

多分辨率线段提取方法及线段语义分析

宋 欣^{1,2} 王正玑^{1,3} 程章林¹

¹(中国科学院深圳先进技术研究院 深圳 518055)

²(中国科学院大学深圳先进技术学院 深圳 518055)

³(中国科学技术大学软件学院 苏州 215123)

摘 要 线段是一种组成几何体的基本元素, 蕴含着非常丰富的几何信息。从图像中提取完整、连续且具有语义信息的线段对恢复场景的几何结构具有重要意义。该文提出了一种多分辨率线段提取方法, 并对线段进行语义分析以区分轮廓线段和纹理线段。该方法首先运用多分辨率思想进行线段提取, 然后结合深度神经网络技术对线段进行语义分析, 最后对线段进行聚类合并得到最终结果。在线段连续性和完整性方面, 该文提出的方法与当前常用的线段提取方法相比具有明显优势; 在语义分析准确性方面, 该文提出的方法在测试集上的像素精度高达 97.82%。

关键词 线段提取; 语义分析; 多分辨率

中图分类号 TG 156 文献标志码 A

On the Extraction of Line Segments with Multi-Resolution and Their Semantic Analysis

SONG Xin^{1,2} WANG Zhengji^{1,3} CHENG Zhanglin¹

¹(Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China)

²(Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen 518055, China)

³(School of Software Engineering, University of Science and Technology of China, Suzhou 215123, China)

Abstract Line segment is the essential element of geometry objects, which contains very rich geometric information. Extracting complete and continuous line segments with semantic information from an image is of great significance for restoring the geometry structure of a scene, yet challenging. This paper proposes a multi-resolution segment extraction approach, which performs semantic analysis on the line segments to distinguish the contour and the texture line segments. This approach first extracts line segments with multi-resolution thought, then combines the deep neural network technology to perform semantic analysis on line segments, and finally clusters the line segments to get the final result. In terms of line segment continuity and integrity, the proposed approach has obvious advantages compared with the commonly used line segment extraction methods. In terms of semantic analysis accuracy, the pixel accuracy of the proposed approach on the test set is achieves 97.82%.

Keywords line segment extraction; semantic analysis; multi-resolution

收稿日期: 2018-03-16 修回日期: 2018-04-10

基金项目: 国家自然科学基金项目(61502471); 863 项目(2015AA016401); 中国科学院国际合作项目(GJHZ1862)

作者简介: 宋欣, 硕士研究生, 研究方向为计算机图形学; 王正玑, 硕士研究生, 研究方向为机器学习; 程章林(通讯作者), 博士, 副研究员, 研究方向为可视计算, E-mail: zl.cheng@siat.ac.cn.

1 引 言

线段是一种常见的几何元素, 在一些计算机辅助设计软件中常常作为基本的几何元素用于人造物体的设计。这使得生活中的一些人造物体, 如工业零件、建筑、家具等都具有丰富的线段特征。因此, 线段的检测和提取在数字图像处理、计算机视觉领域具有重要意义。

与各种点特征相比, 线段特征是一种更具有全局特性的特征, 不仅具有位置信息, 还具有方向信息, 在很多情况下鲁棒性更佳。线段特征已被广泛应用于场景理解^[1-3], 几何与姿态估计^[4,5], 同步定位与地图构建^[6-8]及三维重建^[9,10]等方面。

当前的许多线段提取方法都是从图像的灰度梯度角度来考虑的。线段特征在图像上的主要特点是, 在某个方向上图像的灰度梯度较大, 而在与之垂直的方向上灰度梯度较小。线段提取比较经典的一项工作是 Von Gioi 等^[11]提出的线段检测器 (Line Segment Detector, LSD)。该方法首先通过计算图像中每个像素的灰度梯度和灰度明暗变化的分界线生成一个单位向量场, 再根据向量场中各个向量的方向将向量场分割成不同的连通区域, 即在同一个连通区域中的向量方向具有一致性。然后, 在这些连通区域上拟合出候选线段。最后, 根据 Helmholtz 原则^[12]来剔除错误检测。该方法能够得到亚像素精度的结果, 并且能够很好地控制错误检测的数量。

除了直接在原始图像上提取线段外, 还有一些线段提取方法通过先从原始图像上提取一张边缘图 (Edge Map), 然后在边缘图上来提取线段^[13,14]。基于统计的霍夫变换 (Hough Transform, HT) 是在边缘图上提取线段常用的方法, 但 HT 所提取到的是满足有足够多的边缘点支持的无限长直线, 需要将其分割后才能得到线段。这样不仅容易造成过多的错误检测, 还使提取到的线段非常细碎、不完整和不连续。针对以上问题,

2015 年, Lu 等^[14]提出了 CannyLines 方法, 该方法使用了自适应参数的 Canny 算子来提取边缘图, 并加入了线段扩展和合并的操作, 提高了线段的连续性和完整性。

虽然线段提取方法的效果已越来越好, 但从本质上说, 无论是在原始图像上提取线段还是在边缘图上提取线段, 都是基于局部梯度信息, 均存在以下两个问题: (1) 由于光照不均或者遮挡等因素的影响, 线段提取的结果往往不完整、不连续; (2) 不能区分不同语义的线段。传统的基于梯度的线段提取方法认为图像灰度梯度大的地方就是线段, 但图像上灰度梯度大却有着两种不同的语义。其中, 第一种是由于纹理或者光照等因素造成的颜色突变, 第二种是不同物体的轮廓。在本文中, 分别称这两种不同语义的线段为纹理线段和轮廓线段。

以上两个问题的存在, 使得传统基于梯度的线段提取方法不能满足某些应用的需求, 如立体匹配和三维重建等。完整连续的线段能够保证立体匹配的效果, 语义信息则可以加速线段匹配的速度, 并且能够减少错误匹配的数量。因此, 区分两种不同语义的线段在基于线段的三维重建中具有重要的意义。尽管这两种不同语义的线段在图像上的表征十分相似, 但在三维空间中却有着不同的几何意义。其中, 由纹理和光照等因素产生的线段常常分布在一个平面多边形的内部, 而物体轮廓的线段则起着支撑物体形状的作用。这两种不同语义的线段之间是一种非常复杂的关系, 很难通过设计一些简单的规则来进行区分。而神经网络具有非常强大的描述能力, 能够表达非常复杂的关系。

因此, 本文提出了一种线段提取方法, 该方法在高分辨率的图像上进行线段提取, 提高了线段的完整性和连续性, 并且结合神经网络技术对线段进行语义上的区分。该方法的主要步骤如下: 首先, 在不同分辨率下提取线段; 其次,

使用深度神经网络分析线段语义; 最后, 对线段进行聚类合并。该方法具有以下两个创新点:

(1) 采用多分辨率提取线段再聚类合并的方法提高了线段的连续性和完整性; (2) 将传统基于梯度的线段提取方法与深度学习相结合, 分析不同线段的语义。

2 多分辨率线段提取

图像上线段特征的提取结果与分辨率有很大关系。一般来说, 在高分辨率下提取线段的结果更加精确, 但由于高频噪声及纹理的影响, 线段的连续性和完整性不够好。而在低分辨率下提取线段则会损失一部分精度, 但由于高频噪声及纹理的减少, 提取的线段会更加完整。因此, 在高分辨率和低分辨率下提取线段有不同的优缺点。为了发挥不同分辨率下提取线段的优势, 保证提取线段的连续性和完整性, 本文将在多个分辨率

下提取线段, 并在后面的步骤中对不同分辨率下提取的线段进行聚类合并。

本文采用 LSD 方法^[11]来提取特定分辨率下的线段。其中, LSD 方法提取的线段为矢量形式, 用线段的起点和终点坐标表示, 如线段 $L_i = s_i e_i$ (s_i 、 e_i 分别为线段的起点和终点)。分辨率的设置根据具体情况调整, 一般来说设置为 2~3 个分辨率。分辨率的范围为原始图像分辨率的 0.5~2 倍。其中, 超过 1 倍分辨率的图像通过对原始图像进行上采样得到, 上采样过程中使用 3 次样条插值方法进行插值。如图 1 所示, 在一些实例中, 从原始图像上无法提取到线段, 但在上采样后的图像上却可以提取到线段。而低于 1 倍分辨率的图像则是通过对原始图像进行高斯滤波和下采样得到。过低的分辨率将会带来较大的误差, 过高的分辨率则会增加计算复杂度。由图 2 可以看到, 线段的完整性随着分辨率的降低而提高, 但精度却在下降。

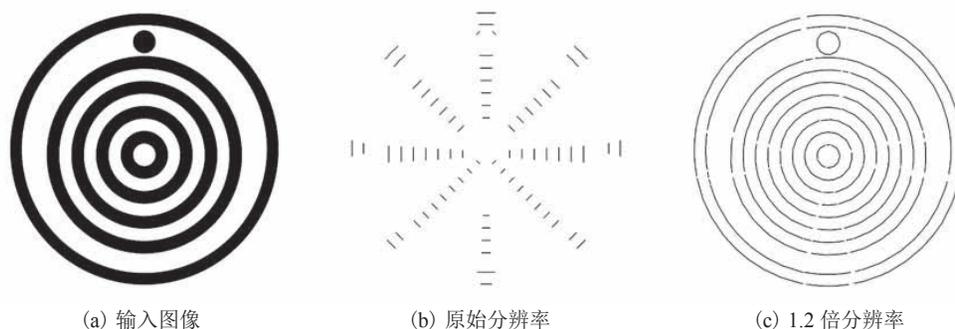


图 1 上采样对线段提取结果的影响

Fig. 1 Effect of upsampling on line segment extraction results

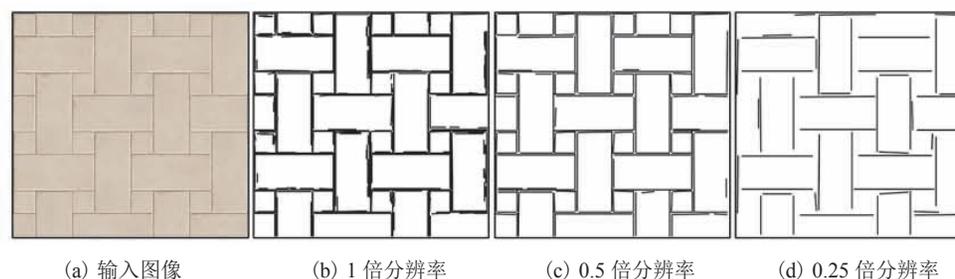


图 2 不同分辨率下线段提取的结果

Fig. 2 The results of line segment extraction at different resolutions

3 基于神经网络的线段语义分析

单层的神经网络也称作感知机，是最简单的神经网络。由于感知机只能处理线性可分的问题，对于很多复杂的问题需要使用多层神经网络来解决。多层神经网络可以看作是多个感知机单元的组合。其中，前端第一层叫做输入层，中间的叫隐含层，末端的叫输出层。在 2006 年，Hinton 和 Salakhutdinov^[15]发现具有多个隐含层的神经网络学习特征的能力更强，并且可以通过逐层预训练的方法来缓解训练复杂性，自此越来越多的科研人员投入到深度神经网络的研究当中。

卷积神经网络在计算机视觉领域应用非常广泛。与普通的神经网络相比，卷积神经网络使用了大量的卷积层。1990 年，Le Cun 等^[16]提出的 LeNet-5 多层卷积神经网络在手写字体识别上取得了非常高的准确率。2012 年，Krizhevsky 等^[17]提出的 AlexNet 卷积神经网络在图像分类竞赛中取得了最低的错误率，并远远超过第二名的成绩。2015 年，Long 等^[18]提出了全卷积神经网络 (Fully Convolutional Network, FCN)，能够接受任意大小的图像作为输入，且输出结果也是一张图像，大小与输入图像相关。由于全卷积神经网络能对输入图像进行像素级别的分类和预测，故非常适合处理像素级别图像语义分割的任务。

线段的语义与图像上像素的语义有紧密关系，可以由线段周围的像素语义得到线段的语义。而获取图像上像素的语义是一个像素级别的图像语义分割问题，即图像上所有的像素可以分为两类：物体轮廓边缘的像素和其他像素。这两种像素构成了一副轮廓图。

由于全卷积神经网络非常适合处理像素级别的语义分割问题，所以本文采用全卷积视觉几何组 (Fully Convolutional Network-Visual Geometry Group, FCN-VGG) 网络来对图像进行语义分割。FCN-VGG 网络由视觉几何组 (Visual

Geometry Group, VGG) 网络^[19]进行结构调整得到。与 VGG 网络相比，FCN-VGG 去掉了最后的 3 个全连接层，但加上了反卷积层。

深度神经网络的精度和泛化能力与训练样本的数量有很大关系，但这样大量的样本很难获取且标注成本很高。因此，本文首先采用 VGG 网络在 ImageNet 数据集上训练得到的权值作为 FCN-VGG 网络特征提取部分的初始权值；然后，使用手工标注的 219 张图像进行调优。这样不仅节省了大量的训练时间，还节约了数据标注的成本。图 3 为数据集标注的实例。其中，图 3 (a) 为输入图像；图 3 (b) 为标注的轮廓图，轮廓图中白色像素表示轮廓像素，黑色像素表示其他像素。

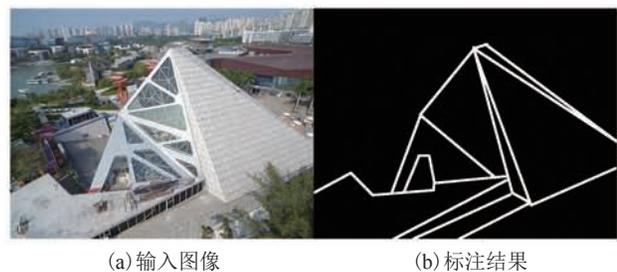


图 3 图像标注实例

Fig. 3 Image annotation example

在训练过程中，将训练集中的 219 张图像随机分为两部分。其中，154 张为训练集，65 张为测试集。本文采用 TensorFlow 框架对 FCN-VGG 网络进行训练，每次训练图片的 Batch 大小为 4，采用随机梯度下降优化算法进行优化。在训练过程中，训练集上的损失和测试集上的损失如图 4 所示。由图 4 可以看到，随着迭代次数的增加，训练集上的损失逐渐变小，测试集上的损失则先逐渐变小，但在 3 000 次迭代后呈现出增长趋势，说明此时神经网络模型存在过拟合的风险。因此，在后面的步骤中采用 3 000 次迭代的神经网络模型进行预测。

由于 FCN-VGG 卷积神经网络得到的是一副带有语义信息的图像，即轮廓图，需要将图像中

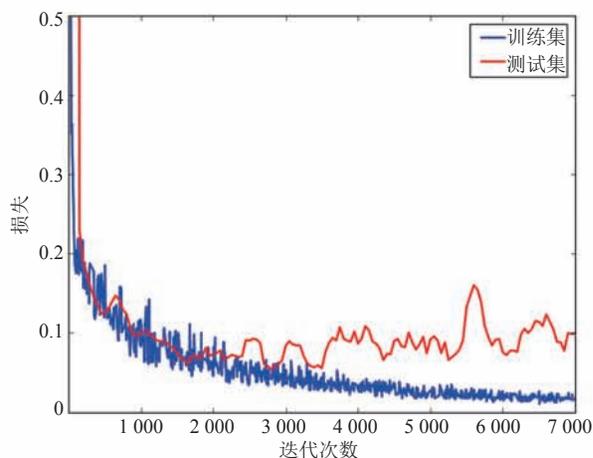


图 4 损失曲线

Fig. 4 Loss curve

的语义信息融合到线段中。如图 5 所示, 对于线段 L_i , 首先从起点开始到终点结束均匀地选取 m 个采样点, 分别为 p_1, p_2, \dots, p_m ; 然后, 统计落在轮廓边缘区域内的点的数量 n 。最后, 若 n/m 大于阈值 r , 则线段 L_i 为轮廓边线段, 否则为纹理线段。

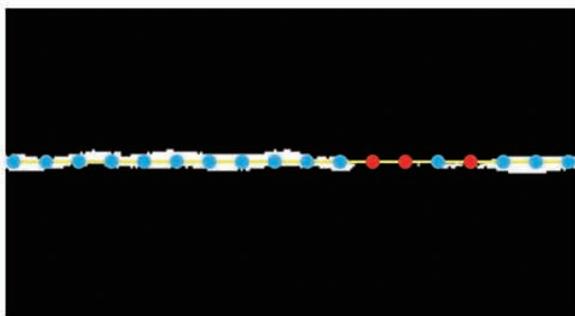


图 5 从图像中获取线段语义

Fig. 5 Get segment semantics from the image

4 线段聚类及合并

为了保证线段的连续性和完整性, 需要对一些线段进行连接、合并。本文采用聚类的方法来决定哪些线段应当合并到一起。对于聚类问题, 两个非常关键的因素需要考虑, 一是距离的定义, 二是聚类算法的选择。如图 6 所示, 通过观察发现影响线段是否应该合并的因素有 3 个:

(1) 线段之间夹角的大小; (2) 线段之间垂直间距的大小; (3) 线段之间平行间距的大小。对于线段之间的距离定义, 本文参考了 Lee 等^[20]提出的轨迹聚类算法中使用的轨迹距离, 并加以修改使之满足线段聚类的要求。图 7 为本文线段之间距离定义的示意图。本文定义的线段之间的距离包含 3 个部分: 角度距离 d_θ 、平行距离 d_p 和垂直距离 d_v 。



图 6 影响线段合并的因素

Fig. 6 Factors affecting line segment consolidation

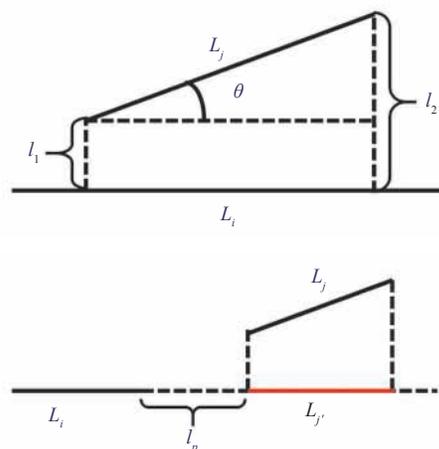


图 7 线段距离定义示意图

Fig. 7 Segment distance definition

角度距离的定义如公式(1)所示。其中, θ 是线段 L_i 和 L_j 之间的夹角; $\|L_j\|$ 表示线段 L_j 的长度。本文中, 线段的聚类不考虑线段方向的影响, 即线段 $e_i s_j$ 和线段 $s_i e_j$ 表示同一条线段。

$$d_\theta = \|L_j\| \sin(\theta) \quad (1)$$

垂直距离的定义如公式(2)所示。假设 L_i 的长度比 L_j 长, 则 l_1 和 l_2 为线段 L_j 的两个端点到 L_i 所在直线的距离。线段 L_i 和 L_j 之间的垂直距离定义为 l_1 和 l_2 的 2 阶 Lehmer 平均数。

$$d_v = \frac{l_1^2 + l_2^2}{l_1 + l_2} \quad (2)$$

平行距离的定义如公式(3)所示。假设 L_i 的长度比 L_j 长, 将 L_j 向 L_i 所在的直线方向上投影得到线段 $L_{j'}$ 。线段 L_i 和 L_j 之间的平行距离为 L_i 与 $L_{j'}$ 之间的最小距离 l_p 。当 L_i 与 $L_{j'}$ 相互交错时, L_i 与 L_j 之间的平行距离为 0。

$$d_p = \min(l_p, 0) \quad (3)$$

最终两条线段之间的距离为以上 3 部分的加权和。

$$d(L_i, L_j) = \alpha d_o + \beta d_v + \gamma d_p \quad (4)$$

其中, α 、 β 、 γ 为各项距离的权重, 且满足以下两个条件。

$$\begin{aligned} 0 < \alpha, \beta, \gamma < 1 \\ \alpha + \beta + \gamma = 1 \end{aligned} \quad (5)$$

定义好线段之间的距离后需要确定聚类算法, 本文采用基于密度的有噪声应用空间聚类 (Density-Based Spatial Clustering of Applications with Noise, DBSCAN)^[21] 算法对线段进行聚类。DBSCAN 算法需要两个参数, 分别是计算局部密度的半径 r 和局部密度阈值 n 。这两个参数很容易根据具体应用情况进行调节。此外, DBSCAN 算法具有抗噪声的功能。因此, DBSCAN 算法非常适合线段的聚类。

聚类完成后需要对同一类的线段进行合并。虽然同一类的线段可能来自不同分辨率的图像, 但从高分辨率图像上提取的线段精度更高, 故合并时选择该类中从最高分辨率的图像中提取的线段进行合并。对于给定的需要合并的线段集合 $C = \{L_1, L_2, \dots, L_n\}$, 首先需要计算这些线段的几何中心 o 和平均方向 v 。其中, 计算平均方向时需要保证集合中的线段方向一致, 即对任意的两条线段 L_i 和 L_j , 其方向向量的内积要为非负值。在具体计算过程中, 可以将所有的线段和第一条线段 L_1 的方向保持一致即可。然后, 计算合并后的线段。具体计算过程请参考算法 1。

如图 8 所示, 几何中心 o 和方向 v 可以确定合并后的线段所在的直线 l , 将集合 C 中每一条

算法 1 线段合并算法

输入: 待合并线段集合 $C = \{L_1, L_2, \dots, L_n\}$, 对于任意的 $L_k \in C$ 有 $L_k = s_k e_k$

输出: 合并后线段 L

```

 $o = s_1 + e_1$ 
 $v = \overline{s_1 e_1}$ 
for  $k = 2, 3, \dots, n$  do
     $o = o + s_k + e_k$ 
    if  $v \cdot \overline{s_k e_k} < 0$  then
         $v = \overline{v - s_k e_k}$ 
    else
         $v = \overline{v + s_k e_k}$ 
    end if
end for
 $o = o / 2n$ 
 $v = v / \|v\|$ 
 $t_{\max} = -\infty$ 
 $t_{\min} = \infty$ 
for  $k = 1, 2, \dots, n$  do
     $t_s = \overline{o s_k} \cdot v$ 
     $t_e = \overline{o e_k} \cdot v$ 
     $t_{\max} = \max(t_{\max}, t_s, t_e)$ 
     $t_{\min} = \min(t_{\min}, t_s, t_e)$ 
end for
 $s = o + t_{\min} v$ 
 $e = o + t_{\max} v$ 
 $L = se$ 

```

线段的端点向直线 l 上投影, 则这些投影点分布在点 o 的两侧。记在点 o 两侧距离 o 点最远的两个投影点为 s 和 e , 即为合并后的线段的起点和终点。图 8 中, 黄色线段 L 为合并后的线段。

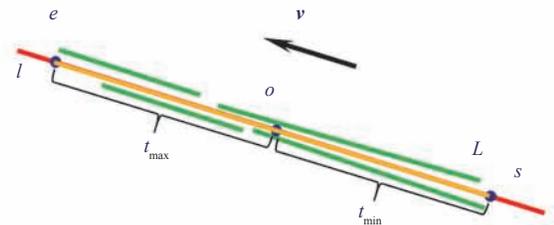


图 8 线段合并示意图

Fig. 8 Segment consolidation

5 实验结果

与其他传统的线段提取方法相比, 本文提出

的线段提取方法主要特点为: 在多分辨率的图像下进行线段提取, 并进行语义分析以区分边缘轮廓线段和纹理线段, 在此基础上对线段进行了聚类合并, 提高了线段的完整性和连续性。因此, 本文将从两个方面分析实验结果: (1) 线段的完整性和连续性; (2) 线段语义分析的准确性。

5.1 线段的完整性和连续性

为了评估提取线段的完整性和连续性, 本文选取了几个不同的实例进行测试, 并与相关工作进行比较。由于参与比较的 CannyLine^[14]和 LSD^[11]线段提取方法没有对线段进行语义分析,

所以在进行线段完整性和连续性评估时不考虑线段的语义区别。不同方法提取线段的结果如图 9 所示。其中, 9(a) 为输入图片, 9(b)~9(d) 分别为 CannyLine、LSD 和本文方法提取的线段。通过观察发现, 本文方法提取的线段完整性和连续性最好, 断裂线段的数量最少, 视觉效果最好。

为了定量评估线段的完整性和连续性, 本文选取了 10 张真实场景的图片进行测试。其中, 5 张为室内场景, 5 张为室外场景。并对这 10 个实例中不同方法提取到的线段平均长度进行统计, 结果如表 1 所示。从表 1 的数据可以看到, 在

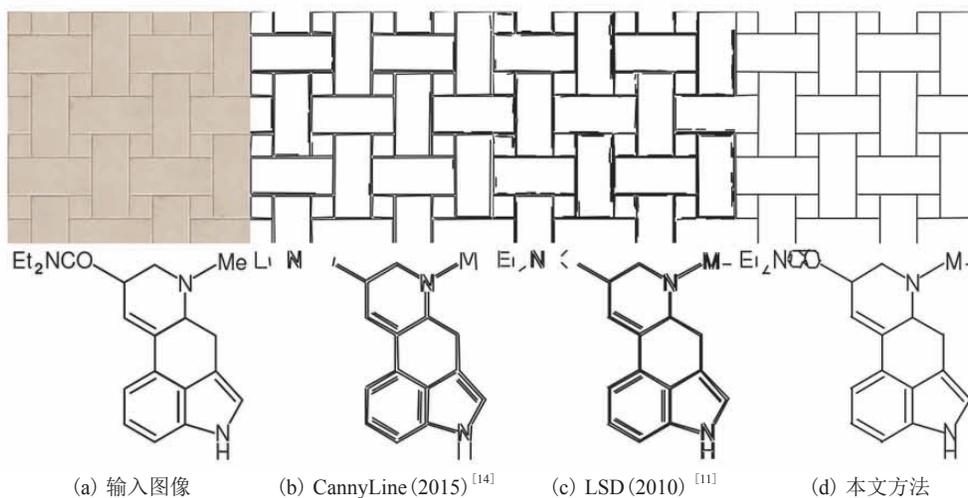


图 9 不同方法提取线段的结果

Fig. 9 Results of different line segment extraction method

表 1 不同方法线段提取结果的完整性比较

Table 1 Integrity comparison of different methods

场景	图片编号	CannyLine ^[14]		LSD ^[11]		本文方法	
		线段条数	平均长度	线段条数	平均长度	线段条数	平均长度
室内场景	1	641	41.40	1 154	32.08	316	67.39
	2	393	50.62	717	36.42	298	59.44
	3	442	39.90	595	30.24	272	50.14
	4	201	71.21	278	48.04	149	82.22
	5	486	37.11	933	22.95	337	46.57
室外场景	1	608	32.22	1 443	24.49	334	46.32
	2	456	32.00	568	21.93	306	42.03
	3	252	34.67	327	24.52	117	49.25
	4	316	37.16	604	24.07	312	38.58
	5	301	40.22	477	30.55	180	50.90

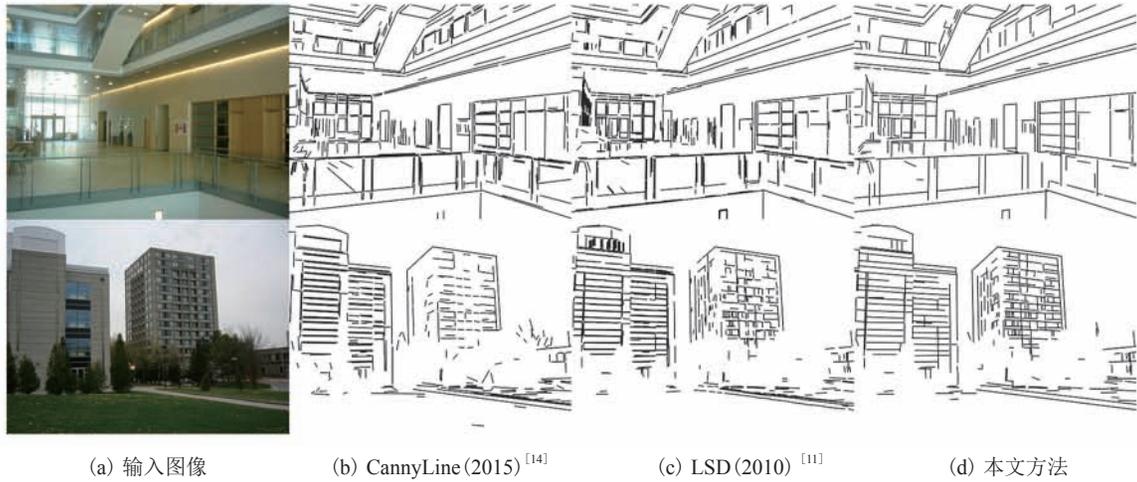


图 10 室内和室外场景中不同线段提取方法的结果

Fig. 10 Result of different line segment extraction methods in indoor and outdoor scene

10 个实例中，本文方法提取的线段平均长度最长。图 10 为两个具有代表性的实例的线段提取结果。

本文提出的方法可以非常灵活地满足不同应用的需求。同时，还可以通过改变线段聚类合并的参数来调整最终线段提取的结果。如图 11 所示，改变聚类半径的参数可以调整最终线段提取的边缘效果。

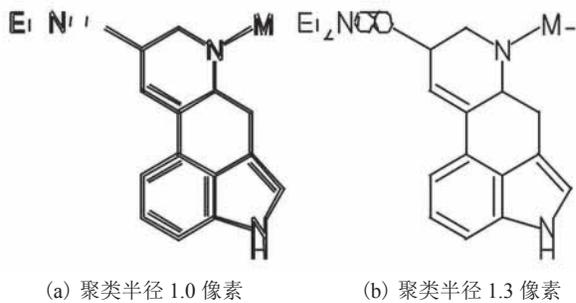


图 11 不同参数对结果的影响

Fig. 11 The effect of different parameters

5.2 语义分析的正确率

由本文第 3 部分可知，线段的语义来源于图像上像素的语义，而获取图像上像素的语义是一个二类语义分割问题。因此，采用评估图像语义分割常用的像素精度、平均精度和平均交并比 3 个指标对本文语义分析的正确率进行分析，结果

见表 2。设 n_{ij} 表示本身是 i 类像素且被预测为 j 类像素的像素数目， n_{cl} 为像素总类别数（本文中， $n_{cl}=2$ ）， t_i 表示 i 类像素的总数目，即 $t_i = \sum_j n_{ij}$ ，那么以上 3 个指标的定义如下。

$$(1) \text{ 像素精度: } \frac{\sum_i n_{ii}}{\sum_i t_i}, \text{ 表示预测正确的像素占总像素的比例。}$$

$$(2) \text{ 平均精度: } \frac{1}{n_{cl}} \sum_i \frac{n_{ii}}{t_i}, \text{ 表示各类像素预测正确率的平均值。}$$

$$(3) \text{ 平均交并比: } \frac{1}{n_{cl}} \sum_i \frac{n_{ii}}{t_i + \sum_j n_{ji} - n_{ii}}, \text{ 表示预测结果和真实结果的交集与并集比值的平均值。}$$

由表 2 可以看出，本文所采用的图像语义分割方法具有非常高的精度，其中像素精度高达 97.82%，能够满足线段语义分析的要求。

表 2 图像语义分割精度

Table 2 Image semantic segmentation accuracy

像素精度(%)	平均精度(%)	平均交并比(%)
97.82	87.87	81.97

图 12 为测试集中 3 个线段语义分析的实例。其中，每一行表示一个实例，从左到右分别

是输入的图像实例、人工标注的真实结果、深度神经网络预测的结果和将语义信息融合到线段上的结果。图中红色线段表示轮廓线段, 黑色线段表示纹理线段。

为了测试神经网络的泛化能力, 本文收集了一些不同风格的场景图像(这些图像既不属于训练集也不属于测试集), 并使用上述神经网络对其进行预测, 结果如图 13 所示。从图 13 可以

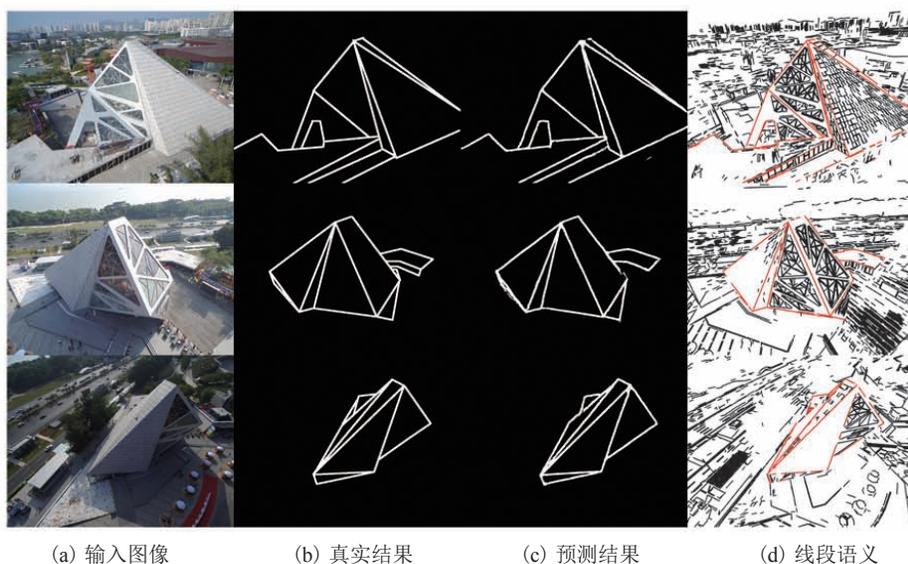


图 12 线段语义预测结果

Fig. 12 Result of line segment semantic prediction

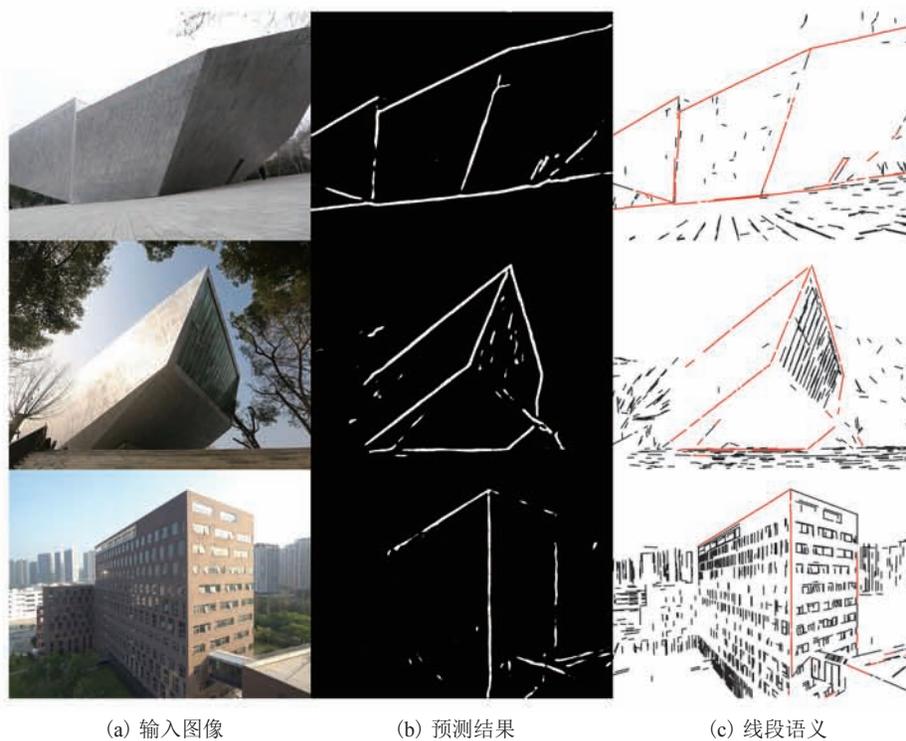


图 13 不同风格场景预测结果

Fig. 13 Prediction results of different style scenes

看出, 与训练集中风格越相近的场景预测结果越好。这表明该神经网络具有一定的泛化能力, 但泛化能力有限。

6 结 论

本文提出了一种基于语义分析的线段提取方法, 并使用深度神经网络对线段进行语义分析。与 CannyLine^[14] 和 LSD^[11] 方法相比, 本文方法不仅在线段完整性和连续上有明显提高, 而且还能够区分轮廓线段和纹理线段, 对基于线段的立体匹配和三维重建等应用具有较高的价值。但本文方法仍存在一定的局限性: 首先, 由于使用 LSD 方法对特定分辨率的图像进行线段提取, 所以最终线段提取的效果还是会受到 LSD 方法的局限影响; 其次, 用于训练深度神经网络的数据集较小, 还存在泛化能力不够的问题。在未来, 可以深入运用多分辨率思想, 脱离 LSD 方法的束缚, 进一步提高线段的完整性和连续性。此外, 还可以增加训练数据集, 增强神经网络的泛化能力, 提高线段语义分析的准确性, 并将线段提取工作放入卷积神经网络, 实现端对端的线段检测算法。

参 考 文 献

- [1] Zhao YB, Zhu SC. Scene parsing by integrating function, geometry and appearance models [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2013: 3119-3126.
- [2] Tretyak E, Barinova O, Kohli P, et al. Geometric image parsing in man-made environments [J]. International Journal of Computer Vision, 2012, 97(3): 305-321.
- [3] Khan SH, Bennamoun M, Sohel F, et al. Geometry driven semantic labeling of indoor scenes [C] // European Conference on Computer Vision, 2014: 679-694.
- [4] Elqursh A, Elgammal A. Line-based relative pose estimation [C] // Proceedings of the 2001 Conference on Computer Vision and Pattern Recognition, 2011: 3049-3056.
- [5] Zhang LL, Xu C, Lee KM, et al. Robust and efficient pose estimation from line correspondences [C] // Asian Conference on Computer Vision, 2012: 217-230.
- [6] Uehara, K, Saito H, Hara K. Line-based SLAM considering directional distribution of line features in an urban environment [C] // Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, 2017: 255-264.
- [7] Zhou HZ, Zou DP, Pei L, et al. Structslam: visual slam with building structure lines [J]. IEEE Transactions on Vehicular Technology, 2015, 64(4): 1364-1375.
- [8] Pumarola A, Vakhitov A, Agudo A, et al. PL-SLAM: real-time monocular visual SLAM with points and lines [C] // 2017 IEEE International Conference on Robotics and Automation (ICRA), 2017: 4503-4508.
- [9] Hofer M, Maurer M, Bischof H. Line3D: efficient 3D scene abstraction for the built environment [C] // German Conference on Pattern Recognition, 2015: 237-248.
- [10] Wang JL, Fang T, Su QK, et al. Image-based building regularization using structural linear features [J]. IEEE Transactions on Visualization and Computer Graphics, 2016, 22(6): 1760-1772.
- [11] Von Gioi RG, Jakubowicz J, Morel JM, et al. LSD: a fast line segment detector with a false detection control [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(4): 722-732.

- [12] Desolneux A, Moisan L, Morel JM. Edge detection by Helmholtz principle [J]. *Journal of Mathematical Imaging and Vision*, 2001, 14(3): 271-284.
- [13] Akinlar C, Topal C. EDLines: a real-time line segment detector with a false detection control [J]. *Pattern Recognition Letters*, 2011, 32(13): 1633-1642.
- [14] Lu XH, Yao J, Li K, et al. Cannylines: a parameter-free line segment detector [C] // 2015 IEEE International Conference on Image Processing, 2015: 507-511.
- [15] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks [J]. *Science*, 2006, 313(5786): 504-507.
- [16] Le Cun Y, Boser B, Denker JS, et al. Handwritten digit recognition with a back-propagation network [C] // *Advances in Neural Information Processing Systems*, 1990: 396-404.
- [17] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks [C] // *Proceedings of the 25th International Conference on Neural Information Processing Systems*, 2012: 1097-1105.
- [18] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [C] // *IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 3431-3440.
- [19] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J/OL]. arXiv:1409.1556.
- [20] Lee JG, Han JW, Whang KY. Trajectory clustering: a partition-and-group framework [C] // *Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data*, 2007: 593-604.
- [21] Ester M, Kriegel HP, Sander J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise [C] // *Proceedings of the 1996 International Conference on Knowledge Discovery and Data Mining (KDD'96)*, 1996, 96(34): 226-231.