凡点附近切换为灵活的步长。

如果想要使用角度学习的常规方法适应这种灵活步长, 则必须仔细设计额外的步长学习过程。相反, 基于我们提出的“点”表示, 无需添加移动距离检测分支, 基于点的检测器即可通过训练以移动距离编码的基于点的监督来学习性能良好的灵活步长, 如 Fig. 2(c) 所示

**轨迹探索.** 在迭代探索的框架中, 每个步骤都可能带来微小的错误。受强化学习中长期奖励和经验重播机制的启发, 我们认为可以一次直接预测移动的轨迹, 而不是仅预测一步。我们通过将向下采样的下一移动预测循环发送回下一移动检测器 (沙漏块) 最多  $T$  次来实现这一点。值得一提的是, 给定一个航空图像作为输入, 我们只提取一次图像特征。通过使用递归机制, 神经网络将获得对未来轨迹的更长视线并减少总误差。

### 3.3. 分割线索

与专注于局部的下步行动的探索机制 (如 Fig. 2) 不同, 分割技术具有更全局的关注点。对比 [1] 在 Fig. 1(c) 的结果和我们的结果 (d), 我们可以看到, 如果没有如下所述的全局的关于在何处进行长远探索的知识, 该探索将导致道路和路口错位。

**道路分割线索.** 道路分割的目标是从航空影像中提取道路的中心线 [19, 5]。如 Fig. 1(b) 所示, 道路中心线可以从宏观的角度更好地表示道路图的拓扑。在这里, 我们将解释在我们的方法中使用道路分割的两个关键思想。首先, 迭代探索方法主要关注局部下一步动作的位置, 但缺乏对道路区域 (即真实道路所在的位置) 的全面了解。特别的, 如 Fig. 7(e), (g) 和 (h) 所示, 通过道路分割对道路区域进行全局指导可以减少与实际道路的错位。其次, 以下一步预测的形式, 可以将道路分割视为探索点的理想选择。因此, 道路分割可以在探索之前提供适当的指导和中心线。

**路口分隔线索.** 路口分割, 恰好适用于我们的灵活步长方法, 可以指导路口前方的下一步行动预测。由于航拍图像中的路口通常为区域形式, 因此路口分

割提示可以帮助网络在训练过程中准确地学习最佳路口位置。例如，在没有路口提示的帮助下，当几个路段在路口区域相遇时，探索方法可能会发现更少或更多的十字路口。如 Fig. 7(e), (f) 和 (h) 所示，如果没有路口分割作为支持，探索时间会在一个复杂路口上挣扎，而在路口的指导下，预测出的图更有条理。与使用道路分割的原因类似，路口分割可以给出路口位置的先验信息，这有助于神经网络更好地识别距离模式并确定达到路口的精确坐标的步长。

### 3.4. 网络结构

在我们的网络设计中，如 Fig. 5 中所示，采用了 VGG 主干的侧面融合  $\mathcal{F}$  提取四分之一尺度的航空影像的金字塔特征。为了使神经网络预测更稳健的前进方向，我们从推断的图显式生成探索路径分割  $\mathcal{W}$  作为中间输入。我们利用侧面输出特征融合  $\mathcal{F}$  生成四分之一尺度的道路和路口分段预测，分别用  $\hat{\mathcal{R}}$  和  $\hat{\mathcal{J}}$  表示。我们采用道路监管  $\mathcal{R}$  和交叉路口监管  $\mathcal{J}$  来引导骨干网络学习道路的基本表示。沙漏块是一个金字塔特征自动编码器，基本上将道路主干特征和  $\mathcal{W}$  作为输入，通过快速下采样和上采样来融合特征。当采用分割提示作为隐式指导时，我们的沙漏块的输入将进一步通过分段提示特征图连接为  $(\mathcal{F}, \hat{\mathcal{R}}, \hat{\mathcal{J}}, \mathcal{W})$ 。在这里，为了简单起见，我们在表示  $\hat{\mathcal{R}}$  和  $\hat{\mathcal{J}}$  时不区分中间分割特征图和输出预测。 $\hat{\mathcal{R}}$  和  $\hat{\mathcal{J}}$  的隐式指导将充当输入，并为该设计带来良好的可解释性。在沙漏块之后，下一步  $\hat{\mathcal{A}}'$  的高斯分布将由实际高斯图  $\mathcal{A}$  的侧面监督来保证。我们网络的解码器部分旨在利用高级别和低级别的道路信息来放大和完善预测，从而帮助精确生成预测分布。最后，将应用监督  $\mathcal{A}$  来保证精准分配  $\hat{\mathcal{A}}$ 。此外，由于多任务的联合学习，我们的方法是端到端地训练的，不需要单独的网络来获取起点。

为了反复预测  $T$  步，我们将最终预测下采样为四分之一尺度，并通过与上述沙漏块的输入串联使用来实现重用。我们最初使用占位符  $\hat{\mathcal{A}}_0$  来确保特征通道的一致性。给定  $(\mathcal{F}, \hat{\mathcal{R}}, \hat{\mathcal{J}}, \mathcal{W}, \sum_{i=1}^1 D(\hat{\mathcal{A}}_i), \dots, \sum_{i=1}^t D(\hat{\mathcal{A}}_i))$ ，我们获得下步移动的概率图  $\hat{\mathcal{A}}_{t+1}$ 。在这里， $D(\cdot)$  表示下采

样操作。因此，我们可以循环获得  $T$  个时间步长中的  $\hat{\mathcal{A}}_t$ 。

如果道路分割在  $T$  个时间步长之前就遇到了一个路口，不妨设在  $t = k, k < T$  的位置，我们在计算损失函数时将会忽略  $t > k + 1$  的步长，这是因为探索  $t = k + 1$  之后的监督对于顶点连接将是不确定的。

至于网络架构设计的细节，沙漏模块是通过带有残差连接的 4 层下采样和 4 层上采样构成的。每层包含两个 Conv-ReLU 层，内核大小为 3。每个解码器块将根据前一个块计算出的 32 通道的主干特征和 32 通道的下步移动特征求和，然后是两个  $3 \times 3$  卷积层。我们使用标准的二元交叉熵损失来分别优化  $\hat{\mathcal{R}}, \hat{\mathcal{J}}$ , and  $\sum_{t=0}^T \hat{\mathcal{A}}_t$ 。总的损失函数为：

$$\mathcal{L} = \sum_{t=0}^T \left( L(\hat{\mathcal{A}}_t, \mathcal{A}_t) + L(U(\mathcal{A}'_t), \mathcal{A}_t) \right) + \lambda_1 L(\hat{\mathcal{R}}, \mathcal{R}) + \lambda_2 L(\hat{\mathcal{J}}, \mathcal{J}), \quad (1)$$

其中  $L(X, Y)$  是预测矩阵  $X$  和 ground-truth 矩阵  $Y$  之间的二元交叉熵损失。 $U(\cdot)$  表示上采样函数， $\lambda$  是用来平衡多类损失的参数。在 Equ. (1) 中， $\mathcal{T}$  由  $\min(k + 1, T)$  决定。损失函数中  $\lambda_1$  和  $\lambda_2$  的值均设为 1。

### 3.5. 实现细节

跟随 RoadTracer 的方法 [1]，我们动态地生成 ground-truth 的下一步来训练神经网络。这里，在训练过程中，我们采用空的图进行探索，并采用监督图来指导下一步行动的坐标。为了确保训练过程的独立同分布，我们从不同的航拍图像中采样了一批训练补丁，并在每个训练图像的起点集中应用了随机取出的方法。由于随机图探索本身就是一种数据增强技术，因此我们不应用额外的数据增强。

RP-Net 使用  $256 \times 256$  分辨率的航拍图像的裁剪补丁进行训练，这是在精度和效率之间进行权衡的结果。我们采用 Pytorch 框架和发布的 VGG-16 模型 [20] 作为初始化。我们使用 Adam 优化器 [13] 对网络进行了 102,400 次迭代训练。我们的初始学习率为  $1e - 3$ ，并且我们在 40960 次迭代中将学习率

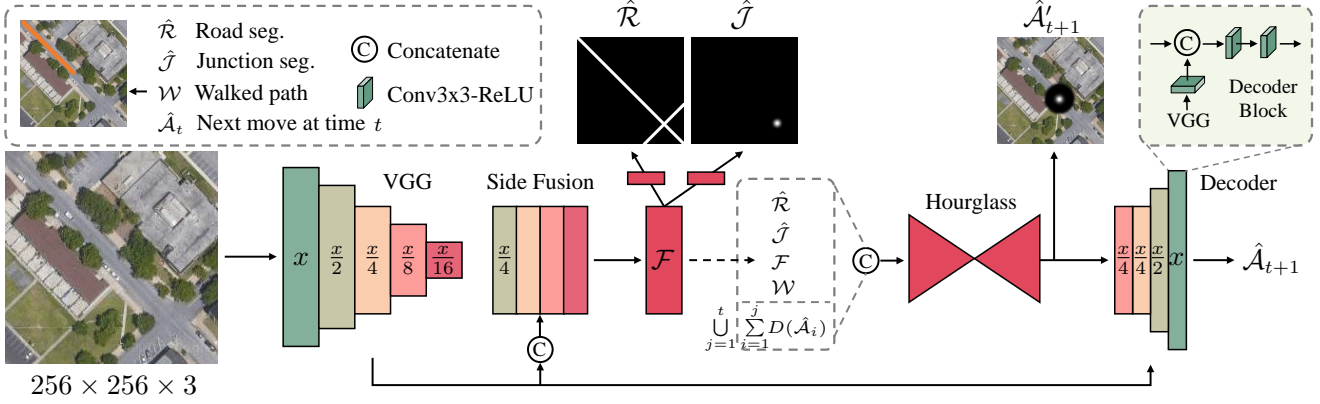


图 5. RP-Net 网络结构概述。矩形是不同尺度下的特征图。每个特征图的颜色表示相应 VGG 阶段的用法。循环部分将  $(\mathcal{F}, \hat{\mathcal{R}}, \hat{\mathcal{J}}, \mathcal{W}, \sum_{i=1}^1 D(\hat{\mathcal{A}}_i), \sum_{i=1}^2 D(\hat{\mathcal{A}}_i), \dots, \sum_{i=1}^t D(\hat{\mathcal{A}}_i))$  作为输入，并按照时间序列输出下步动作  $\hat{\mathcal{A}}_{t+1}$ 。

降低了 0.1 倍。我们使用 24 的 batch 大小来使用 2 个 NVidia Titan XP 显卡训练模型。训练数据没有使用数据增强处理。在推断时，我们使用 0.4 为阈值进行从下一个移动概率图到坐标的转换。总时间步长  $T$  的值是图像大小和步长大小之间的权衡。我们在这个工作中使用  $T = 4$  来提供足够的轨迹长度，并且确保下一个动作同时在输入图像内。应该注意的是，较大的图像尺寸在保持批处理大小的同时需要更多的 GPU 资源。在训练阶段，固定步长为 20 像素，可变步长可在 10 到 30 像素之间调整，目的是使其可以动态拟合距当前路段末端的距离。由于我们的分割线索，我们不需要额外的网络来提取起点，因为我们可以从路口分割和道路分割的峰值输出获得  $S$ 。我们网络中的参数数量为 20M，小于 [1] 中 21M（用于生成起点）加 26M（用于迭代探索）的参数数量。

## 4. 实验

我们在 RoadTracer 数据集 [1] 上定量和定性地验证了我们的方法，并在本节中提供了详细信息。

### 4.1. 评估指标

我们在道路校准和图形连通性上评估了我们的结果。依照 [17, 19]，我们采用像素度量来研究道路中心线蒙版的像素级别对齐方式。在这里，道路中心线遮罩是通过在 2D 地图上以 8 像素的固定宽度绘制道路图来生成的。道路宽度被视为道路中心线的松弛。

为了更好地评估道路连通性和拓扑结构，我们依据 [1] 评估路口级的精度召回率。[1] 中的  $F_{correct}$  和  $F_{error}$  可以通过将  $1 - F_{error}$  当做精度，将  $F_{correct}$  当做召回率的方式进一步计算 F-score。我们统一采用平均 F-score 来表示像素度量和路口度量的精度和召回率的综合性能，其中像素度量为 ‘P-F1’，路口度量为 ‘J-F1’。有关像素度量和路口度量的更多细节将在补充材料中进行解释。

平均路径长度相似性度量 (APLS) 由 [24] 引入。APLS 度量分别具有来自预测图和 ground-truth 图的所有对应节点对，可研究它们之间的最短路径长度差：

$$APLS = \frac{1}{N} \sum \left( \frac{2}{\frac{1}{S_{\hat{G} \rightarrow G}} + \frac{1}{S_{G \rightarrow \hat{G}}}} \right), \quad (2)$$

其中

$$S_{\hat{G} \rightarrow G} = 1 - \frac{1}{M} \sum \min \left( 1, \frac{|L(a, b) - L(\hat{a}, \hat{b})|}{L(a, b)} \right) \quad (3)$$

是  $\hat{G}$  到  $G$  最短路径长度分数。在 Equ. (3) 中， $M$  是  $\hat{G} \rightarrow G$  中独立的路径数量。 $L(\hat{a}, \hat{b})$  和  $L(a, b)$  代表路径长度，其中  $(\hat{a}, \hat{b})$  在  $\hat{G}$  中， $(a, b)$  在  $G$  中。

在 Equ. (2) 中， $N$  是属于数据集的图像数量。

### 4.2. 与最新技术的比较

我们将我们的方法与以前的最新技术进行了比较 [16, 1, 2]。如 Tab. 1 所示，我们观察到，在所有这三个评估指标上，我们的方法均优于最新技术。特别

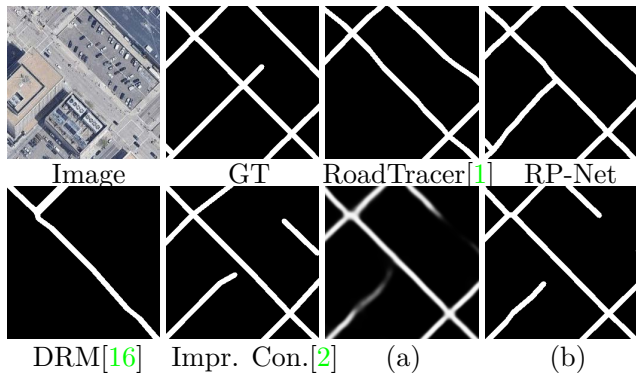


图 6. Tab. 1中提到的各种方法的定性比较。(a)从 RP-Net-Seg 生成的细分结果。,(b)从 RP-Net-Seg 生成的图。使用后处理。

是,我们的方法在路口度量上具有优越的性能,显示了基于点的方法在恢复道路交叉口特征方面的优势。

对于比较中使用的方法,我们在 Fig. 6中提出了定量可视化在补充材料中将进行更多的可视化比较。由于在分割监督中缺少连通性信息,因此当发生阴影或遮挡时,神经网络具有间歇性输出。RoadTracer[1]是一个更鲁棒的从航空影像中生成图的方法,但是由于角度学习设计的固定步长,该方法没有考虑路口的移动。请注意,RoadTracer的结果缩小了输入比例,因此我们也通过缩小 ground-truth 来公平地评估其性能,并在 Tab. 1中将其报告为“RoadTracer-256”。[2]方法是基于分割的方法,同时也采用了复杂的硬编码后处理。尽管路段分割可以考虑连通性并生成性能良好的 mask,但复杂的后处理又会导致几何变形。具有 Res2Net[8]主干的版本稍微牺牲了像素级分数的高分辨率,但由于多尺度聚合和自适应接收场 [32],因此获得了更好的连接性分数。

### 4.3. 消融研究

我们通过逐步应用来研究方法的改进。如 Tab. 2所示,最初,我们执行基线实验时没有使用灵活步长,分割提示和轨迹探索。

**灵活步长.** 通过在基准线方法上利用灵活的步长,可以改善像素级道路对齐方式和交叉路口连接性。

Method	P-F1	J-F1	APLS
DeepRoadMapper [16] <sup>†</sup>	56.85	29.05	21.27
RoadTracer [1]	55.81	49.57	45.09
RoadTracer-256 [1]	59.69	52.19	-
ImprovedConnectivity [2] <sup>†</sup>	73.35	55.21	56.89
RP-Net-Seg. <sup>†</sup>	71.61	50.16	49.68
RP-Net-Full	73.69	62.36	61.14
RP-Net-Full+Res2Net	72.56	63.13	64.59

表 1. 在 RoadTracer 道路数据集上的性能对比 ‘P-F1’ and ‘J-F1’ 分别表示像素级和路口级的 F-score。<sup>†</sup> 意味着使用来自 [1]的后处理方法。RP-Net-Full+Res2Net 表示额外使用 Res2Net 作为骨架的特征提取器。

在这里,可以从“J-F1”得分中揭示出路口方向的连通性。具体来说,路口度量提高了 7.58%,这主要受益于精确的路口位置。如 Fig. 8展示的,路口滑动情况明显受到抑制。

**分割线索.** 我们通过执行基于可变步长的方法来研究道路分割和路口分割的有效性。我们首先仅应用道路分割提示,即对融合的 VGG 特征和道路分割特征的级联进行侧面监督,作为下一个运动预测变量的输入,正如 Sec3 中所述。

只使用道路分割线索,像素级指标提升了 11.86%,路口级像素提升了 12.38%。其次,我们仅应用路口分割提示,这有助于精确地确定路口的位置,并进一步帮助灵活的步进方法找到准确的步长。与单独使用道路分割提示相比,两个指标均得到类似的改善,但像素指标的改善却较小。这是我们预期内的结果,因为错误的路口位置始终会导致错误的道路路线,但事实却并非如此。因此,使用道路分割来实施道路路线调整会产生更明显的效果。

但是,只有结合两个分割提示,我们才能观察到 APLS 指标的大幅提高。其背后的原因是这两个分割提示具有互补的特点。道路分割提示引导的改进主要集中于道路中心线的对齐,有助于避免因不在道路上的起点的而中止。另一方面,路口分割专注于从输入图像中的复杂路口区域提供统一且准确的

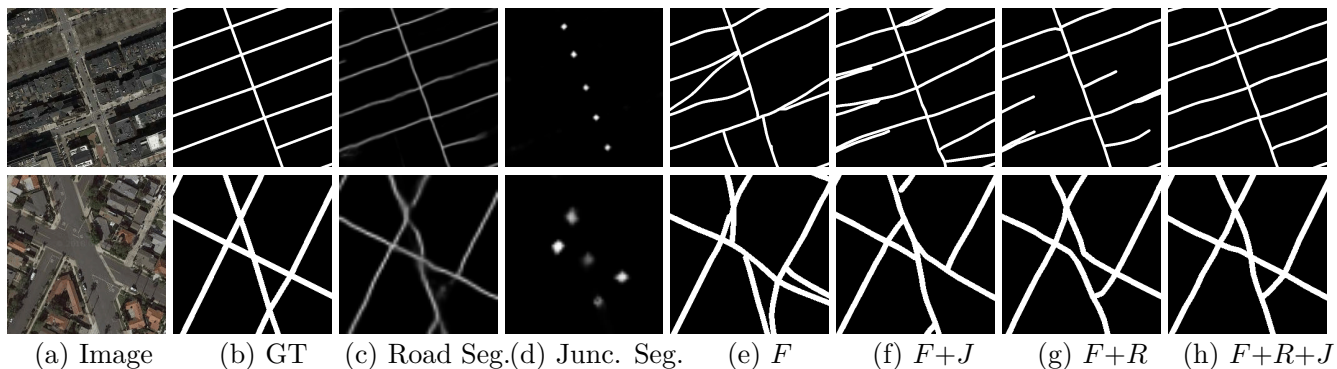


图 7. 带有分割提示指导的道路图的可视化。我们用  $F$  表示使用可变步长,  $R$  表示使用道路分割线索。(a) 航空图像, (b) ground-truth 分割结果, (c) 预测道路分割, (d) 预测路口分割, (e) 使用  $F$  的图, (f) 使用  $F+J$  的图, (g) 使用  $F+R$  的图, (h) 使用  $F+R+J$  的图。

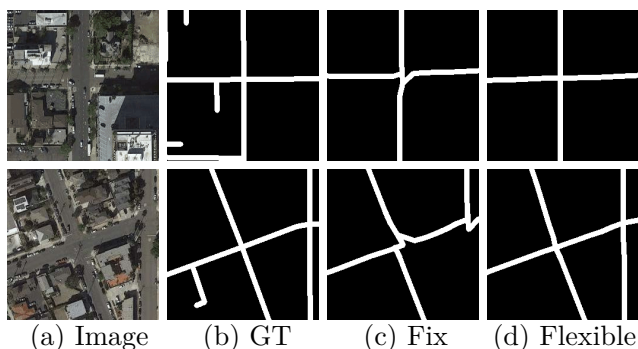


图 8. 固定步长和灵活步长生成的图的可视化。(a) 航空图像, (b) ground-truth 分割结果, (c) 固定步长预测图, (d) 灵活步长预测图。

Flexible Step	Road Seg.	Junc Seg.	Traj. Exp.	P-F1	J-F1	APLS
				53.28	36.25	34.69
✓				56.42	43.83	46.22
✓	✓			68.28	56.21	49.46
✓		✓		61.28	55.49	50.75
✓	✓	✓		69.81	59.42	57.28
✓	✓	✓	✓	73.69	62.36	61.14

表 2. 我们提出的方法的逐步改进。请注意, 所有实验共享从完整模型中获得的相同起点集。

路口, 这有助于现有路线图正确链接到新路口。它们都有助于增强路线图中的连通性。

实验结果表明, 这三个评估指标均具有改进的性能。我们基于点的迭代探索方案自然可以使我们以统一和兼容的方式利用分割提示。它还使神经网络输出具有良好的对齐和连接性。

**轨迹探索.** 我们将轨迹探索方案与采用灵活步长和两种分割提示但没有递归机制并且直接输出多达  $T$  个通道的“顺序步进”点估计的模型进行比较。

我们在直接前向模型中设置  $T = 1$ , 并记录其路口度量值。我们在直接前向模型的训练阶段设置  $T = 4$ , 并通过分别使用输出的前 1-4 个通道构建最终图来四次评估路口得分。我们还仅通过使用前 1-4 次递归输出在 RP-Net 上四次评估路口得分。

结果展示在 Tab. 3 中, 表明了递归方案有助于进一步改善我们的网络性能。

**后处理和完整的 RP-Net 方案对比.** 值得注意的是, 我们的网络还允许通过在解码器上执行分段监督来恢复道路分割, 而无需其他监督。具有航空图像的原始尺寸的分割输出可以通过常规的后处理技术来跟随, 以生成路线图。我们评估了此后处理方案的性能, 并将其与我们的完整 RP-Net 方案进行比较。

与其他基于分割的方法一样 [16, 15], 我们使用形态学细化和 RDP 算法生成道路分割, 以生成图。之后, 应用了诸如短边修剪和消除小孔 [16] 的技术。评估分数以“RP-Net-Seg.”线的形式记录在 Tab. 1 中, 它们比使用迭代 RP-Net 更低, 尤其是在路口度量和 APLS 度量上。在 Fig. 6 (a) 和 (b) 中也给出了定性比较, 其中模糊的分割会导致路线图中断。使用我们设计的完整网络结构, RP-Net 方案显示了出更好的连接性和一致性。

Traj. Exp.	Train $T$	Test $T$	J-F1
×	1	1	59.19
×	4	1	60.71
×	4	2	60.69
×	4	3	60.70
×	4	4	60.44
✓	4	1	60.66
✓	4	2	61.60
✓	4	3	61.92
✓	4	4	62.36

表 3. 轨迹探索消融研究。Train  $T$  表示监督通道数量。Test  $T$  表示推断时仅使用输出的前  $T$  个通道进行图形构建。我们根据“J-F1”分数评估在轨迹探索中采用的不同步数的性能。

## 5. 总结

在本文中，我们提出了使用灵活步长和分割线素的基于点的迭代图探索方法。在各种指标上进行的实验表明，与 RoadTracer 数据集上的最新方法相比，我们的方法在路线图对齐和连通性方面提供了显著改进。将来，我们计划进一步研究轨迹探索，并研究在探索中进行全局 [28] 优化的可能性。我们的高质量道路检测还可以作为检测航空图像中的离散目标的先验知识 [9, 12]。

## 致谢

这项研究得到了新一代 AI 重大项目的资助，项目编号为 2018AAA0100400，NSFC (61922046) 和天津自然科学基金 (17JCJQJC43700)。

## 参考文献

[1] F. Bastani, S. He, S. Abbar, M. Alizadeh, H. Balakrishnan, S. Chawla, S. Madden, and D. DeWitt. Road-tracer: Automatic extraction of road networks from aerial images. In CVPR, 2018. 1, 3, 4, 5, 6, 7

[2] A. Batra, S. Singh, G. Pang, S. Basu, C. Jawahar, and M. Paluri. Improved road connectivity by joint learning of orientation and segmentation. In CVPR, pages 10385–10393, 2019. 1, 3, 6, 7

[3] D. Chai, W. Forstner, and F. Lafarge. Recovering line-networks in images by junction-point processes. In CVPR, pages 1894–1901, 2013. 2

[4] A. Chaurasia and E. Culurciello. Linknet: Exploiting encoder representations for efficient semantic segmentation. In IEEE Visual Commun. Image Process., 2017. 2

[5] 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. Rem. S., 55(6):3322–3337, 2017. 2, 4

[6] S. Das, T. Mirnalinee, and K. Varghese. Use of salient features for the design of a multistage framework to extract roads from high-resolution multispectral satellite images. IEEE Trans. Geosci. Rem. S., 49(10):3906–3931, 2011. 2

[7] D. H. Douglas and T. K. Peucker. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. Cartographica: Int. J. Geog. Inform. Geovisualization, 10(2):112–122, 1973. 1, 2

[8] 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 TPAMI, 2020. 7

[9] L. Han, P. Tao, and R. R. Martin. Livestock detection in aerial images using a fully convolutional network. Computational Visual Media, 5(2):221–228, 2019. 9

[10] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In CVPR, pages 770–778, 2016. 2

[11] S. Hinz and A. Baumgartner. Automatic extraction of urban road networks from multi-view aerial imagery. ISPRS J. Photogramm. Rem. S., 58(1-2):83–98, 2003. 2

[12] Q. Hou, M.-M. Cheng, X. Hu, A. Borji, Z. Tu, and P. Torr. Deeply supervised salient object detection with short connections. IEEE TPAMI, 41(4):815–828, 2019. 9

[13] D. Kinga and J. B. Adam. A method for stochastic optimization. In ICLR, volume 5, 2015. 5

[14] Z. Li, J. D. Wegner, and A. Lucchi. Topological map extraction from overhead images. In ICCV, 2019. 3

[15] Y. Liu, M.-M. Cheng, X. Hu, J.-W. Bian, L. Zhang, X. Bai, and J. Tang. Richer convolutional features for

- edge detection. *IEEE TPAMI*, 41(8):1939–1946, 2019. [8](#)
- [16] G. Mátyus, W. Luo, and R. Urtasun. Deeproadmapper: Extracting road topology from aerial images. In *ICCV*, 2017. [1](#), [2](#), [6](#), [7](#), [8](#)
- [17] V. Mnih and G. E. Hinton. Learning to detect roads in high-resolution aerial images. In *ECCV*, pages 210–223. Springer, 2010. [2](#), [6](#)
- [18] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, pages 234–241, 2015. [2](#)
- [19] S. Saito, T. Yamashita, and Y. Aoki. Multiple object extraction from aerial imagery with convolutional neural networks. *Electronic Imaging*, 2016(10):1–9, 2016. [1](#), [2](#), [4](#), [6](#)
- [20] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014. [5](#)
- [21] M. Song and D. Civco. Road extraction using svm and image segmentation. *Photogramm. Eng. Rem. S.*, 70(12):1365–1371, 2004. [2](#)
- [22] R. Stoica, X. Descombes, and J. Zerubia. A gibbs point process for road extraction from remotely sensed images. *IJCV*, 57(2):121–136, 2004. [2](#)
- [23] E. Türetken, F. Benmansour, and P. Fua. Automated reconstruction of tree structures using path classifiers and mixed integer programming. In *CVPR*, pages 566–573. IEEE, 2012. [2](#)
- [24] A. Van Etten, D. Lindenbaum, and T. M. Bacastow. Spacenet: A remote sensing dataset and challenge series. *arXiv preprint arXiv:1807.01232*, 2018. [3](#), [6](#)
- [25] C. Ventura, J. Pont-Tuset, S. Caelles, K.-K. Maninis, and L. Van Gool. Iterative deep learning for road topology extraction. *BMVC*, 2018. [1](#)
- [26] J. D. Wegner, J. A. Montoya-Zegarra, and K. Schindler. A higher-order crf model for road network extraction. In *CVPR*, pages 1698–1705, 2013. [2](#)
- [27] J. D. Wegner, J. A. Montoya-Zegarra, and K. Schindler. Road networks as collections of minimum cost paths. *ISPRS J. Photogramm. Rem. S.*, 108:128–137, 2015. [2](#)
- [28] H. Wu, X. Lyu, and Z. Wen. Automatic texture exemplar extraction based on global and local textureness measures. *Computational Visual Media*, 4(2):173–184, 2018. [9](#)
- [29] F. Yu and V. Koltun. Multi-scale context aggregation by dilated convolutions. In *ICLR*, 2016. [2](#)
- [30] T. Zhang and C. Y. Suen. A fast parallel algorithm for thinning digital patterns. *Commun. ACM*, 27(3):236–239, 1984. [1](#)
- [31] Z. Zhang, Q. Liu, and Y. Wang. Road extraction by deep residual u-net. *IEEE Geosci. Rem. S. L.*, 2018. [2](#)
- [32] W. Zhen, S. Yao, and J. Lin. Learning adaptive receptive fields for deep image parsing networks. *Computational Visual Media*, 4(3):1–14, 2018. [7](#)
- [33] L. Zhou, C. Zhang, and M. Wu. D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. In *CVPR*, pages 182–186, 2018. [2](#)