

深度霍夫变换用于语义线检测

赵凯*, 韩琦*, 张长彬, 徐君, 程明明†

摘要—我们关注于在自然场景中检测有意义的线结构(又称语义线)这一基本任务。许多先前的方法将该问题视为目标检测的特例并调整现有的目标检测算子进行语义线检测。然而, 这些方法忽略了线的固有特性, 从而导致性能欠佳。线比其他复杂物体有更简单的几何特性, 因此可以使用几个参数将其紧凑地参数化表示。为了更好地利用线的属性, 在本文中, 我们将经典霍夫变换技术结合到了深度学习的特征表示中, 并提出了用于线检测的一站式端到端学习框架。通过使用斜率和截距对线参数化, 我们使用霍夫变换将深度特征表示转换到参数域, 然后在其中进行线检测。具体来说, 我们沿特征图平面上的候选线聚合特征, 然后将聚合的特征分配到参数域中相应的位置。因此, 在空间域中检测语义线被转换为在参数域中检测单个点的问题, 使得后处理步骤(即非最大抑制)更高效。此外, 我们的方法可以轻松提取对精确线检测至关重要的上下文线特征。除了提出的方法, 我们还设计了一种评价指标来评价线检测的质量, 并构建了用于线检测任务的大规模数据集。在我们提出的数据集和另一个公共数据集上的实验结果证明, 我们的方法优于以前的最先进技术。数据集和源代码已开源<https://mmcheng.net/dhtline/>。

Index Terms—语义线检测, 霍夫变换, CNN, 深度学习。



1 引言

在数字图像中检测语义线在计算机视觉中有悠久的历史。线结构的组织是将视觉信号转换为有用的中间概念以进行视觉解释 [3] 的前期但必不可少的步骤。尽管很多技术被提出以检测显著物体 [4], [5], [6], [7], [8] 和显著区域 [9], [10], [11], 但是很少有关于检测突出的/揭示结构的线结构的工作。最近的一项研究提出了检测突出的直线结构(称为“语义线”)的问题, 这些直线结构可以揭示自然图像的语义结构。识别语义线对于计算机图形学和计算机视觉应用至关重要, 例如摄影构图 [12], [13], 结构保留的图像处理 [14], [15], 美学评价 [16], [17], [18], [19], 车道检测 [20] 和艺术创作 [21], [22], [23], [24]。如 Fig. 1 所示, Liu 等人 [12] 提出根据黄金分割率使用“突出线”裁剪图片。检测这些“语义线”可以帮助拍出具有视觉上令人愉悦的摄影构图的图像。

霍夫变换 [25], [26] 是具有代表性的线检测技术, 最初被提出用来检测气泡室照片中的直线 [27]。由于其简单和有效性, HT 被用来检测数字图像中的线 [25], 并被 [26] 进一步扩展用于检测其他规则图形如圆和矩形。霍夫变换关键思想是将置信度从图像域投影到参数域, 然后在参数域中通过识别局

部最大响应来检测参数域中的形状。在线检测中, 图像域中的一条线可由其参数表示, 即参数空间中的斜率和截距。霍夫变换在图片中沿直线收集置信度并在参数空间中累积置信度到单个点。因此, 图像域中的线检测被转换为了检测参数空间中的峰值响应问题。基于经典霍夫变换 [28], [29], [30], [31] 的线检测算法通常检测连续的直线边缘而忽略线结构中的语义信息。此外, 这些方法对光线变化和遮挡敏感。因此, 如 Fig. 1(d) 所示这些结果通常都有噪声并且包含不相关的线 [32]。

卷积神经网络(CNN)在多种计算机视觉任务中均取得了显著成功。最近的部分研究 [34], [35] 提出了基于 CNN 的方法进行线检测。具体来说, 他们将线检测视为一种特殊的目标检测, 并且使用现有的目标检测算法, 如 faster R-CNN [36] 或 CornerNet [37], 进行线检测。受限于对线的 ROI 池化和非最大抑制, 两种方法 [34] and [35] 在运行时间方面都比较低效。此外, ROI 池化 [38] 仅沿着一条直线聚集特征, 而许多最近的研究表明丰富的上下文信息对许多任务, 如视频分类 [39] 和语义分割 [40], 是至关重要的。这一点将在 Sec. 6.6 进行验证, 其中我们通过实验验证了仅沿一条直线聚集特征会导致次优结果。

将有力的 CNNs 整合进霍夫变换对于语义线检测是一个有希望的方向。一个简单将 CNN 与霍夫变换组合的方式是使用基于 CNN 的边缘检测器进行边缘检测 [41], [42], 然后在边缘图上应用标准霍夫变换。但是, 这两个组成部分有不同的优化目标, 我们的实验表明这会导致次优结果。在本文中, 我们提出将霍夫变换以端到端的形式整合进 CNN 中, 从而使得我们方法中的每个部分都有相同的优化目标。我们的方法首先使用基于 CNN 的编码器提取像素级别特征表示, 然

本文是 IEEE TPAMI 论文 [1] 的中文翻译版。

* 前两名学生对本文贡献相等。

† M.M. Cheng is the corresponding author (cmcheng@nankai.edu.cn)。

- Kai Zhao, Qi Han, Chang-Bin Zhang, and Ming-Ming Cheng are with the TKLNDST, College of Computer Science, Nankai University, Tianjin, China, 300350.
- Jun Xu is with the School of Statistics and Data Science, Nankai University, Tianjin, China, 300071.
- A preliminary version of this work has been presented in [2].

Manuscript received August 11, 2020.

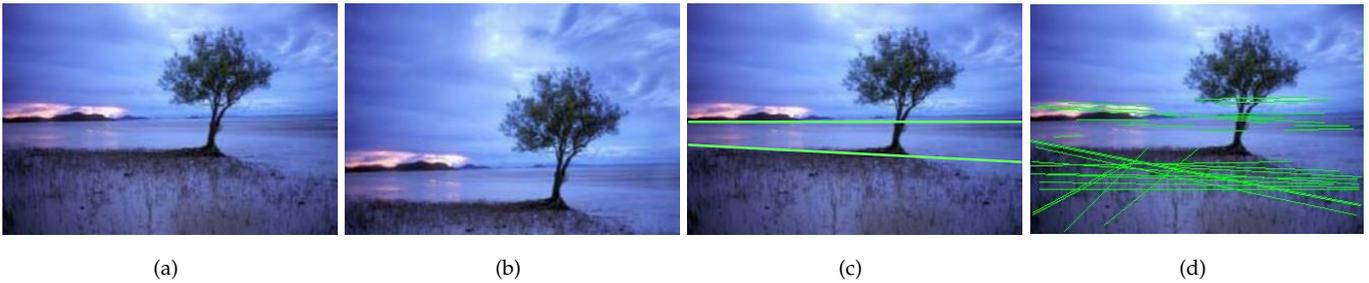


图 1. [12]中的示例图片揭示了语义线可能有助于摄影构图。(a): 任意角度拍摄的图像。(b): 符合黄金分割率原则 [21], [33]的照片, 该照片根据 [12]描述的方法使用图片中所谓的“突出线”获得。(c): 我们的检测结果是干净的, 并且仅包含对摄影构图有潜在帮助的少数有意义的线。(d): 使用经典线检测算法得到的线检测结果通常只集中在精细的细直线边缘上。

后在深度特征表示上使用霍夫变换将特征表示从特征空间中转换到参数空间中。然后全局线检测问题就转换为了在变换后的特征中简单地检测峰值响应的问题, 使得问题变得更加简单。举例来说, 耗时的非最大抑制(NMS)可被简单地替换为在参数空间中计算相邻区域的质心, 这使得我们的方法非常有效地进行实时的线检测。此外, 在检测阶段, 我们在变换后的特征上使用几个卷积层聚集附近线的上下文感知特征。因此, 最终的决策不仅取决于一条线, 也取决于周围线的信息。如图 1(c)所示, 我们的方法可以检测到干净的、有意义的和突出的线, 有助于摄影构图。

为了更好地评估线检测方法, 我们引入了合理的评价标准来评价检测到的线和对应的真实线之间的匹配程度。尽管 [34]提出了使用相交区域来评价一对线的相似性的评价指标, 但是这一度量指标可能会导致模棱两可和令人误解的结果。最后, 我们收集了6,500张精心标注的图片构建了一个大规模数据集用于语义线检测。这个新的数据集, 又称NKL (是NanKai Lines的简称), 包括多种多样的场景, 并且在图片数量和标注的线方面比现有的SEL [34]数据集规模大得多。

本文的贡献可总结为以下几点:

- 我们提出了将一个端到端框架, 将CNN的特征学习能力与霍夫变换相结合, 得到了高效的实时语义线检测方案。
- 为了促进语义线检测有关的研究, 我们使用6,500张图片构建了一个新的数据集, 其规模和多样性都优于现有的SEL数据集 [34]。
- 我们提出了合理的评价标准用以度量两条线的相似性。相比于原来的基于IOU的评价标准 [34], 我们的评价标准有直观的解释并且易于实现, 详见Sec. 4。
- 公开数据集上的评估结果表明, 我们的方法大幅度超过现有的方法。

[2]中有我们方法的初步版本。在扩展的工作中, 我们引入了以下三个主要的改进:

- 我们提出了全新的“边缘引导的微调”模块来调整线的位置, 并且在精细的边缘信息协助下获得了更好的

线检测性能。该部分详见Sec. 3.5。

- 我们提出了用于语义线检测的大规模数据集, 详见Sec. 5。新的数据集, 又称NKL(为NanKai Lines的简称), 共包括6,500张图片, 每张图片由多名经验丰富的标注人员标注。
- 评估时, 我们采用最大二部图匹配来匹配真实数据和检测到的线。匹配过程剔除了冗余的真实正确(true positives)匹配, 以便于使每一条真实的线最多与一条检测到的线相关联, 反过来亦是如此。

本文其余部分组织如下: Sec. 2总结了相关工作。Sec. 3详细说明了提出的深度霍夫变换方法。Sec. 4描述了提出的评价标准, 该标准用于检测一对线之间的相似性。Sec. 5介绍了我们新构建的数据集。Sec. 6展示了实验细节和对比结果。Sec. 7总结全文。

2 相关工作

数字图像中线检测的研究工作可以追溯到计算机视觉研究的早期阶段。在此, 我们首先简述我们用于线检测的最根本的工具之一霍夫变换 [25] (HT)的发展过程。然后, 我们介绍部分最近的基于CNN的线检测的方法。最后, 我们总结用于语义线检测的方法和数据集。

2.1 霍夫变换

霍夫变换(HT)最初被提出来用来进行气泡室照片的机器分析。它使用直线的斜率-截距进行参数化, 导致无限的变换空间(因为斜率可以是无穷大)。[25]使用角度-半径而不是斜率-截距来扩展HT, 而且从概念上讲与二维Radom变换 [43]相似。随后, Ballard等人 [26]将数字图像中HT的概念泛化至可以定位任意形状, 如椭圆和圆形。例如, 通过使用角度和半径进行参数化, 线检测可以通过在有限参数空间中票决边缘置信度和寻找峰值响应实现。通常, 随着Canny [44]和Sobel [45]等边缘检测器, 检测到的线即变换后参数空间中的最大局部响应点。最近两个工作采用了HT的核心思想, 分别是使

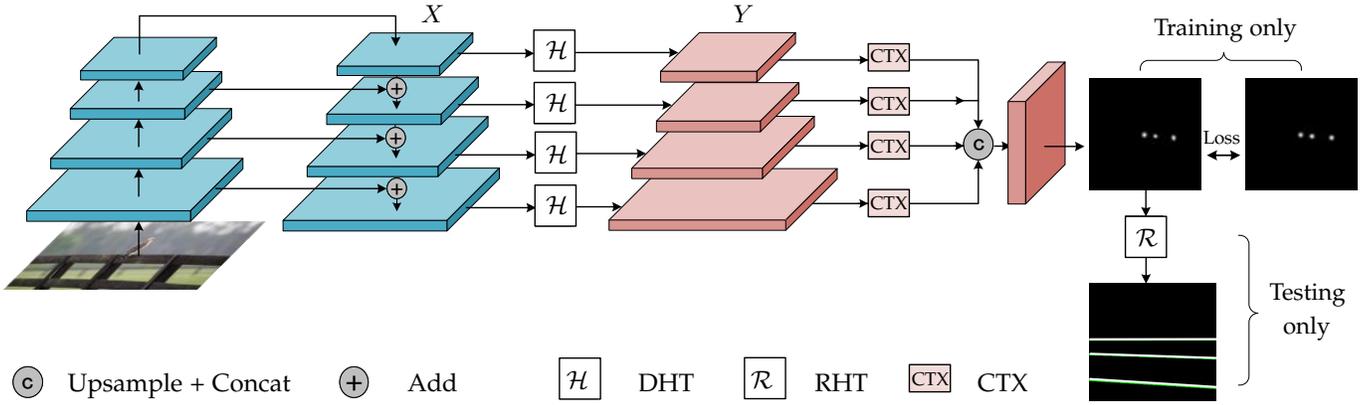


图 2. 我们提出方法的流程图。DHT为深度霍夫变换的缩写，RHT代表逆向霍夫变换。CTX为上下文敏感线检测器，包含多层卷积层。

用截距和方向参数化CNNs的输出来预测表面网格的 [46]或对三维形状进行凸分解的 [47]。

尽管HT在线检测中是成功的，但是HT仍存在较高的计算代价和不稳定的性能的缺点。为了加速HT票决，Nahum等人 [31]提出了“概率霍夫变换”在线上随机选取样本点，而 [48]使用图片的梯度方向选取票决点。同时，[28], [49]的工作采用了基于核方法的霍夫变换来进行霍夫票决，该方法在共线像素上使用椭圆高斯核来加速原始HT。John等人 [29], [30]将输入图片分为层次化的图像块，然后在这些块中独立地应用HT。Illingworth等人 [50]使用由粗到细的累积和搜索策略在霍夫参数空间中识别主要峰值。[51]在正则化框架中解决了线检测的问题，抑制了与非线性图像特征相对应的噪声和杂波的影响。霍夫票决方案在许多其他任务中也被采用，如检测点云中三维形状的重心 [52]和识别图像对应点 [53]。

2.2 线段检测

尽管鲁棒且可并行化，霍夫变换不能直接应用于线段检测，因为其不能识别线段的端点。概率霍夫变换 [31]在票决方案中使用随机采样，并通过定位样品点位置重建线段。但该方法仍偏向于检测长直线。除去霍夫变换，还有许多其他研究用于检测线段。Burns等人 [3]使用边缘方向作为引导提取线段。主要优点是梯度方向可以帮助识别低对比度的线条和端点。Etemadi等人 [54]在给定的边缘图中建立链，通过遍历这些链来提取线段。Chan等人 [55]使用量化的边缘方向来搜索和合并短的线段。Gioi等人 [56]在不调节参数的前提下提出了线性时间复杂度的线段检测器(LSD)，并被许多后续研究 [32], [57], [58]所采用。

2.3 基于CNN的线检测

最近，CNNs在计算机视觉任务中带来了显著的进步，并被应用于线检测。这些方法或者集中于直线检测，如语义线检测 [34], [59]，或者集中于线段检测，如线框解析 [35], [60], [61], [62]。Lee等人 [34]遵循faster-RCNN [36]的两分支流

程，提出了直线检测框架来检测图像中有意义的语义直线。一条分支验证线的存在，另一条分支进一步回归线的详细位置。Zhang等人 [35]采用了CornerNet [37]的概念，在室内场景中通过提取成对的关键点来提取线段。Huang等人 [60]提出了两头网络来预测线和节点以进行线框解析。这是通过添加一个线提取子网络来扩展 [61]的方法。Zhou等人 [61]提出了端到端的结构在线框解析任务中进行详细的线段检测。

以上方法均通过LoI池化单独地沿每条线聚合深度特征表示来提取线级别的特征向量，从而导致不充分的上下文信息。

2.4 语义线检测

辅助图像摄影的有意义的直线最初由 [34]提出，称为“语义线”。[34]将语义线检测视为目标检测的一种特例。它首先使用LoI池化提取候选线的CNN特征表示，该方法首先沿整条直线对特征进行双线性插值。然后线特征表示由一个分类器和回归器验证，与Faster-RCNN [38]相似。图像中的候选线是所有唯一的线。两条线的评价指标重叠度(IoU)由 [34]提出用来评价图像中两条直线的相似度。这一指标可能在部分场景下导致模棱两可的定义，我们在Sec. 4中详述。此外，Lee等人 [34]收集了一个语义线检测数据集，共包括1,700张室外场景图片，大部分为自然景观。

3 方法

在本节中，我们详细介绍提出的用于语义线检测的深度霍夫变换。我们的方法主要包括四部分：1) 一个CNN编码器用于提取像素级别深度特征表示；2) 深度霍夫变换用于将深度特征表示从空间域转换到参数域；3) 一个线检测器负责检测参数空间中的线；4) 一个逆向霍夫变换将检测到的线映射回图像空间。所有的组成部分整合进一个端到端的框架中，可以在单个步骤中实现前向推理和反向传播训练。Fig. 2为流程的图解，详细的结构可见于我们的补充材料中。

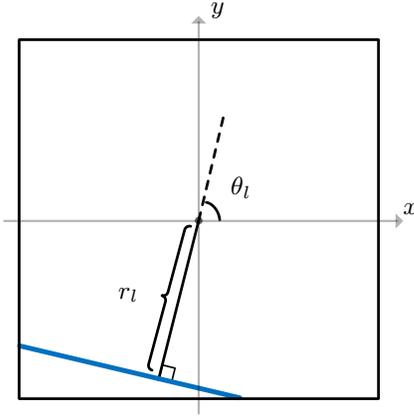


图 3. 一条直线可由截距 r_l 和斜率 θ_l 参数化表示。

3.1 线参数化和反参数化

如图 3 所示，给定一张二维图像 $I_{H \times W} \in \mathbb{R}^{H \times W}$ ，我们设定原点为图像中心。在二维平面上，一条直线 l 可由两个参数进行参数化：一个方向参数 $\theta_l \in [0, \pi)$ 代表 l 与 x 轴的角度和一个距离参数 r_l 表示 l 和原点的距离。显然 $\forall l \in I, r_l \in [-\sqrt{W^2 + H^2}/2, \sqrt{W^2 + H^2}/2]$ 。

给定图像 I 上的任意直线 l ，我们可以用以上方程进行参数化，也可以将任意有效的 (r, θ) 对逆向映射为一条线实例。我们定义线到参数和逆向映射为：

$$\begin{aligned} r_l, \theta_l &\leftarrow P(l), \\ l &\leftarrow P^{-1}(r_l, \theta_l). \end{aligned} \quad (1)$$

显然， P 和 P^{-1} 为双射映射。实际情况中， r 和 θ 量化至离散的分区中以供计算机程序处理。设 r 和 θ 的量化间隔分别为 Δr 和 $\Delta \theta$ 。然后量化过程可由以下公式表示：

$$\hat{r}_l = \left\lceil \frac{r_l}{\Delta r} \right\rceil, \quad \hat{\theta}_l = \left\lceil \frac{\theta_l}{\Delta \theta} \right\rceil, \quad (2)$$

其中 \hat{r}_l 和 $\hat{\theta}_l$ 为量化的线参数。由 Θ 和 R 表示的量化级数的个数为：

$$\Theta = \frac{\pi}{\Delta \theta}, \quad R = \frac{\sqrt{W^2 + H^2}}{\Delta r}, \quad (3)$$

如图 4(a) 所述。

3.2 使用深度霍夫变换的特征变换

3.2.1 深度霍夫变换

给定输入图像 I ，我们首先使用编码器网络提取深层 CNN 特征 $\mathbf{X} \in \mathbb{R}^{C \times H \times W}$ ，其中 C 表示通道数， H 和 W 为空间尺寸。随后，深度霍夫变换 (DHT) 以 \mathbf{X} 为输入并计算变换后的特征 $\mathbf{Y} \in \mathbb{R}^{C \times \Theta \times R}$ 。变换后特征的尺寸 Θ, R 由量化级数决定，如 Eq. (3) 所述。

如图 4(a) 所示，给定图像中任意直线 l ，我们沿着 l 聚集所有像素的特征得到参数空间 Y 中的点 $(\hat{\theta}_l, \hat{r}_l)$ ：

$$\mathbf{Y}(\hat{\theta}_l, \hat{r}_l) = \sum_{i \in l} \mathbf{X}(i), \quad (4)$$

其中 i 为位置索引。 $\hat{\theta}_l$ 和 \hat{r}_l 由直线 l 的参数根据 Eq. (1) 决定，然后根据 Eq. (2) 量化为离散值。

给定量化级数 Θ 和 R ，我们可得到 $\Theta \cdot R$ 个独立的候选线。然后对所有的候选线应用 DHT，它们对应的特征聚集到参数空间 \mathbf{Y} 中的对应位置。值得注意的是 DHT 在特征空间和参数空间中都是顺序不可知的，这使得该操作可以高度并行化。

3.2.2 基于 FPN 的多尺度 DHT

我们提出的 DHT 可以轻易地应用于任意尺度特征。我们使用特征金字塔网络 (FPN) [63] 作为我们的编码器。FPN 有助于提取多尺度的、有丰富语义的特征。

具体说来，FPN 输出 4 个特征图 X_1, X_2, X_3, X_4 ，其对应的分辨率为输入分辨率的 $1/4, 1/8, 1/16, 1/16$ 。然后每个特征图分别由 DHT 模块进行变换，如图 2 所示。由于这些特征图有不同的分辨率，变换后的特征 Y_1, Y_2, Y_3, Y_4 也有不同的尺寸，因为我们在每个阶段都使用相同的量化间隔 (详见 Eq. (3))。为了将变换后的特征融合在一起，我们将 Y_2, Y_3, Y_4 插值到 Y_1 的尺寸，然后将其连接在一起。

3.3 参数空间中的线检测

3.3.1 上下文敏感线检测器

经过深度霍夫变换 (DHT) 后，特征被转换至参数空间，其中网格位置 (θ, r) 对应特征空间中沿整条直线 $l = P^{-1}(\theta, r)$ 的特征。将特征转换入参数空间的一个重要原因是线结构可以更紧凑地表示。如图 4(b) 所示，线 l 附近的其他线也被转换为 (θ_l, r_l) 周围的点。随后，周围线的特征可以通过在参数空间中使用卷积层进行有效的聚集。

在 FPN 每一层，我们使用两个 3×3 卷积层聚集上下文相关线特征。然后，我们对特征图进行插值以匹配不同层的特征图的分辨率，如图 2 所示，并且将插值后的特征连接在一起。最后，在连接后的特征图上应用一个 1×1 卷积层生成点级别的预测。

3.3.2 损失函数

由于预测直接在参数空间中生成，我们也在同一空间中计算损失函数。对于一张训练图像 I ，真实的线首先通过标准霍夫变换转换入参数空间中。然后为了加速收敛，我们使用高斯核对真实数据进行平滑和扩展。在其他许多任务中也使用了相似的技巧，如人群计数 [64], [65] 和道路语义分割 [66]。形式上，设 \mathbf{G} 为参数空间中二值真实数据图， $\mathbf{G}_{i,j} = 1$ 表示参数空间中 i, j 处存在一条线。扩展后的真实数据图为

$$\hat{\mathbf{G}} = \mathbf{G} \circledast K,$$

其中 K 为 5×5 卷积核， \circledast 代表卷积操作。如图 2 所示为一对平滑后的真实数据和预测图的示例图。

最后，我们在参数空间中计算平滑后真实数据和预测图的交叉熵：

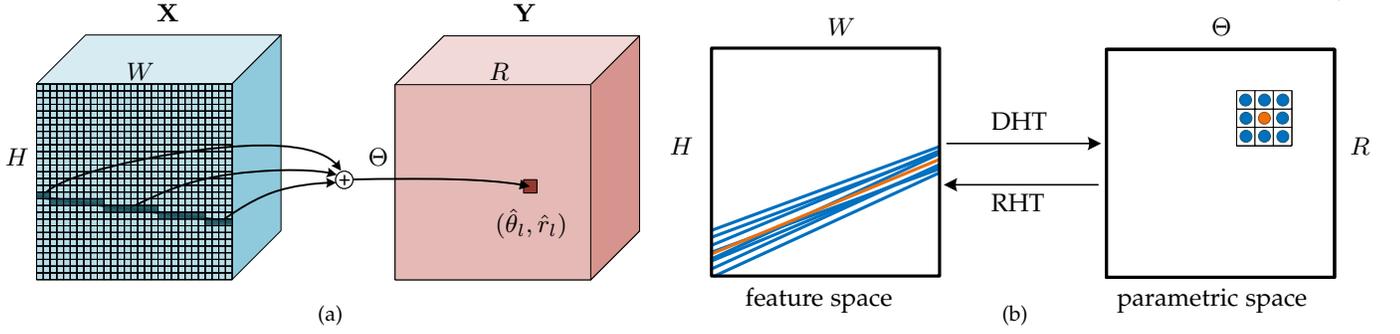


图 4. (a): 特征空间(蓝色, 左边)中沿直线的特征可累积至参数空间(红色, 右边)中的点 $(\hat{r}_l, \hat{\theta}_l)$ 。 (b): 所提出的上下文敏感特征聚合的图示。特征空间(左边)中附近线的特征转换至参数空间(右边)中的相邻点。参数空间中, 简单的 3×3 卷积操作可以轻易地获得中心线(橙色)周围的上下文信息。建议在彩色模式下观看。

$$L = - \sum_i \left\{ \hat{\mathbf{G}}_i \cdot \log(\mathbf{P}_i) + (1 - \hat{\mathbf{G}}_i) \cdot \log(1 - \mathbf{P}_i) \right\} \quad (5)$$

3.4 逆向映射

我们的检测器在参数空间中产生的预测代表存在线的概率。预测图使用阈值(如0.01)进行二值化。然后我们寻找相连的区域并计算对应的重心。这些重心代表参数空间中检测到的线。最后, 所有线根据 $P^{-1}(\cdot)$ 映射回图像空间, 按照Eq. (1)计算。我们将“映射回”的步骤称为“逆向霍夫变换(RHT)”, 如Fig. 2所示。

3.5 边缘引导的线微调

语义线为场景中分割不同区域的突出的结构。因此, 边缘能作为语义线的指引。我们提出通过使用边缘信息对齐线位置来优化检测结果。首先, 我们使用HED [42]计算边缘图 E 。随后, 给定检测到的线 l , l 的边缘密度定义为沿 l 的平均边缘响应:

$$\rho(l) = \frac{\sum_{i \in l} E_i}{|l|}, \quad (6)$$

其中 $|l|$ 为 l 的像素数。为了稳定性考虑, 我们处理Eq. (6)时在 l 两边各扩展一个像素(总宽度为3)。

设 \mathcal{L} 为 l 周围线的集合。这些线通过将 l 的端点顺时针和逆时针移动 δ_r 个像素获得。由于每条线有两个端点, 并且每个端点有 $\delta_r + 1$ 个可能的位置, 集合中元素个数为 $|\mathcal{L}| = (\delta_r + 1)^2$ 。然后最优位置可通过查找 \mathcal{L} 中有最高边缘密度的最优线 l^* 获得:

$$l^* = \arg \max_{l \in \mathcal{L}} \rho(l). \quad (7)$$

在Sec. 6.6.2记录了使用不同 δ_r 时“边缘引导的线微调”的性能。

4 所提出的评价标准

在本节中, 我们详细介绍所提出的衡量相似度的评价标准, 详细地说是衡量图像中两条线相似性的评价标准。首先, 我

们回顾计算机视觉社区中几个广泛应用的评价标准, 然后解释为什么现有评价标准不适用于我们的任务。最后, 我们介绍我们新提出的评价标准, 该标准同时考虑了欧氏距离和角度距离来计算两条线的相似度。

4.1 回顾现有标准

重叠度(IOU)广泛应用于目标检测、语义分割和许多其他任务, 该标准计量检测到的边界框(区域)与真实数据的相似度。Lee等人 [34]采用了原始的IOU应用于线检测, 并提出了基于线的IOU来评价检测到的线的质量。详细地说, 两条线的相似度由两条线的相交区域除以图像区域来度量。以Fig. 5(a)为例, 线 m 和 n 的相似性为 $\text{IOU}(m, n) = \text{area}(\text{red}) / \text{area}(I)$ 。

然而, 我们发现基于IOU的评价标准不适合该任务, 并且在某些场景下可能会导致不合理或模棱两可的结果。如Fig. 5(a)所示, 两对具有相似结构的线(m, n 和 p, q)可能会有非常不同的IOU分数。如Fig. 5(b)所示, 即使人也无法确定哪个区域(红色或蓝色)应该被用作线IOU的相交区域。

还有其他评价标准, 如Earth Mover's Distance (EMD) [67]和Chamfer distance(CD) [68], 可用于度量线的相似度。然而, 这些度量标准需要将线栅格化为像素, 然后计算像素级别距离, 该操作效率较低。

为了弥补不足之处, 我们提出了简单而高效的度量标准来度量参数空间中两条线的相似度。我们提出的评价标签比EMD和CD高效得多。Sec. 6.4中的量化对比表明我们提出的评价标准可得到与EMD和CD非常相似的结果。

4.2 提出的评价标准

我们提出的标准, 称作EA-score, 同时计算了一对线的欧氏距离(Euclidean distance)和角度距离(Angular distance)。设 l_i, l_j 为一对待量度的线, 对应的角度距离 S_θ 定义为两条线间的角度:

$$S_\theta = 1 - \frac{\theta(l_i, l_j)}{\pi/2}, \quad (8)$$

其中 $\theta(l_i, l_j)$ 为 l_i 和 l_j 间的角度。欧氏距离定义为:

$$S_d = 1 - D(l_i, l_j), \quad (9)$$

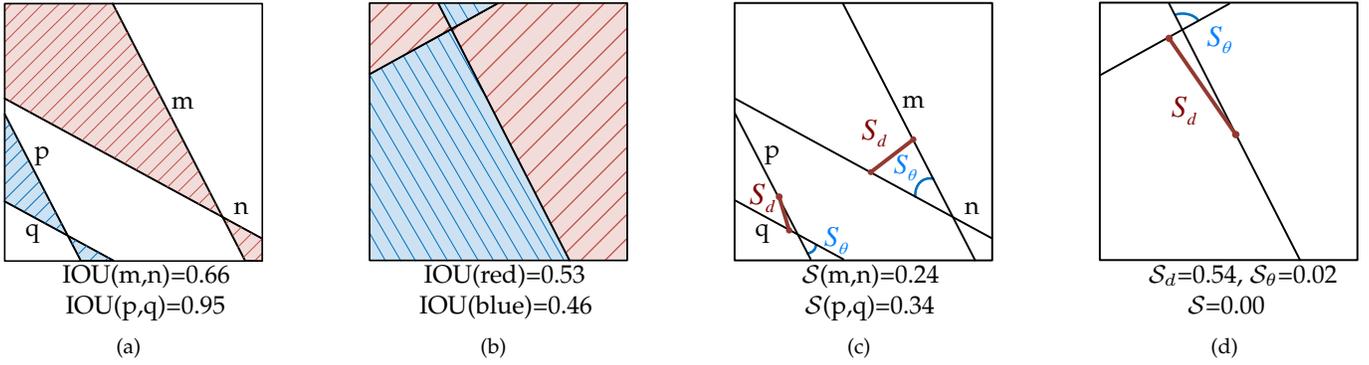


图 5. (a): 两对有相似相对位置的线可能有非常不同的IOU分数。(b): 即使人也无法分清哪个区域(蓝色或红色)应被视为基于IOU的评价标准 [34]的相交区域。(c)和(d): 我们提出的评价标准考虑了一对线的欧氏距离和角度距离, 得到了一致且合理的分数。建议在彩色模式下观看。

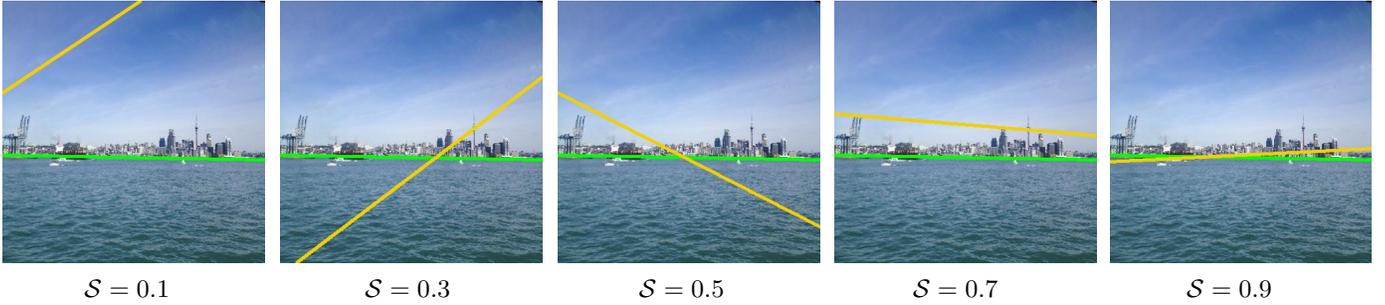


图 6. 不用EA-scores(Eq. (10)中的 S)的实例线。EA-scores越大两条线越相似。

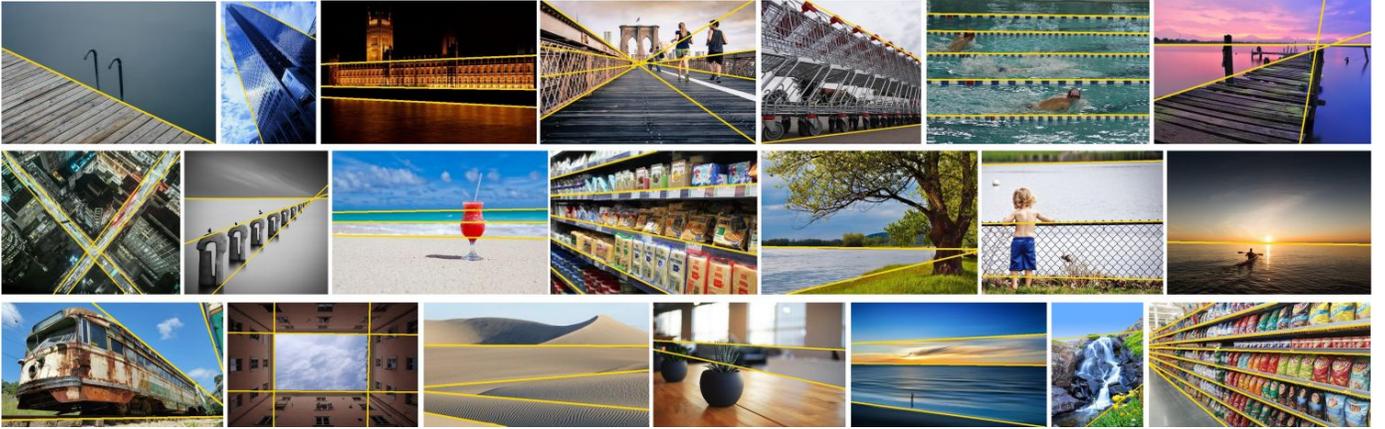


图 7. NKL中的示例图片和标注(黄线)。NKL中的图片包含多样的场景和丰富的线标注。

其中 $D(l_i, l_j)$ 为 l_i 和 l_j 中点间的欧氏距离。注意我们在计算 $D(l_i, l_j)$ 前将图像正则化为单位尺寸。

Fig. 5(c)和Fig. 5(d)为 S_d 和 S_θ 的示例。最后, 我们提出的EA-score为:

$$S = (S_\theta \cdot S_d)^2. \quad (10)$$

Eq. (10)使用二次方以便使得数值过大时更敏感且更有区分度。

Fig. 6中包括数个示例直线对和对应的EA-score。

5 NKL:语义线检测数据集

据我们所知, 目前仅有一个数据集SEL [34]用于语义线检测。SEL包括1,715张图片, 其中175张图片用于测试其他用于训

练。为了弥补基于CNN的大规模模型和现有数据集规模的差距, 我们收集了用于语义线检测的新数据集。

新的数据集, 名叫NKL(为NanKai Lines的简称), 包括6,500张图片, 在场景和线数量方面有更丰富的多样性。NKL中每张图片都由多名经验丰富的标注人员标注, 以保证标注的质量。数据集开源, 可见于我们的项目页面。

5.1 数据收集和标注

NKL中所有的图像都在网上使用海洋、草地等特定关键词爬取获得。经过版权检查后, 我们仔细地筛选出至少包括一条以上语义线的图像。由于语义线标注是主观的并且取决于标注人员, 因此每张图片首先由三位知识丰富的标注人员进行

表 1
SEL [34]和NKL中图像和线的数量。

Dataset	Total #images, #lines	Training #images, #lines	Evaluation #images, #lines
SEL [34]	1,715, 2,791	1,541, 2,493	175, 298
NKL (Ours)	6,500, 13,148	5,200, 10,498	1,300, 2,650

标注, 然后由其他人验证。一条线仅当所有3名标注人员结果一致时被认为是正确的。然后不一致的线由其他两名标注人员重审。总的来说, 对每条线, 最少由3人最多由5人进行标注, 并且每条线仅当由超过3人标注时被认为是正例。

5.2 数据集统计数据

5.2.1 图像和语义线数量

NKL中共有13,148条语义线, 而SEL [34]数据集中共有2,791条语义线。Tab. 1分别统计了两个数据集中的图像数和语义线数。

Fig. 8总结了NKL和SEL [34]数据集中每张图片语义线数量的柱状图。NKL数据集中超过半数的图片(67%, 4,356/6,500)包括至少一条语义线, 而SEL的比例仅为45.5%。

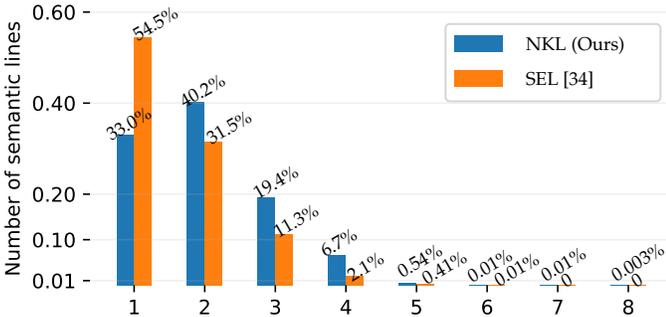


图 8. 线数量的柱状图表。我们数据集中的线比SEL分布更均匀。

5.2.2 多样性分析

为了分析SEL和NKL数据集的多样性, 我们将所有图片输入在Place365 [69]上预训练的ResNet50 [70]网络, 然后收集输出作为分类标签。Fig. 9为结果。

Place365 [69]数据集共包括365种类别, 我们在SEL数据集种共得到167种类别而在NKL中有327类。此外, 如Fig. 9所示, NKL数据集的场景标签比SEL数据集分布更均匀。举例来说, SEL数据集中, 前三个主要类别(天空, 野地, 沙漠)占全部的25%以上。而在NKL, 前三个主要类别少于全部的20%。

6 实验

在本节中, 我们介绍我们系统的实现细节, 并展示与现有方法的实验结果对比。

6.1 实现细节

我们的系统使用PyTorch [71]框架实现, 并且Jittor [72]版本的实现也已开源。由于我们提出的深度霍夫变换(DHT)高度并行化, 我们使用原生CUDA编程实现了DHT, 并且所有其他部分基于框架级别的Python API实现。我们在单个RTX 2080 Ti GPU上进行所有实验。

6.1.1 网络结构

我们使用两个有代表性的网络结构ResNet50 [70]和VGGNet16 [73]作为骨干网络, 并使用FPN [63]提取多尺度深度特征表示。对于ResNet网络, 我们按照现有工作 [74], [75], [76]的常用做法, 在最后一层使用空洞卷积 [77]提高特征图的分辨率。

6.1.2 超参数

Sec. 3.3.2使用的高斯核尺寸定义为 5×5 。所有图像尺寸调整为(400,400), 然后封装为数量为8的mini-batch。我们使用不包含权重衰减的Adam优化器 [78]对所有模型训练30轮。学习率和动量分别设为 2×10^{-4} 和0.9。量化间隔 $\Delta\theta, \Delta r$ 详见Sec. 6.3和Eq. (12)。

6.1.3 数据集和数据增强

我们的实验在SEL [34]数据集和提出的NKL数据集上进行。两数据集的统计数据细节详见Sec. 5。按照 [34]初始设置, 我们在所有实验中仅使用左右翻转进行数据增强。

6.2 评估流程

我们使用准确率、召回率和F-measure评价检测到的线的质量。第一步是匹配检测到的线与真实的线。

设 \mathcal{P} 和 \mathcal{G} 分别为预测的线和真实的线的集合。 p_i 和 g_j 为单个预测的线和真实的线。我们首先使用二部图匹配算法来匹配 \mathcal{P} 和 \mathcal{G} 中的线。设 $G = \{V, E\}$ 为一个二部图¹。顶点集 V 可被分为两个不相交的、独立的集合, 在我们的实验中为 \mathcal{P} and \mathcal{G} :

$$V = \mathcal{P} \cup \mathcal{G}$$

$$\mathcal{P} \cap \mathcal{G} = \emptyset.$$

E 中每条连线代表在某一相似性度量下一对线间的相似性。除了提出的EA-score, 我们还使用其他两个广泛使用的度量标准: earth mover's distance (EMD) [67]和Chamfer distance (CD) [68], 如Sec. 4所述。注意, 我们根据EMD和CD可能的最大值将两者正则化到[0, 1]范围内(对EMD和CD, 当两条直线在对角线上缩为一点时达到最大值)。

给定图 $G = \{V, E\}$, 二部图匹配结果是保证任意两条边之间没有公共点的前提下得到的边集。在我们的任务中,

1. https://en.wikipedia.org/wiki/Bipartite_graph

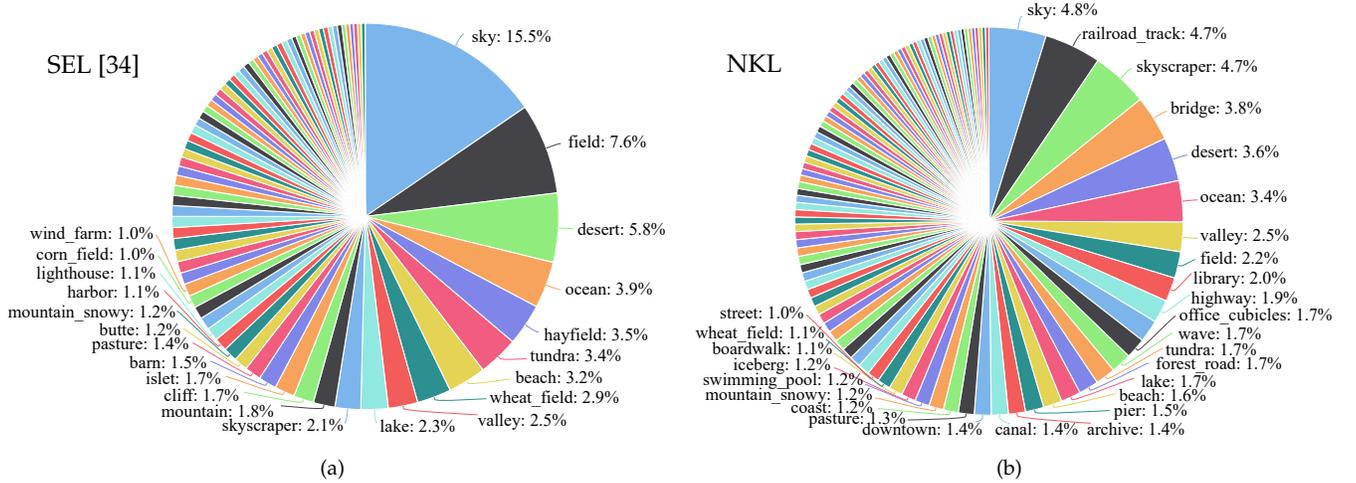


图 9. SEL (a)和NKL (b)数据集的类别分布。类别标签通过Places365预训练模型获得。NKL数据集中包括327(共365)种场景标签，不同于SEL数据集的167种。NKL的标签比SEL分布更均匀。

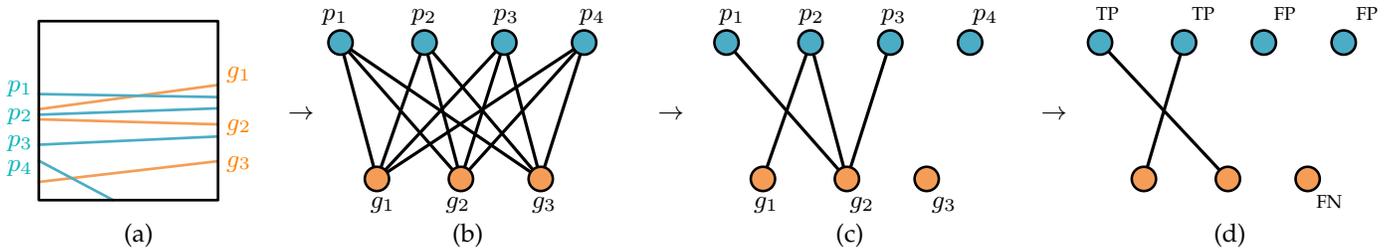


图 10. 评价过程中的二部图匹配算法图解。(a) 包含3个真实的线(g_1, g_2, g_3)和4个预测(p_1, p_2, p_3, p_4)示例图像。(b) 对应的二部图。一对节点间的边缘代表两条线间的相似度(Eq. (10)中的 S)。(c) 通过二部图最大匹配，每个子图中的节点与另一个子图中的节点有不超过1条连线。(d) true positive (TP), false positive (FP) and false negative (FN)。

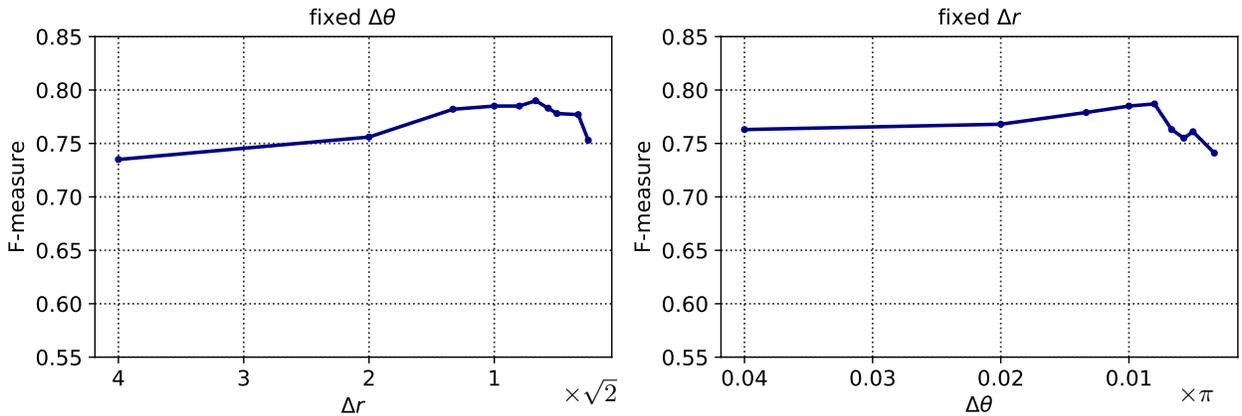


图 11. 左图：固定角度量化间隔 $\Delta\theta = \pi/100$ ，使用不同距离量化间隔 Δr 时的性能。量化间隔 Δr 越大代表量化级数 R 越少。右图：固定距离量化间隔 $\Delta r = \sqrt{2}$ ，使用不同角度量化间隔 $\Delta\theta$ 时的性能。

给定预测的线的集合 \mathcal{P} 和真实线的集合 \mathcal{G} ，我们寻找合适的匹配，以保证每条真实线 g_i 对应于不超过一条检测到的线 p_j ，反过来亦是如此。这一问题，二部图最大匹配，可使用经典的Hungarian方法 [79]在多项式时间复杂度内解决。

匹配 \mathcal{P} 和 \mathcal{G} 后，我们可以据此计算true positive (TP), false positive (FP)和false negative (FN)。如图 10所示，预测的直线对(p_1, p_2)与真实直线对(g_2, g_1)匹配时被视为true positive。预测的线(p_3)与真实直线间不存在匹配时被视为false posi-

tive，真实直线(g_3)不存在对应的预测的线时被认为是false negative。

最后，准确率(Precision)、召回率(Recall)和F-measure为：

$$P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN}, F = \frac{2PR}{P + R}. \quad (11)$$

我们对预测数据和真实数据对采用一系列阈值 $\tau = 0.01, 0.02, \dots, 0.99$ 。对应地，我们得到了一系列准确率、召

回率和F-measure。最后，我们使用平均准确率、召回率和F-measure评价性能。我们使用EMD [67], CD [68]和我们提出的EA评价标准用作量化比较。在消融实验中，为了简便我们仅用EA作为评价标准。

6.3 调整量化间隔

Eq. (2)中的量化间隔 $\Delta\theta$ 和 Δr 是影响性能和运行效率的重要因素。量化间隔越大量化级数越小，即 Θ 和 R ，模型也越快。量化间隔越小，量化级数越多，总体计算代价越大。

我们在SEL [34]数据集上使用坐标下降法寻找合适的量化间隔，以使得计算高效且功能上高效。注意为了简便我们使用EA-score度量线的相似性。在第一阶段，我们固定角度量化间隔为 $\Delta\theta = \pi/100$ ，寻找 Δr ，结果如图Fig. 11(a)所示。根据Fig. 11(a)，随着 Δr 的下降模型性能首先缓慢上升随后下降，转折点在 $\Delta r = \sqrt{2}$ 附近。在第二阶段，我们固定 $\Delta r = \sqrt{2}$ 然后使用不同 $\Delta\theta$ 训练。与Fig. 11(a)相似，Fig. 11(b)结果表明随着 $\Delta\theta$ 下降模型性能首先平滑地上升，随后快速振荡下降。因此，转折点 $\Delta\theta = \pi/100$ 为角度量化的合适的选择。

总结来说，我们使用 $\Delta\theta = \pi/100$ 和 $\Delta r = \sqrt{2}$ 进行量化，对应的量化级数为：

$$\Theta = 100, R = \sqrt{\frac{W^2 + H^2}{2}}, \quad (12)$$

其中 H, W 为DHT中需要转换的特征图的尺寸。

6.4 定量比较

我们对我们的方法与SLNet [34]和使用HED [42]作为边缘检测器的经典霍夫线检测 [25]。注意我们使用SEL [34]训练集并使用其线标注作为真实边缘来训练HED边缘检测器。

6.4.1 在SEL数据集上的结果

Tab. 2总结了在SEL数据集 [34]上的结果。使用VGG16或ResNet50作为骨干网络，我们的方法持续大幅度超过SLNet和HT+HED。除了Tab. 2，我们画出了F-measure *v.s.* 阈值和准确率*v.s.*召回率曲线。Fig. 12表明，我们的方法在大多数阈值下取得了比其他方法更高的F-measure。

6.4.2 NKL数据集上的结果

我们展示了在新构建的NKL数据集上的结果。由于SLNet [34]并未开源训练代码，我们仅与HED+HT比较我们的方法。如Tab. 2所示，我们提出的方法明确领先于基线方法(HED边缘检测器 + 霍夫变换)

6.4.3 运行效率

在本节中，我们对不同方法进行运行效率基准测试，包括不同迭代次数的SLNet [34]、经典霍夫变换和我们提出的方法。

SLNet [34]和HT均需要使用HED [42]边缘检测器作为预处理步骤。SLNet中的非最大抑制(NMS)需要边缘图引导，经典霍夫变换需要以边缘图为输入。此外，SLNet使用微调网络迭代地增强结果。因此，推理速度也与迭代次数相关。相反，我们的方法仅需要一次前向传播便可得到结果，而且NMS可简化为在参数空间中计算相连区域的重心。

Tab. 3中的结果表明，我们的方法在速度上大幅度超过其他对比方法。即使仅使用一次迭代，SLNet仍慢于我们的方法。

6.5 定性比较

在此我们给出我们的方法与SLNet和HED+HT的几个示例结果。如图Fig. 13所示，对比于其他方法，我们的结果更符合真实数据和人类认知。除了Fig. 13，我们在补充材料中提供所有的检测结果。

6.6 消融研究

在本节中，我们消融研究我们方法的每个组成成分。

6.6.1 DHT中的组成部分

我们首先研究“深度霍夫变换”的组成部分。具体地说，他们是：(a)深度霍夫变换(DHT)模块，详见Sec. 3.2; (b)多尺度(MS)DHT结构，详见Sec. 3.2.2; (c)上下文敏感(CTX)线检测器，详见Sec. 3.3.1。实验结果见于Tab. 4。

我们首先使用朴素ResNet50和DHT模块构建基线模型。然后，我们分别验证多尺度(MS)策略的效果和上下文感知线检测器的效果。我们分别将MS和CTX加在基线模型后，分别验证其性能。Tab. 4中的结构表明MS和CTX可以提升基线模型的性能。

最后，我们使用所有组件组合成我们完整的方法，该模型在所有组合中取得了最好的性能。本节中的实验结果清晰地表明，每个组成部分对于我们的方法都有帮助。

6.6.2 边缘引导的微调

在此我们消融研究“边缘引导的微调”模块(缩写为ER)。首先，我们测试使用不同 δ_r 时DHT+ER的性能。参数 δ_r 控制ER(Eq. (7)中的 \mathcal{L})的搜索空间。实验用ResNet50作为骨干网络在SEL数据集上进行。Tab. 5中结果表明随着 δ_r 提升，性能首先提升随后饱和。由于峰值性能在 $\delta_r = 5$ 时出现，我们设 $\delta_r = 5$ 为最优性能。设 δ_r 为5后，我们测试了我们方法使用不同骨干网络和数据集时，有和没有ER的性能。

Tab. 6中的结果清晰地表明，边缘引导的微调模块可以在不同骨干模型和数据集下有效地提升检测结果。

表 2

在SEL [34]和NKL数据集上的定量比较。在SEL [34]数据集上，我们的方法(不使用ER)在平均F-measure大幅度超过其他方法。‘CD’, ‘EMD’, 和‘EA’为不同的评价标准，详见Sec. 4。

Dataset	Method	CD			EMD			EA		
		Avg. P	Avg. R	Avg. F	Avg. P	Avg. R	Avg. F	Avg. P	Avg. R	Avg. F
SEL [34]	HED [42] + HT [25]	0.491	0.578	0.531	0.461	0.543	0.498	0.356	0.420	0.385
	SLNet-iter1 [34]	0.740	0.905	0.812	0.723	0.888	0.797	0.654	0.803	0.721
	SLNet-iter5 [34]	0.826	0.841	0.834	0.810	0.824	0.817	0.735	0.747	0.741
	SLNet-iter10 [34]	0.858	0.821	0.839	0.840	0.804	0.822	0.762	0.729	0.745
	Ours (VGG16)	0.841	0.835	0.838	0.830	0.824	0.827	0.756	0.774	0.765
	Ours (ResNet50)	0.886	0.815	0.849	0.878	0.807	0.841	0.819	0.755	0.786
NKL	HED [42] + HT [25]	0.301	0.878	0.448	-	-	-	0.213	0.622	0.318
	Ours (VGG16)	0.750	0.864	0.803	0.726	0.837	0.778	0.659	0.759	0.706
	Ours (ResNet50)	0.766	0.864	0.812	0.743	0.839	0.789	0.679	0.766	0.719

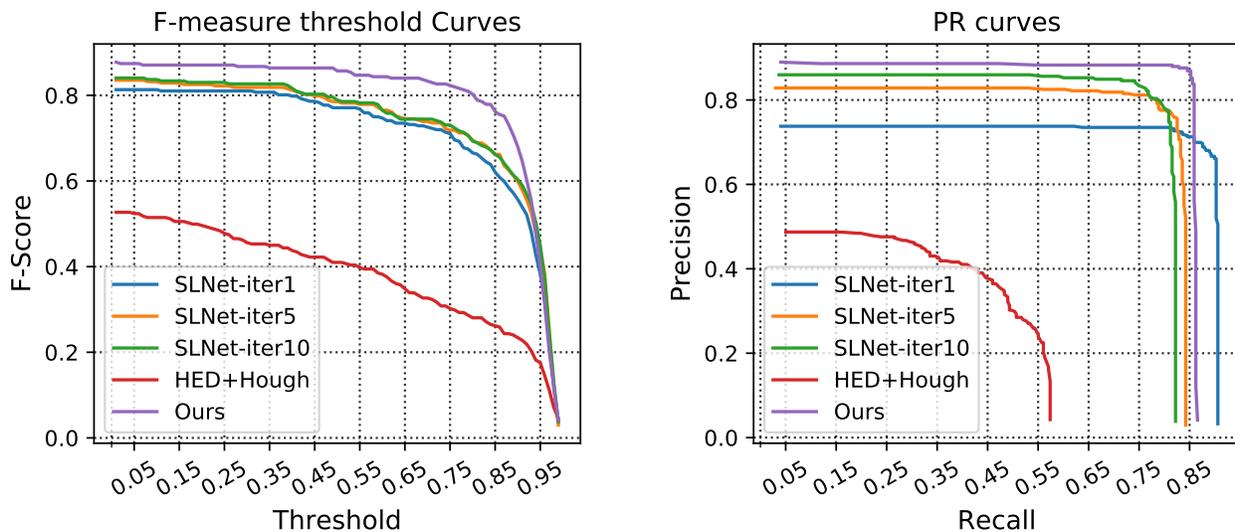


图 12. 左图：不同阈值下的F-measure。右图：准确率-召回率曲线。我们的方法大幅度超过SLNet [34]和经典霍夫变换 [25]。此外，即使经过10轮的位置细化，SLNet仍展示出更差的性能。

表 3

量化速度比较。我们的方法(不使用ER)在前向传播时远快于其他两个对比方法。此外，我们的方法不需要任何额外处理，如边缘检测。因此，我们的方法能以49FPS运行，这比其他两个方法快得多。

Method	Network forward	NMS	Edge	Total
SLNet-iter1 [34]	0.354 s	0.079 s	0.014 s	0.447 s
SLNet-iter3 [34]	0.437 s	0.071 s	0.014 s	0.522 s
SLNet-iter10 [34]	0.827 s	0.068 s	0.014 s	0.909 s
HED [42] + HT [25]	0.014 s	0.117 s	0.024 s	0.155 s
Ours (VGG16)	0.03 s	0.003 s	0	0.033 s
Ours (ResNet50)	0.017 s	0.003 s	0	0.020 s

表 4

对每个组成部分的消融实验。MS代表使用多尺度特征的DHTs，如Sec. 3.2.2所述；CTX代表上下文敏感聚合，如Sec. 3.3.1所述。

DHT	MS	CTX	F-measure
✓			0.664
✓	✓		0.758
✓		✓	0.771
✓	✓	✓	0.786

7 结论

在本文中，我们提出了一个简单而高效的方法用于自然场景中的语义线检测。通过将CNNs的较强的学习能力与经典霍夫变换结合，我们的方法能有效地捕捉直线的复杂纹理和丰



图 13. SEL数据集上的不同方法检测结果示例。相比于SLNet [34]和经典霍夫变换 [25]，我们的结果与真实数据更一致。

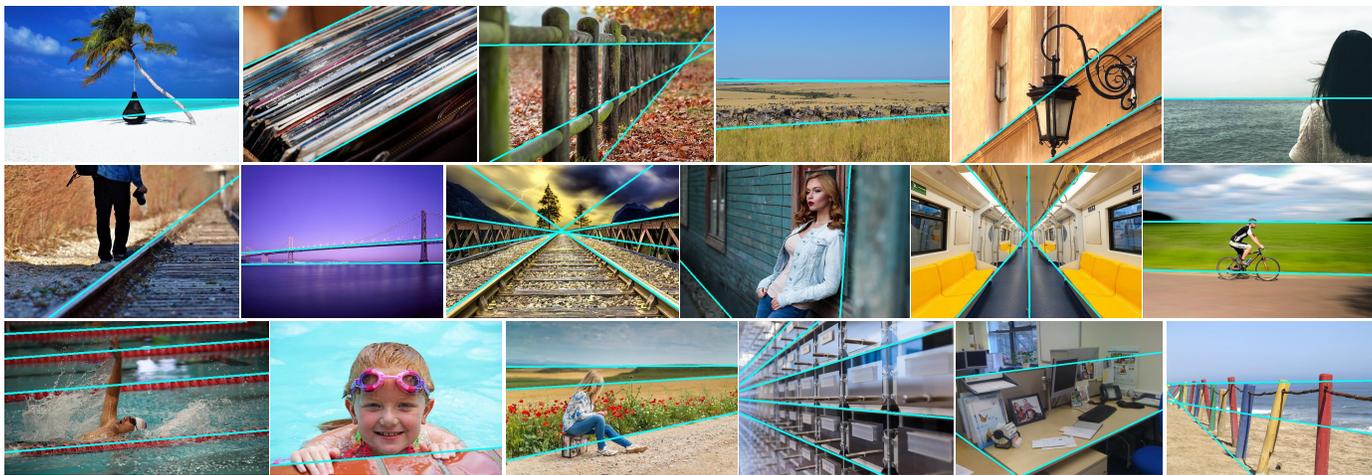


图 14. 在NKL数据集上的检测结果。我们的方法可获得视觉上符合人类感知的结果。

富的上下文语义信息。为了更好的检测一对线的相似性，我们设计了新的评价标准，同时考虑到线间的欧氏距离和角度距离。此外，我们构建了新的用于语义线检测的数据集来填补现有数据集规模和现代CNN模型复杂度的距离。定量和定性结果均表明，我们的方法在检测质量和速度方面大幅度超过原有方法。

致谢

该项研究得到了新一代AI重大项目(2018AAA0100400)，国家自然科学基金(61922046,61620106008,62002176)，中国教育部部科技创新计划和天津自然科学基金会(17JCQJC43700)的资助。

表 5

使用不同 δ_r 时DHT+ER的性能。模型以ResNet50为骨干网络在SEL数据集上训练/测试。 $\delta_r = 0$ 代表不使用ER的原始DHT方法。

δ_r	Precision	Recall	F-measure
0	0.8190	0.7530	0.7861
1	0.8199	0.7561	0.7866
3	0.8208	0.7569	0.7874
5	0.8214	0.7574	0.7880
7	0.8213	0.7573	0.7878
9	0.8212	0.7571	0.7877

表 6

使用不同骨干网络和数据集时，模型在有和没有ER($\delta_r = 5$)时的性能。

Dataset	Arch	Edge	P	R	F	F@0.95
SEL [34]	VGG16		0.756	0.774	0.765	0.380
	VGG16	✓	0.758	0.777	0.770	0.439
	Resnet50		0.819	0.753	0.786	0.420
	Resnet50	✓	0.821	0.757	0.788	0.461
NKL	VGG16		0.659	0.759	0.706	0.434
	VGG16	✓	0.664	0.765	0.711	0.472
	Resnet50		0.679	0.766	0.719	0.459
	Resnet50	✓	0.684	0.771	0.725	0.486

参考文献

- [1] K. Zhao, Q. Han, C.-B. Zhang, J. Xu, and M.-M. Cheng, "Deep hough transform for semantic line detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
- [2] Q. Han, K. Zhao, J. Xu, and M.-M. Cheng, "Deep hough transform for semantic line detection," in *Eur. Conf. Comput. Vis.*, 2020, pp. 750–766.
- [3] J. B. Burns, A. R. Hanson, and E. M. Riseman, "Extracting straight lines," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 4, pp. 425–455, 1986.
- [4] K. Zhao, S. Gao, W. Wang, and M.-M. Cheng, "Optimizing the F-measure for threshold-free salient object detection," in *Int. Conf. Comput. Vis.*, Oct 2019, pp. 8849–8857.
- [5] Q. Hou, M.-M. Cheng, X. Hu, A. Borji, Z. Tu, and P. Torr, "Deeply supervised salient object detection with short connections," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 4, pp. 815–828, 2019.
- [6] S.-H. Gao, Y.-Q. Tan, M.-M. Cheng, C. Lu, Y. Chen, and S. Yan, "Highly efficient salient object detection with 100k parameters," in *Eur. Conf. Comput. Vis.*, 2020.
- [7] A. Borji, M.-M. Cheng, Q. Hou, H. Jiang, and J. Li, "Salient object detection: A survey," *Computational Visual Media*, vol. 5, no. 2, pp. 117–150, 2019.
- [8] W. Wang, J. Shen, J. Xie, M.-M. Cheng, H. Ling, and A. Borji, "Revisiting video saliency prediction in the deep learning era," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 220–237, 2021.
- [9] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, "Global contrast based salient region detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 569–582, 2015.
- [10] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2014, pp. 2814–2821.
- [11] R. Fan, M.-M. Cheng, Q. Hou, T.-J. Mu, J. Wang, and S.-M. Hu, "S4net: Single stage salient-instance segmentation," *Computational Visual Media*, vol. 6, no. 2, pp. 191–204, June 2020.
- [12] L. Liu, R. Chen, L. Wolf, and D. Cohen-Or, "Optimizing photo composition," *Comput. Graph. Forum*, vol. 29, no. 2, pp. 469–478, 2010.
- [13] M. Freeman, *The photographer's eye: composition and design for better digital photos*. CRC Press, 2007.
- [14] M.-M. Cheng, X.-C. Liu, J. Wang, S.-P. Lu, Y.-K. Lai, and P. L. Rosin, "Structure-preserving neural style transfer," *IEEE Trans. Image Process.*, vol. 29, pp. 909–920, 2020.
- [15] S.-M. Hu, F.-L. Zhang, M. Wang, R. R. Martin, and J. Wang, "Patchnet: a patch-based image representation for interactive library-driven image editing," *ACM Trans. Graph.*, vol. 32, no. 6, pp. 1–12, 2013.
- [16] K. Ko, J.-T. Lee, and C.-S. Kim, "Pac-net: pairwise aesthetic comparison network for image aesthetic assessment," in *IEEE Int. Conf. Image Process.* IEEE, 2018, pp. 2491–2495.
- [17] J.-T. Lee, C. Lee, and C.-S. Kim, "Property-specific aesthetic assessment with unsupervised aesthetic property discovery," *IEEE Access*, vol. 7, pp. 114 349–114 362, 2019.
- [18] S. Kong, X. Shen, Z. Lin, R. Mech, and C. Fowlkes, "Photo aesthetics ranking network with attributes and content adaptation," in *Eur. Conf. Comput. Vis.* Springer, 2016, pp. 662–679.
- [19] L. Mai, H. Jin, and F. Liu, "Composition-preserving deep photo aesthetics assessment," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2016, pp. 497–506.
- [20] R. Fan, X. Wang, Q. Hou, H. Liu, and T.-J. Mu, "Spinnet: Spinning convolutional network for lane boundary detection," *Computational Visual Media*, vol. 5, no. 4, pp. 417–428, 2019.
- [21] B. Krages, *Photography: the art of composition*. Simon and Schuster, 2012.
- [22] S.-M. Hu, K. Xu, L.-Q. Ma, B. Liu, B.-Y. Jiang, and J. Wang, "Inverse image editing: Recovering a semantic editing history from a before-and-after image pair," *ACM Trans. Graph.*, vol. 32, no. 6, pp. 1–11, 2013.
- [23] T. Chen, M.-M. Cheng, P. Tan, A. Shamir, and S.-M. Hu, "Sketch2photo: Internet image montage," *ACM Trans. Graph.*, vol. 28, no. 5, pp. 1–10, 2009.
- [24] S.-H. Zhang, Z.-P. Zhou, B. Liu, X. Dong, and P. Hall, "What and where: A context-based recommendation system for object insertion," *Computational Visual Media*, vol. 6, pp. 79–93, 2020.
- [25] R. O. Duda and P. E. Hart, "Use of the hough transformation to detect lines and curves in pictures," Sri International Menlo Park Ca Artificial Intelligence Center, Tech. Rep., 1971.
- [26] D. Ballard, "Generating the hough transform to detect arbitrary shapes," *Pattern Recog.*, vol. 13, no. 2, 1981.
- [27] P. V. Hough, "Method and means for recognizing complex patterns," 1962, uS Patent 3,069,654.
- [28] L. A. Fernandes and M. M. Oliveira, "Real-time line detection through an improved hough transform voting scheme," *Pattern Recog.*, vol. 41, no. 1, pp. 299–314, 2008.
- [29] S. B. Yacoub and J.-M. Jolion, "Hierarchical line extraction," *IEE Proceedings-Vision, Image and Signal Processing*, vol. 142, no. 1, pp. 7–14, 1995.

- [30] J. Princen, J. Illingworth, and J. Kittler, "A hierarchical approach to line extraction based on the hough transform," *Computer vision, graphics, and image processing*, vol. 52, no. 1, pp. 57–77, 1990.
- [31] N. Kiryati, Y. Eldar, and A. M. Bruckstein, "A probabilistic hough transform," *Pattern Recogn.*, vol. 24, no. 4, pp. 303–316, 1991.
- [32] C. Akinlar and C. Topal, "Edlines: A real-time line segment detector with a false detection control," *Pattern Recogn.*, vol. 32, no. 13, pp. 1633–1642, 2011.
- [33] S. Caplin, *Art and Design in Photoshop*. Elsevier/Focal, 2008.
- [34] J.-T. Lee, H.-U. Kim, C. Lee, and C.-S. Kim, "Semantic line detection and its applications," in *Int. Conf. Comput. Vis.*, 2017, pp. 3229–3237.
- [35] Z. Zhang, Z. Li, N. Bi, J. Zheng, J. Wang, K. Huang, W. Luo, Y. Xu, and S. Gao, "PPGnet: Learning point-pair graph for line segment detection," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2019.
- [36] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Adv. Neural Inform. Process. Syst.*, 2015, pp. 91–99.
- [37] H. Law and J. Deng, "Cornernet: Detecting objects as paired keypoints," in *Eur. Conf. Comput. Vis.*, 2018, pp. 734–750.
- [38] R. Girshick, "Fast r-cnn," in *Int. Conf. Comput. Vis.*, 2015, pp. 1440–1448.
- [39] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2018, pp. 7794–7803.
- [40] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, "Ccnnet: Criss-cross attention for semantic segmentation," in *Int. Conf. Comput. Vis.*, 2019, pp. 603–612.
- [41] Y. Liu, M.-M. Cheng, X. Hu, J.-W. Bian, L. Zhang, X. Bai, and J. Tang, "Richer convolutional features for edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 8, pp. 1939 – 1946, 2019.
- [42] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Int. Conf. Comput. Vis.*, 2015, pp. 1395–1403.
- [43] J. Radon, "über die bestimmung von funktionen durch ihre integralwerte längs gewisser mannigfaltigkeiten," *Classic papers in modern diagnostic radiology*, vol. 5, p. 21, 2005.
- [44] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 679–698, 1986.
- [45] I. Sobel, "An isotropic 3x3 image gradient operator," *Presentation at Stanford A.I. Project*, 2014.
- [46] Z. Chen, A. Tagliasacchi, and H. Zhang, "Bsp-net: Generating compact meshes via binary space partitioning," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2020, pp. 45–54.
- [47] B. Deng, K. Genova, S. Yazdani, S. Bouaziz, G. Hinton, and A. Tagliasacchi, "Cvxnet: Learnable convex decomposition," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2020, pp. 31–44.
- [48] F. O’gorman and M. Clowes, "Finding picture edges through collinearity of feature points," *IEEE Trans. Computers*, vol. 25, no. 4, pp. 449–456, 1976.
- [49] F. A. Limberger and M. M. Oliveira, "Real-time detection of planar regions in unorganized point clouds," *Pattern Recogn.*, vol. 48, no. 6, pp. 2043–2053, 2015.
- [50] J. Illingworth and J. Kittler, "The adaptive hough transform," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 9, no. 5, pp. 690–698, 1987.
- [51] N. Aggarwal and W. C. Karl, "Line detection in images through regularized hough transform," *IEEE Trans. Image Process.*, vol. 15, no. 3, pp. 582–591, 2006.
- [52] C. R. Qi, O. Litany, K. He, and L. J. Guibas, "Deep hough voting for 3d object detection in point clouds," in *Int. Conf. Comput. Vis.*, 2019, pp. 9277–9286.
- [53] J. Min, J. Lee, J. Ponce, and M. Cho, "Hyperpixel flow: Semantic correspondence with multi-layer neural features," in *Int. Conf. Comput. Vis.*, 2019, pp. 3395–3404.
- [54] A. Etemadi, "Robust segmentation of edge data," in *International Conference on Image Processing and its Applications*. IET, 1992, pp. 311–314.
- [55] T. Chan and R. K. Yip, "Line detection algorithm," in *Int. Conf. Pattern Recog.*, vol. 2. IEEE, 1996, pp. 126–130.
- [56] R. G. Von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall, "Lsd: A fast line segment detector with a false detection control," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 4, pp. 722–732, 2008.
- [57] C. Akinlar and C. Topal, "Edcircles: A real-time circle detector with a false detection control," *Pattern Recogn. Let.*, vol. 46, no. 3, pp. 725–740, 2013.
- [58] H. Feng, Z. Jiang, F. Xie, P. Yang, J. Shi, and L. Chen, "Automatic fastener classification and defect detection in vision-based railway inspection systems," *IEEE T Instrum. Meas.*, vol. 63, no. 4, pp. 877–888, 2013.
- [59] T. Ahmad, P. Campr, M. Čadik, and G. Bebis, "Comparison of semantic segmentation approaches for horizon/sky line detection," in *International joint conference on neural networks*. IEEE, 2017, pp. 4436–4443.
- [60] K. Huang, Y. Wang, Z. Zhou, T. Ding, S. Gao, and Y. Ma, "Learning to parse wireframes in images of man-made environments," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2018, pp. 626–635.
- [61] Y. Zhou, H. Qi, and Y. Ma, "End-to-end wireframe parsing," in *Int. Conf. Comput. Vis.*, 2019, pp. 962–971.
- [62] N. Xue, T. Wu, S. Bai, F. Wang, G.-S. Xia, L. Zhang, and P. H. Torr, "Holistically-attracted wireframe parsing," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2020, pp. 2788–2797.
- [63] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2017, pp. 2117–2125.
- [64] W. Liu, M. Salzmann, and P. Fua, "Context-aware crowd counting," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2019, pp. 5099–5108.
- [65] Z.-Q. Cheng, J.-X. Li, Q. Dai, X. Wu, and A. G. Hauptmann, "Learning spatial awareness to improve crowd counting," in *Int. Conf. Comput. Vis.*, 2019, pp. 6152–6161.
- [66] Y.-Q. Tan, S. Gao, X.-Y. Li, M.-M. Cheng, and B. Ren, "Vecroad: Point-based iterative graph exploration for road graphs extraction," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2020.
- [67] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover’s distance as a metric for image retrieval," *Int. J. Comput. Vis.*, vol. 40, no. 2, pp. 99–121, 2000.
- [68] G. Borgefors, "Distance transformations in digital images," *Computer vision, graphics, and image processing*, vol. 34, no. 3, pp. 344–371, 1986.
- [69] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, "Places: A 10 million image database for scene recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017.
- [70] 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.
- [71] 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. 8024–8035.

- [72] 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 Science*, vol. 63, no. 222103, pp. 1–21, 2020.
- [73] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Int. Conf. Learn. Represent.*, 2015.
- [74] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2017, pp. 2881–2890.
- [75] 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.
- [76] 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.
- [77] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *Int. Conf. Learn. Represent.*, 2016.
- [78] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Int. Conf. Learn. Represent.*, 2015.
- [79] H. W. Kuhn, "The hungarian method for the assignment problem," *Naval research logistics quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955.



Jun Xu 分别于2011年和2014年在南开大学数学科学院取得学士学位和理学硕士学位，于2018年在香港理工大学计算机系取得博士学位。他目前是南开大学统计与数据科学学院讲师。他的主页为<https://csjunxu.github.io/>。



Kai Zhao 在上海大学取得学士和硕士学位。目前就读于南开大学计算机学院，导师为程明明教授他的研究方向包括统计学习和计算机视觉。



Ming-Ming Cheng 于2012年在清华大学取得博士学位。随后在牛津大学与Prof. Philip Torr合作担任两年研究员。他目前为南开大学教授，领导媒体计算实验室。他的研究方向包括计算机图形学、计算机视觉和图像处理。他获得的荣誉奖项包括ACM中国心性奖、IBM全球SUR奖、CCF-英特尔青年教师研究计划等。



Qi Han 目前在南开大学计算机学院攻读硕士学位，导师为程明明教授。他于2019年在西安电子科技大学取得学士学位。他的研究方向包括深度学习和计算机视觉。



Chang-Bin Zhang 目前在南开大学计算机学院攻读硕士学位，导师为程明明教授。他于2019年在中国矿业大学取得学士学位。他的研究方向包括深度学习和计算机视觉。