

doi: 10.12052/gdutxb.200006

基于差分响应图的无监督特征点检测网络

林璟怡, 李 东, 胡晓瑞

(广东工业大学 自动化学院, 广东 广州 510006)

摘要: 为了突破基于人工设计的特征点检测器的性能限制, 提出了一种新的数据驱动的基于差分特征响应图的无监督特征点检测网络。该网络使用不同尺度的卷积核计算差分输出, 利用差分响应图的绝对值大小筛选出大量特征点, 并评判这些特征点重要性程度。同时采用旋转、光照、模糊等多种图像变换训练检测器, 获得相应特征不变性, 使得该网络更适用于小规模数据集训练。通过在 3 个常用数据集上进行实验, 并与现有经典算法进行定性与定量的对比分析, 结果表明, 基于差分响应图的无监督特征点检测网络能较好地完成特征点检测任务, 所需训练时间更短、边缘定位更准确、数据集规模依赖性更低, 优化特征点检测性能。

关键词: 特征点; 无监督网络; 检测

中图分类号: TP391.4 文献标志码: A 文章编号: 1007-7162(2020)06-0063-08

An Unsupervised Feature Point Detection Network Based on Difference Response Graph

Lin Jing-yi, Li Dong, Hu Xiao-rui

(School of Automation, Guangdong University of Technology, Guangzhou 510006, China)

Abstract: In order to break through the performance limitations of hand-craft feature point detectors, a new data-driven unsupervised feature point detection network based on differential feature response graphs is proposed. The network calculates the differential output using convolution kernels of different scales, uses the absolute value of the differential response map to select a large number of feature points and judges the importance of these feature points. The detector is trained by using various image transformations such as rotation, illumination, and blur to obtain corresponding feature invariance, making the network more suitable for small-scale data set training. Through experiments on three commonly used data sets, and qualitative and quantitative comparison analysis with existing classic algorithms, the experimental results show that the proposed unsupervised feature point detection network based on the differential response graph can complete the feature point detection well. The network locates object edges more accurately, requires less training time, and has less dependency on dataset size. The network has optimized and improved feature point detection performance.

Key words: feature points; unsupervised network; detection

图像特征点检测是计算机视觉研究中的一个关键性任务。随着图像处理与模式识别技术的发展, 准确充分地提取原图蕴含信息的图像特征, 对物体识别、图像匹配、视觉跟踪、三维重建等课题的发展有着很大的裨益, 因此特征点检测是一项关键且基础的研究任务。

图像特征点指的是周围包含丰富的局部图像特征的像素点, 常见的特征类型有边缘点、角点、纹理、斑点等^[1]。传统的特征点检测方法包括Moravec算法、Harris算法、DoG算法^[2-4]等。Moravec算法考察局部图像平移前后的像素值差异, 并将差异最大的局部图像的中心像素作为特征点。然而该方法只具备弱旋

收稿日期: 2020-01-07

基金项目: 国家自然科学基金资助项目(61503084); 广东省自然科学基金资助项目(2016A030310348)

作者简介: 林璟怡(1994-), 男, 硕士研究生, 主要研究方向为图像处理、模式识别

通信作者: 李东(1983-), 男, 副教授, 博士, 硕士生导师, 主要研究方向为模式识别、机器学习、人脸识别和机器视觉, E-mail: dong.li@gdut.edu.cn

转不变性,其在实际任务中的性能表现较差。Harris算法^[5]计算局部图像方向导数得出图像特征变换最剧烈/缓和的两个方向,根据这两个方向的梯度值区分无纹理区域的点、边缘点或是角点。然而由于Harris算法对于图像的尺度较敏感,该算法在实际复杂环境下的图像特征提取应用中表现亦不佳。LoG和DoG算法将局部区域中的图像导数的极值点作为特征点,无法兼顾特征点的精确定位和图像的良好去噪,在实际应用中难以取舍。尺度不变特征转换方法(Scale-invariant Feature Transform, SIFT)^[6]算法使用高斯差分金字塔构造尺度空间,并在尺度空间中定位极值点以选取特征点。然而该算法无法对边缘光滑的目标准确提取特征点。加速稳健特征(Speeded Up Robust Features, SURF)^[7]算法对SIFT算法进行了改进,通过计算Hessian矩阵并构造尺度空间以进行特征点的选取与定位。然而SURF算法并未解决SIFT算法在尺度空间中各个层之间的尺度值不够紧密的问题。

除基于人工设计的特征点检测方法^[8-10],近年来也涌现出越来越多基于神经网络学习的特征点检测方法^[11-12]。然而现阶段缺少带有特征点标注的大规模数据集。其原因在于,首先,对于不同的任务场景,所需的最优特征点类型可能不同,即无法准确找到最优特征点对数据集进行标注。其次,对于密集的特征点生成来说,人工标注是一件非常困难的事情。这也使得很多基于学习的特征点检测器都依赖人工设计的检测器来生成训练数据,或将其作为网络的一部分一起参与训练。而后者将限制基于学习的检测器的性能进一步发展。为了解决以上问题,Quad-Network^[13]提出了一种基于数据驱动的无监督特征点检测方法,该思想也被LF-Net^[14]和SuperPoint^[15]作为特征点检测的重要部分来参与整个图像匹配任务网络的训练。以上工作说明了无监督的学习方法能够很好地完成特征点检测任务。

本文提出一种基于差分响应图的无监督特征点检测网络,主要的工作如下:

(1) 提出一种新的无监督特征点检测网络,训练特征点检测器对以每一个像素点为中心构成的局部图像块计算一个响应值。考察每一个图像块的响应值在当前图像的所有图像块中的大小关系,依据响应值大小使用非极大值抑制的方法在响应图中挑选特征点。在训练方式上,使用卷积核对整幅图像做卷积操作,省去了Quad-Network中对图像块进行随机选取、裁剪的步骤,减少了训练时间。

(2) 对特征点检测器前端进行优化,将简单线性滤波器替换为由不同尺度的卷积核进行差分运算得到的差分卷积核。在本文后续章节的分析与实验结果表明,利用差分响应有利于更精准地定位物体边缘,并缓解了物体边缘选取的特征点聚集的现象。

(3) 与现有方法相比,本文所提出的方法采用旋转、光照、模糊等多种图像变换训练检测器,获得相应特征不变性,该网络更适用于小规模数据集训练,摆脱对大数据集规模的依赖。实验结果表明,即便只使用具有足够丰富细节信息的单幅图像进行训练,并对图像进行合理变换,即可得到性能优良的特征点检测器。

1 本文方法

假设1 已知集合 X, Y 。若存在映射关系 F 与操作 H ,对于任意元素 $x, y(x \in X, y \in Y)$ 使得

$$y = F \otimes x \quad (1)$$

$$H(x) = H(y) \quad (2)$$

成立,则称操作 H 能匹配 X, Y 中任意一对具有 F 关系的元素,其中 $\otimes$ 符号表示映射操作。

1.1 主体思想

对于特征点检测任务来说,特征点的可重复性指的是:同一特征点可在该场景的不同图像中被检测器检测出来。特征点的可重复率是特征点检测任务中的一项重要指标。

根据假设1,假设存在两幅图像 I_i 和 I_j ,其映射关系符合单应矩阵 F_{ij} 。若存在完美的检测器 H ,其对任意一对匹配的像素点 $x \in I_i, y \in I_j$ 输出的值都应相等,此时所检测出的特征点的可重复率达到了极限值。然而,由于相机参数和噪声等原因,让所有的像素点都能正确匹配的概率极小。

综上所述,本文将特征点对于多对匹配的像素点响应值——对应相等的问题,近似为,多对匹配的像素点对特征检测器的输出响应值,在不同图像中的大小排序相近的问题,即将值比较问题转化成排序比较问题。基于该思想,检测器将对每一个以像素点为中心的局部图像块输出一个响应值,该响应值在整体图像所有响应值的排序位置决定该像素点的重要性程度。对于两幅同一三维场景下的不同图像来说,要求多对匹配点所输出的响应值的大小排序顺序相近。而响应值在全局范围内排序队列前端部分和末端部分的像素点,在不同的图像中重复率比排序队列中其他像素点更大,即选取响应值最大及

响应值最小的多对像素点作为特征点。

在检测器的选取问题上,本文使用不同尺度下卷积核的差分输出来计算响应值。并在后续章节中分析差分卷积核的相关性质。

1.2 网络结构及模型训练

图1展示了本文方法的网络架构,使用的深度学习框架为PyTorch。图1(a)为训练过程的步骤。输入图像 I 后网络将对其进行随机变换 t 并得到图像 $t(I)$,其次使用同一个差分卷积核模型将对两幅图像输出各自响应图。PyTorch中的评判器模块会根据预定义的损失函数计算得到误差梯度并反向传递回模型。参数更新模块会根据反向传递的误差梯度对模型的参数进行修改,最终得到训练完毕的模型将用于实际的测试与应用。

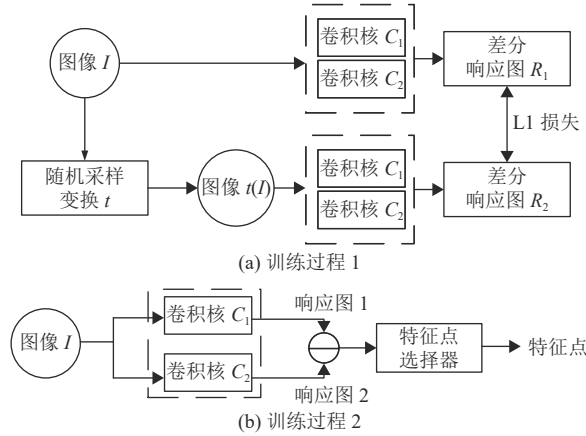


图1 本文网络架构

Fig.1 The proposed network architecture

如图1(b)所示,在测试过程中,使用训练后的检测器对图像输出一幅差分响应图,然后在全局范围内挑选响应值最大和最小的一部分像素点作为特征点。在选取的方式上使用常见的非极大值抑制方法。以选取的特征点的重复率评判特征点检测器的性能优劣。

1.2.1 采样随机变换与仿射变换

由于缺乏大规模的含有真实标注的数据集,在特征点检测任务中如何设计基于学习的网络是一件困难的事情。在使用反传网络来训练模型参数的时候,由真实标注和计算结果构成的误差会反向传播回模型并更新模型的参数以达到训练的效果。在特征点检测任务中除了需要考虑如何得到真实标注外,还需扩增训练的数据量以避免造成过拟合。本文使用采样随机变换的方法来获得真实的标注。

如图2所示,可对图像进行随机变换操作,例如

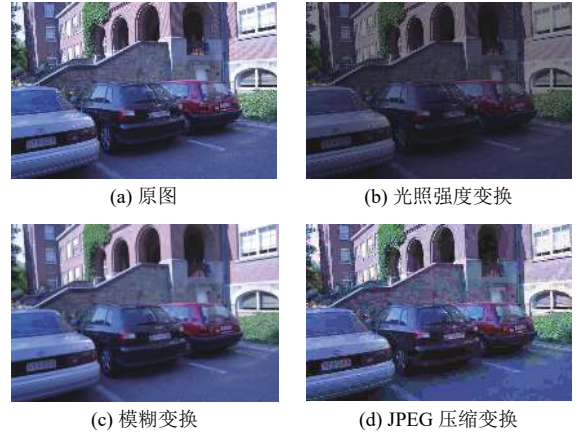


图2 原图与经过光照强度变换、模糊变换与JPEG压缩变换后的图像

Fig.2 The original image and its three transformed images (brightness, blur, JPEG compress)

光照强度变换、模糊变换等。该变换操作并不改变像素的相对位置,即二维图像每个像素点对应的三维位置信息是固定和已知的,并且所得到的新图像与原图的差异可用来训练模型获得相应的特征不变性。若模型对同一场景的不同图像的响应分布越相近,则说明这些图像之间包含的变换关系对该特征点检测网络的干扰影响越小,代表着这些图像变换的鲁棒性越强。而模型中的卷积操作并不会影响位置属性,于是图1(a)所示的模型输出的两幅差分响应图之间的位置信息也是一致的,即在训练过程中获得了对应的真实标注,而该标注信息将被用于计算网络后端的损失函数的输出。

由于本文使用小规模的数据来进行训练,需要进行数据扩增以防止训练过拟合。本文对原图及进行随机变换后的新图像采用随机的仿射变换。对图像进行仿射变换,相当于图像中每个像素点坐标乘以一个仿射变换矩阵,如式(3)所示。仿射变换包括一些常见的变换,例如平移、旋转、缩放等。可通过设置矩阵的参数 a_{11} 至 a_{23} 来实现特定的变换。式(4)与式(5)分别代表旋转与缩放操作,其中 θ 代表旋转角度, a_x 与 a_y 代表在两个坐标轴上的缩放尺度。使用仿射变换矩阵进行运算后,需将新坐标的第3个维度的值归一化,从而得到仿射变换后的坐标值 (x', y') 。

由于在同一个训练批次中仿射变换矩阵是已知的,可通过对网络模型输出的响应图进行逆仿射变换以使得新的响应图符合原图的位置分布。该操作执行后真实标注的对应情况不受影响。

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (3)$$

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (4)$$

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} a_x & 0 & 0 \\ 0 & a_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (5)$$

1.2.2 差分卷积核

与以往使用单个线性滤波器作为检测器的方法不同,本文使用两个不同尺度的卷积核输出的差分来计算得到响应图,如图3所示。不同尺度的卷积核表示感受野大小的不同。图3(a)为原图,图3(b)与图3(c)分别为原图经过较小尺度和较大尺度的卷积核运算后输出的结果,其中假设两个卷积核参数服从同一分布。在该可视化方法中,响应值越大的点所显示的颜色越浅。在图3(b)上可发现该选定的卷积核对建筑物纹理与树阴的响应值有较大不同,并且两者的分界较为清晰。而图3(c)也显示了建筑物纹理与树阴的响应值的不同,但两者的分界较为模糊。图3(b)和图3(c)表明线性卷积核对于物体的纹理的输出较为稳定。对两幅响应图进行批规范化操作,即将响应图规范化至符合特定均值与方差的分布,再对其进行差分运算,得到图3(d)。如图3(d)所示,建筑物纹理的颜色与树阴区域的颜色相近,表明了两者的响应值大小相近。而在物体边缘部分像素点的响应值将往最大及最小的趋势变化(如建筑物与树阴的分界区域响应值变大,建筑物与天空的分界区域响应值变小)。这是由于大尺度的卷积核对物体的边缘不敏感,在进行差分运算时,相比于边缘位置的像素点,非边缘像素点的响应会减去与其更为相近的值,而使得其差分响应接近于零,而边缘点的响应值会处于全局范围内最大及最小的部分。因此,本文方

法选取差分响应图中最大及最小的一部分作为特征点。

为了得到每一个像素点的响应值,首先根据二维卷积公式

$$f * g = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u, v) g(x-u, y-v) du dv \quad (6)$$

其离散二维卷积形式可写为

$$H_1(i, j) = \sum_{k=0}^{n_1} \sum_{l=0}^{n_1} I(k, l) C_1(i-k, j-l) \quad (7)$$

$$H_2(i, j) = \sum_{k=0}^{n_2} \sum_{l=0}^{n_2} I(k, l) C_2(i-k, j-l) \quad (8)$$

式(7)~(8)中,图像 $I \in \mathbb{R}^{h_1 \times h_2}$, 卷积核 $C_1 \in \mathbb{R}^{n_1 \times n_1}$, $C_2 \in \mathbb{R}^{n_2 \times n_2}$ 。 i 和 j 为像素点的坐标, n_1 和 n_2 分别为卷积核的尺寸, h_1 和 h_2 为图像的尺寸。

选取响应图 H_1 与 H_2 的差分作为输出,可以得到该网络的输出 $H(i, j)$ 为

$$H(i, j) = H_1(i, j) - H_2(i, j) \quad (9)$$

$$H(i, j) = \sum_{k=0}^{n_1} \sum_{l=0}^{n_1} I(k, l) C_1(i-k, j-l) - \sum_{k=0}^{n_2} \sum_{l=0}^{n_2} I(k, l) C_2(i-k, j-l) \quad (10)$$

式(6)~(10)为像素点的差分响应值计算步骤。检测器输出的差分响应图可简记为 $H(I|w)$, 其中 I 为输入的图像, w 为网络参数的权重。本文通过计算损失函数并进行误差反向传播,使用的深度学习框架在训练时会根据反向传播的误差值对卷积神经网络参数权重 w 进行自动更新。当网络参数训练完毕后,网络对输入的图像生成响应图,并根据响应值选取特征点。

1.2.3 损失函数

当图像经过某种非仿射变换后(如模糊、光照等),特征点检测器对变换前后的图像所输出的响应图的分布越接近,代表着检测器对于该变换的鲁棒性越强,其所选取的特征点的重复率也越高。

假设存在图像变换操作集 T , $t \in T$ 为随机采样的图像变换。图像 I_j 和 I_i 具有 t 映射关系,即 $I_j = t(I_i)$ 。利用检测器 H 计算响应图 $H(I_i|w)$ 和 $H(I_j|w)$ 。当 $H(I_i|w)$ 与 $H(I_j|w)$ 的L1距离越接近,则表示网络对 t 变换的鲁棒性越强。因此构造损失函数为

$$L(w) = \sum_{t \in T} \lambda_t \| \text{BN}(H(I_i|w)) - \text{BN}(H(t(I_i)|w)) \|_1 \quad (11)$$

其中BN代表Batch Normalization操作。 λ_t 为超参数,用于调整不同变换操作对损失函数的贡献。目前超

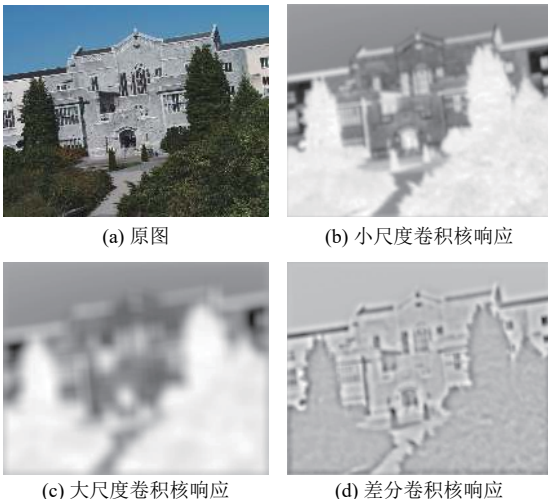


图3 原图及卷积核响应图

Fig.3 The original image and response maps

参数 λ_i 的值需通过多次实验尝试,并根据最优实验结果和图像数据集的属性决定。例如,假设数据集图像之间均存在 t_1 和 t_2 变换, t_1 的变换程度范围较大而 t_2 的变化程度较轻微,则可考虑调整 λ_{t_1} 与 λ_{t_2} 的比重以增大 t_1 变换对损失函数的贡献,使网络对 t_1 变换的鲁棒性更强。

本文图像变换操作集合 T 包括光照强度变换、模糊变换与JPEG压缩变换。光照强度变换计算方法如式(12)所示,其中 α 为线性的亮度衰减因子,取值范围为 $[0,1]$ 。当 $\alpha = 1$ 时代表图像不进行光照强度变换。

模糊变换的计算方法如式(13)~(15)所示,其中 e 为自然对数常数, σ 参数控制邻域像素点对中心点的影响, n_3 为模糊半径。模糊变换的计算形式与式(7)中离散二维卷积值的计算相似,不同之处为将卷积核换成二维离散高斯函数。

由于JPEG压缩算法的实现较复杂,本文直接使用OPENCV函数库工具对图像的JPEG压缩率进行调整。

$$B(i, j) = \alpha \cdot I(i, j) \quad (12)$$

$$G(i, j) = g(i, j) * I(i, j) \quad (13)$$

$$G(i, j) = \sum_{k=0}^{2n_3+1} \sum_{l=0}^{2n_3+1} I(k, l) \cdot g(i-k, j-l) \quad (14)$$

$$g(i, j) = \frac{1}{2\pi\sigma^2} e^{-\frac{(i-n_3-1)^2 + (j-n_3-1)^2}{2\sigma^2}} \quad (15)$$

2 实验分析

2.1 数据集

本文使用Oxford VGG数据库^[16]进行训练和测试分析,其中包含LEUVEN数据集、UBC数据集和TREES数据集。LEUVEN数据集中包含同一场景中不同光照强度的图像,用于测试特征点检测器的光照不变性。TREES数据集包含同一场景不同模糊程度的图像,UBC数据集包含同一场景不同压缩程度的图像,分别用于测试特征点检测器的模糊不变性与JPEG压缩损失不变性。每个数据集都提供了图像之间的单应矩阵,可用于计算特征点的重复率指标。

2.2 度量标准

本文实验的定量评价指标为特征点的可重复率,其数学描述见定义1。图4能直观地表示可重复率的意义。检测器对存在确定的仿射关系 F 的两图像进行特征点的选取,其中A点与C点在真实的三维空间中为同一点。若A点与C点能同时被检测器选取,



(a) 匹配图一

(b) 匹配图二

图4 特征点检测器在两幅图像选取的特征点

Fig.4 Feature points selected by detector in two images

并且检测的两点的坐标在像素误差阈值允许的条件下满足仿射关系 F ,则称特征点A是可重复的。特征点的可重复率即为在同一图像中可重复的特征点占所有检测出的特征点的比率。

然而只使用可重复率作为评价指标也具有局限性。例如当选取的特征点过于密集形成积聚时,计算的可重复率较高,但是该情况可能造成特征点在后续实际应用中性能不佳。因此本文同时对特征点的分布情况做出定性评价。

定义1 特征点的可重复率。若存在图像 I_i 与 I_j , F_{ij} 与 F_{ji} 分别为从 I_i 到 I_j 和从 I_j 到 I_i 的单应矩阵。 x_i 与 x_j 分别为在 I_i 与 I_j 中观测到的特征点。 $\tilde{x}_i$ 与 $\tilde{x}_j$ 为在 I_i 与 I_j 共有区域中观测到的特征点,其满足

$$\{\tilde{x}_i\} = \{x_i | F_{ij}x_j \in I_i\} \quad (16)$$

$$\{\tilde{x}_j\} = \{x_j | F_{ji}x_i \in I_j\} \quad (17)$$

特征点以误差 ξ 阈值的可重复率 $r(\xi)$ 的定义为

$$r(\xi) = \frac{|R(\xi)|}{\min(n_i, n_j)} \quad (18)$$

式(18)中 $n_i = |\{\tilde{x}_i\}|$ 和 $n_j = |\{\tilde{x}_j\}|$ 分别为 $\tilde{x}_i$ 和 $\tilde{x}_j$ 的数量, $R(\xi)$ 的定义为

$$R(\xi) = \{(\tilde{x}_i, \tilde{x}_j) | \text{dist}(H_{ij}\tilde{x}_i, \tilde{x}_j) < \xi\} \quad (19)$$

2.3 训练细节

本文使用Python语言来编写和训练特征点检测网络,使用的深度学习框架为PyTorch。训练集为LEUVEN数据集中图像细节最完备的一幅图像,测试集为前面所述3个数据集。本文选择无监督的Quad-Network网络和常用的DoG检测器作为对比方法。在Quad-Network算法的实现上,为了公平比较,本文使用该网络的训练方式和损失函数替换本文网络相对应的部分,其余训练操作均一致。本文网络使用Adam算法来优化网络权重,批量化大小(Batch

Size)为8,学习率为0.01。随机光照强度变换选取衰减因子的范围为[0.2,1]。随机模糊变换选取的模糊像素半径为[0, 5], σ 值设置为1。随机JPEG压缩率的选取范围为[0.2,1]。

本文方法使用仿射变换中的旋转与缩放操作进行数据增广。选取随机旋转角度的范围为[0,360],随机缩放因子 a_x 与 a_y 的范围均为[0.9,1.1]。

计算特征点可重复率的误差阈值对特征点检测器的性能做出了要求,当误差阈值越小时,要求检测器对特征点的定位要更精准。本文选取常用的5像素阈值来评判检测器的性能。

2.4 结果分析

图5为3种方法在LEUVEN数据集上的测试结果,红色和蓝色的点为所选择的特征点。图6展示DoG算法选取的候选点及Quad-Netwrok和本文方法输出的响应图。所输出的响应图采用灰度图的方式进行可视化。如图7所示,是3种方法分别在TREES和UBC数据集上选择的特征点。所有实验的特征点重复率数据如表1所示。

由表1可以看出,3种方法在UBC数据集中取得的重复率均较高,且与其余数据集中相差较大。这是由于LEUVEN数据集与TREES数据集中图像包含了仿射变换,而UBC数据集是使用同一幅图像在不同JPEG压缩条件下生成的图像,并不存在仿射变换。传统的DoG方法只有在LEUVEN数据集上的定量性

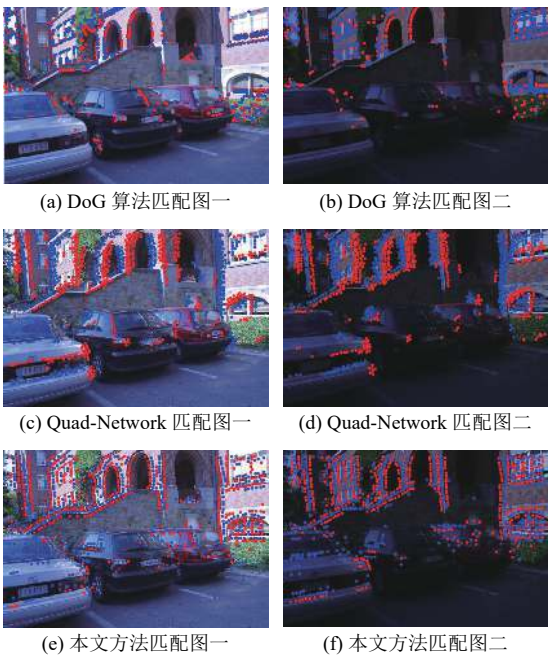


图 5 3种方法在LEUVEN数据集上检测的特征点

Fig.5 Feature points selected by DoG detector, Quad-Netwrok and our method on LEUVEN dataset

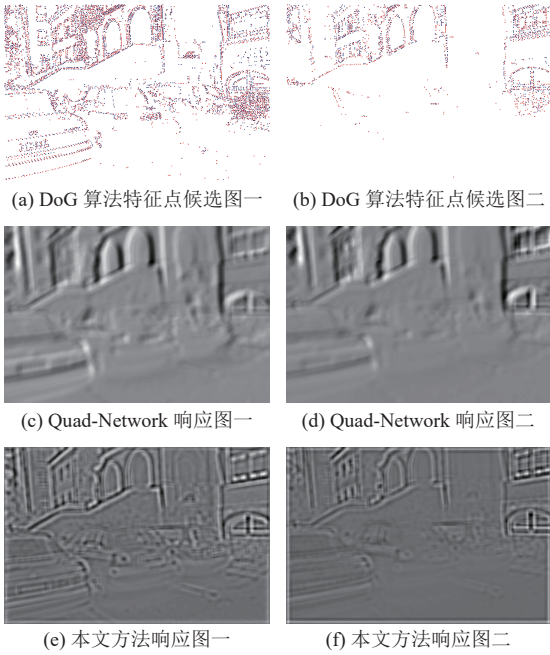


图 6 DoG, Quad-Netwrok和本文方法在LEUVEN数据集上的候选点/响应图

Fig.6 Candidate point/response diagrams of DoG detector, Quad-Netwrok and our method on LEUVEN dataset

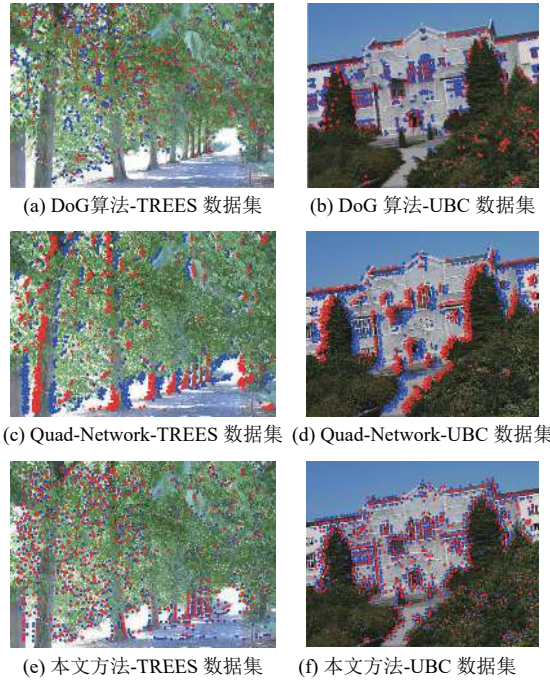


图 7 3种方法分别在TREES数据集(左)和UBC数据集(右)上选择的特征点

Fig.7 Feature points selected by DoG detector, Quad-Netwrok and our method on TREES and UBC datasets

能要比Quad-Netwrok和本文的方法要好。但是,结合图5(a)与图6(a)分析,DoG方法选择的特征点有很大部分聚集在树阴的区域。然而该区域中,点的位置难以辨认与定位,所选择的像素点在特征点检测任务

表 1 特征点重复率
Table 1 Repetition rate of feature points

数据库	实验算法	特征点数量		
		400	700	1000
LEUVEN	DoG	0.581	0.680	0.747
	Quad-Network	0.477	0.579	0.628
	Ours	0.499	0.589	0.651
TREES	DoG	0.227	0.324	0.403
	Quad-Network	0.287	0.356	0.420
	Ours	0.310	0.410	0.500
UBC	DoG	0.912	0.935	0.947
	Quad-Network	0.980	0.990	0.988
	Ours	0.934	0.955	0.948

中被认为是难样本(Hard Sample)。这是由于DoG方法总是选择图像离散函数二阶导数为零的点,即局部变化最剧烈的点,但在全局中该类型的点可能并非最优特征点。综上,虽然DoG检测器能很好地检测出物体的边缘,但也存在着选取大量难样本的风险。

如图5(c)和图5(d),Quad-Network方法对物体的边缘不够敏感,在边缘的响应值大小极为接近,在所测试的3个数据集上都存在特征点聚集情况。由图6(c)与图6(d)可看出,Quad-Network输出的响应图模糊现象及边缘点的积聚现象较为严重。虽然Quad-Network在UBC数据集上取得最高的重复率,但是结合对特征响应图分析,Quad-Network可能会选取远离图像中复杂纹理的像素点,并认为其与所处在复杂纹理的像素点具有相似的性能表现。这种特征点的选取策略可能会对后续任务的性能有损害,例如图像描述子生成任务等。

与DoG方法比较,本文所提出的方法在TREES数据集和UBC数据集中取得的重复率均高于前者。虽然在LEUVEN数据集上的重复率指标与前者相比较低,但缓解了DoG方法优先选择难样本的问题。与Quad-Network方法相比,所提出的方法在LEUVEN和TREES数据集上特征点的重复率均优于前者。此外,本文方法生成了对物体边缘更为敏感,细节更为清晰的特征响应图,见图6(e)与图6(f)。

图8为两种基于学习的方法在测试过程中对LEUVEN数据集检测特征点所耗费的时间曲线图。可以看出本文方法在同等特征点数目检测中所需要的时间远比Quad-Network的要少。因为Quad-Network在使用线性滤波器去对局部图像块计算相应的响应值时,存在着大量耗时的图像裁剪操作,而本文方法使用了不同于Quad-Network的训练模式与

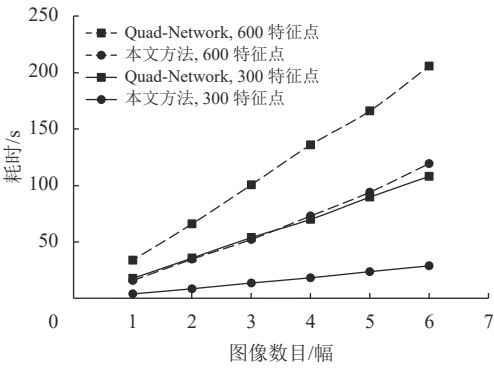


图 8 检测特征点的数量与所消耗时间的关系
Fig.8 The relationship between the number of feature points detected and the time consumed

损失函数,利用卷积核对图像输出特征图以候选特征点,省去了随机采样和裁剪图像块的操作,简化了数据处理流程,使得耗时大大减少。

图9是对本文网络中随机选取的一个尺度的卷积层进行可视化操作。训练得到的卷积核呈中心对称,且卷积参数的均值近似于零。由于未使用非线性的激活函数,当卷积核处于无纹理区域中,即图像像素值近似无变化的区域,其输出的响应值也接近于零;当卷积核处于有纹理区域中,输出的响应值会趋于正无穷及负无穷,远离零点。实验表明,选取响应值最大及最小的一部分像素点为特征点,具有较好的可重复性。

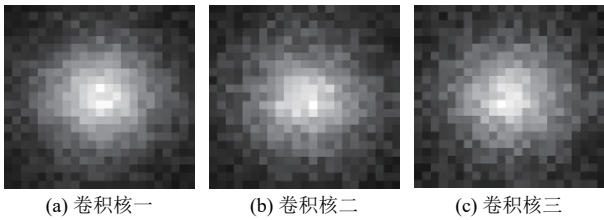


图 9 尺度为23像素的三通道卷积核可视化
Fig.9 Visualization of 3-channel convolution kernels with a scale of 23 pixels

综上,在分别对应光照强度变换、模糊变换和JPEG压缩变换的3个数据集的测试上,本文所提出的方法在3种方法中表现出最佳的综合性能。

3 结语

本文提出了一种新的数据驱动的基于差分特征响应图的无监督特征点检测网络。利用光照、模糊、压缩等变换来准确获取训练数据真实标注,利用随机仿射变换扩充数据集以避免训练过拟合,采用差分卷积核来代替传统的卷积核提取图像特征,获得对边缘更为敏感的特征图。采用了一种新的基于全

局图像的损失函数,简化了数据处理过程。使用定量与定性评价标准在3个数据集上将提出的方法与其他传统经典方法进行实验对比。实验结果表明,所提出的方法能有效地完成特征点检测任务,生成的特征响应图对物体边缘更敏感、细节更清晰,缩短了训练耗时,在小规模数据集上获得更优的检测效果,总体表现出最佳的综合性能。本文方法的局限性在于,未解决含有仿射变换的数据集所检测的特征点可重复率不高的问题,这也是下一步将要研究的问题。

参考文献:

- [1] SCHMID C, MOHR R, BAUCKHAGE C. Evaluation of interest point detectors [J]. *International Journal of Computer Vision*, 2000, 37(2): 151-172.
- [2] CANNY J. A computational approach to edge detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1986, 8(6): 679-698.
- [3] TRAJKOVIM, HEDLEY M. Fast corner detection [J]. *Image and Vision Computing*, 1998, 16(2): 75-87.
- [4] MATAS J, CHUM O, URBAN M, *et al.* Robust wide-baseline stereo from maximally stable extremal regions [J]. *Image and Vision Computing*, 2004, 22(10): 761-767.
- [5] HARRIS C G, STEPHENS M. A combined corner and edge detector[C]//*Proceedings of the 4th Alvey Vision Conference*. Manchester: Citeseer, 1988: 147-151.
- [6] LOWE D G. Distinctive image features from scale-invariant keypoints [J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [7] BAY H, TUYTELAARS T, VAN GOOL L. SURF: Speeded up robust features[C]//*European Conference on Computer Vision*. Heidelberg: Springer, 2006: 404-417.
- [8] 陈晓宁, 赵健, 杨润丰. 基于兴趣点检测的视觉关键词提取新方法研究[J]. *现代电子技术*, 2017, 40(21): 70-73.
CHENG X N, ZHAO J, YANG R F. Study on visual keywords extraction new method based on points of interest detection [J]. *Modern Electronics Technique*, 2017, 40(21): 70-73.
- [9] 李德隆, 刘伟. 基于改进的 SIFT 特征点的双目定位[J]. *广东工业大学学报*, 2017, 34(1): 90-94.
LI D L, LIU W. Object location for binocular stereo vision based on improved SIFT feature [J]. *Journal of Guangdong University of Technology*, 2017, 34(1): 90-94.
- [10] 叶志坚, 王福龙. 一种改进的 FREAK 算法的图像特征点匹配[J]. *广东工业大学学报*, 2017, 34(6): 37-42.
YE Z J, WANG F L. An improved FREAK algorithm for image feature point matching [J]. *Journal of Guangdong University of Technology*, 2017, 34(6): 37-42.
- [11] LENC K, VEDALDI A. Learning covariant feature detectors[C]//*European Conference on Computer Vision*. Amsterdam: Springer, 2016: 100-117.
- [12] VERDIE Y, YI K, FUA P, *et al.* TILDE: A temporally invariant learned detector[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston: IEEE, 2015: 5279-5288.
- [13] SAVINOV N, SEKI A, LADICKY L, *et al.* Quad-networks: unsupervised learning to rank for interest point detection[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Hawaii: IEEE, 2017: 1822-1830.
- [14] ONO Y, TRULLS E, FUA P, *et al.* LF-Net: learning local features from images[C]//*Advances in Neural Information Processing Systems*. San Francisco: Margan Kaufmann, 2018: 6234-6244.
- [15] DETONE D, MALISIEWICZ T, RABINOVICH A. Superpoint: self-supervised interest point detection and description[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. Salt Lake City: IEEE, 2018: 224-236.
- [16] MIKOLAJCZYK K, TUYTELAARS T, SCHMID C, *et al.* A comparison of affine region detectors [J]. *International Journal of Computer Vision*, 2005, 65(1-2): 43-72.

(责任编辑: 王威娜)