doi: 10.12068/j.issn.1005-3026.2017.07.004

基于多特征融合的图像区域几何标记

刘 威, 遇 冰, 周 婷, 袁 淮
(东北大学 研究院, 辽宁 沈阳 110819)

摘 要: 提出一种基于多特征融合的图像区域几何标记方法. 首先,提出了一种新型卷积网络结构——多尺度核卷积网络用于提取像素点的多尺度特征信息,推断像素点的几何类别,并结合图像超像素分割获得图像超像素区域的几何标记;其次,将提取的多尺度特征与超像素区域传统特征相结合,建立超像素区域的特征表达. 最后,建立超像素图像的条件随机场(conditional random field, CRF)模型,对超像素区域的几何类别进行推断. 在公开数据集 Geometric Context(GC)上的实验结果表明,同已有算法相比,所提方法提高了图像区域几何标记的准确率.

关 键 词: 多特征融合;多尺度核卷积网络;图像区域几何标记;特征学习;条件随机场模型
中图分类号: TP 391 **文献标志码**: A **文章编号**: 1005-3026(2017)07-0927-05

Geometric Labeling of Image Regions Based on Combination of Multiple Features

LIU Wei, *YU Bing*, *ZHOU Ting*, *YUAN Huai*
(Research Academy, Northeastern University, Shenyang 110819, China. Corresponding author: LIU Wei, E-mail: lwei@neusoft.com)

Abstract: A geometric labeling method of image regions was proposed based on combination of multiple features. First of all, according to the requirement of multi-scale feature information extraction, a novel network structure—multi-scale kernel convolutional network (MSKCN) was proposed. The multi-scale feature information was used for inferring geometric label of pixel. The geometric labeling of super-pixel regions with the image super-pixel segmentation was achieved. Then a feature representation of super-pixel regions was established by combining multi-scale features proposed and traditional features of super-pixel regions. Finally, a CRF (conditional random field) model was constructed for the super-pixel image to infer geometric label of super-pixel regions with the image super-pixel segmentation. The experiments on public database Geometric Context (GC) indicated that the accuracy of geometric labeling was improved by using the proposed method compared with the existing state-of-art.

Key words: combination of multiple features; multi-scale kernel convolutional network; geometric labeling of image regions; feature learning; conditional random field model

图像区域几何标记是将图像中各个区域标记为不同几何类别的过程,天空(sky)、立体物(vertical)以及地面(support)是三类常见的几何类别. 图像区域几何标记结果常被应用在自动驾驶系统中道路、自由驾驶空间检测等领域[1-2],从而得到广泛的研究.

近年来许多研究者对图像区域几何标记进行了深入的探究. Hoiem 等通过提取图像中多种特征信息作为推理线索,对图像区域的几何标签进行推理标记[3]. Gould 等则通过在图像的外观和结构上建立统一的能量函数,并通过最大化该能量函数获得图像区域的几何标签[4]. Alvarez 等则利用卷积神经网络学习图像特征信息,并与所提出的基于颜色平面融合的纹理特征相融合进行道

收稿日期: 2016-01-01
基金项目: 国家自然科学基金资助项目(61273239); 中央高校基本科研业务费专项资金资助项目(N151802001).
作者简介: 刘 威(1975-),男,辽宁沈阳人,东北大学副教授.

路区域的推理[5]. Lazebnik 等建立了一种非参数化的方法来进行区域几何标记[6],他们随后结合马尔科夫随机场(Markov random field, MRF),利用图像整体和区域匹配的方法进行了类似的工作[7].

自然场景通常具有较高的复杂性,上述方法在标记几何类别时,没有考虑场景中各类别之间的上下文联系以及立体物的深度信息,导致上述方法对图像区域几何类别的标记准确性不高. 为此,本文提出了一种新型的网络结构——多尺度核卷积网络(multi-scale kernel convolutional network, MSKCN),解决立体物深度信息不同带来的影响. 针对图像区域几何标记问题设计了一个整体可训练框架,用于针对性地学习特征表达和分类,并结合图像超像素分割结果获取超像素区域的几何类别标记. 进一步地,利用本文提出的 MSKCN 网络作为特征提取器所获取的特征,与文献[3]中所使用的超像素区域传统特征相结合,建立超像素图像的条件随机场模型[8],推断超像素区域的几何类别. 最后,通过实验验证了所提特征以及模型对图像区域几何标记的有效性.

1 多尺度核卷积网络

1.1 多尺度特征信息提取

卷积神经网络具有较强的特征学习和表达能力,近年来在计算机视觉、语音识别、自然语言处理等多个领域都取得了很好的效果[9]. 由于物体深度不同,导致在图像中呈现出“近大远小”的特点,其所具有的特征也可能出现在不同的尺度下. 这要求在提取目标物的特征时,需要在不同尺度下对图像进行处理,从而获得不同尺度下的特征信息.

对目标物的多尺度特征提取通常有两种方法:

第一种方法对目标物图像进行缩放,通过使用同种尺寸的卷积窗口对不同尺寸的图像进行卷积,获得相应尺度下的特征信息. 若目标物图像为 I,卷积核为 K,则目标物在尺度 s 下的特征 F_s 为

$$F_s = K * I_s. \tag{1}$$

其中,I_s 为目标物图像根据尺度 s 缩放后的图像.

第二种方法不改变目标物图像,但使用不同尺寸的卷积窗口对图像进行卷积,获得不同尺度下的特征信息. 这时,目标物在尺度 s 下的特征 F_s 为

$$F_s = K_s * I. \tag{2}$$

其中,K_s 为根据尺度 s 缩放后的卷积核.

在传统卷积神经网络中,由于其特征提取并不改变目标物图像的尺度,所以属于第二种方法. 其位于同一卷积层上的卷积核尺寸是相同的,对于不同尺度特征信息的提取是通过多层卷积和降采样完成的. 令位于第 n 层的特征在原图像上所对应的尺度是 s_n,则位于第 $n+1$ 层的特征在原图像上所对应的尺度 s_{n+1} 为

$$s_{n+1} = s_n + (k_{n+1} - 1)\prod_{i=1}^{n} d_i. \tag{3}$$

其中:k_n 和 k_{n+1} 分别为第 n 和 $n+1$ 层的卷积 - 降采样窗口尺寸;d_i 为第 i 层的卷积 - 降采样步长,$i = 1, 2, \cdots, n$.

可以看出,传统卷积神经网络是在网络的不同深度上提取不同尺度的特征信息. 按照常规的前馈方法,小尺度信息是无法直接被利用的. 文献[7]通过将各中间层获得的特征图传递到全连接层从而获得多尺度特征表达. 但这样做会使得各个尺度的特征信息维数不均,并且造成全连接层的参数量大幅上升,大大增加网络训练难度和对样本的需求量. 基于以上原因,需要设计一种新型的网络结构使得在参数量没有大幅增加的前提下,将多尺度的特征信息直接应用到全连接层,从而使得依据多尺度线索对实际问题进行推理成为可能.

1.2 多尺度核卷积网络

为了提取目标物多尺度下的特征信息,并且不引入过多的参数,本文提出一种新型的网络结构——多尺度核卷积网络. 此种网络的结构如图 1 所示.

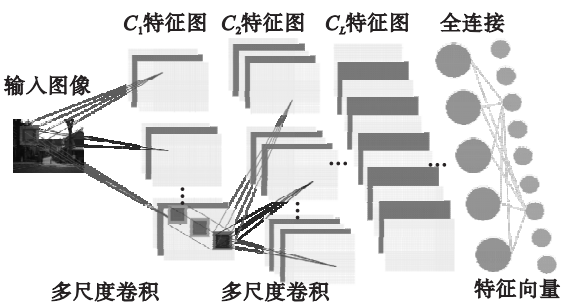

图 1 本文提出的多尺度核卷积网络

Fig. 1 Proposed multi-scale kernel convolutional network

此种网络结构具有如下特点:

1) 网络的同一卷积层内包含有多种尺寸的卷积核. 这样的卷积层可以在同一层网络中提取不同尺度的特征信息. 这相当于直接应用了 1.1 节中提到的第二种提取多尺度特征信息的方法.

2）由于网络可以同时提取多尺度特征信息，所以不再需要 N_s 层网络去提取 N_s 种尺度的特征信息，降低网络的深度.

3）由于多尺度特征信息直接传入全连接层，所以全连接层的参数量不会大幅提升，保证了训练过程的可行性.

若 $l-1$ 层网络的输出特征图组为 $\boldsymbol{F}^{l-1}$，位于卷积层 l 的卷积核 K_i^l 对输入特征图组的作用为

$$\boldsymbol{F}_i^l = f^l(K_i^l * \boldsymbol{F}^{l-1} + b_i^l) . \tag{4}$$

式中：b_i^l 为卷积核 K_i^l 相对应的偏置项；$f^l(\cdot)$ 为 l 层网络的激活函数. 若卷积层 l 共包含 N_l 个卷积核，则输入特征图经过该层后获得的特征图组为

$$\boldsymbol{F}^l = [F_1^l, F_2^l, \cdots, F_{N_l}^l] = [f^l(K_1^l * \boldsymbol{F}^{l-1} + b_1^l), \cdots, f^l(K_{N_l}^l * \boldsymbol{F}^{l-1} + b_{N_l}^l)] . \tag{5}$$

这里，卷积核 K_i^l 为三维卷积核，且对输入的特征图组中的每一张特征图均进行卷积操作：

$$K_i^l * \boldsymbol{F}^{l-1} = [K_{i1}^l * F_1^{l-1}, K_{i2}^l * F_2^{l-1}, \cdots, K_{iN_{l-1}}^l * F_{N_{l-1}}^{l-1}] . \tag{6}$$

其中，$K_{it}^l (t = 1,2,\cdots, N_{l-1})$ 为卷积核的第 t 层的二维卷积核. 位于卷积层 l 的三维卷积核的层数与 $l-1$ 层网络的特征图张数 N_{l-1} 相等.

设第 n 层有 p 张特征图，第 n 层特征在原图像上所对应的尺度为 $\boldsymbol{S}^n = [S_1^n, S_2^n, \cdots, S_p^n]$，则第 $n+1$ 层特征在原图像上所对应的尺度为 $\boldsymbol{S}^{n+1}$，设这一层有 q 个卷积核，卷积核尺度为 $\boldsymbol{k}^{n+1} = [k_1^{n+1}, k_2^{n+1}, \cdots, k_q^{n+1}]$，步长为 d^{n+1}，则 $\boldsymbol{S}^{n+1}$ 可以表示为

$$S_j^{n+1} = \max_{l=1}^{p}(S_l^n) + (k_j^{n+1} - 1)\prod_{i=1}^{n} d_i . \tag{7}$$

其中，$j = 1,2,\cdots,q$.

将上面对于 MSKCN 的特征尺度分析结果与传统卷积神经网络的特征尺度进行对比，可以看出，MSKCN 在其卷积层的每一层均可以同时提取出多个尺度的特征信息. 对于多尺度特征信息的提取，其所需网络深度更小，参数量更小，训练过程也更简单. 与文献[10]相比，本文所提网络结构的全连接层需要的神经元个数更少，更容易避免过拟合.

2　基于多特征融合的图像区域几何标记

2.1　基于 MSKCN 的图像区域几何标记

为了对每个像素点都能给出相应的多尺度特征信息，本文不对原图像进行任何缩放，直接使用原图像进行处理. 同时，为了保持原图像中各个像素点邻域的特征信息，网络不使用降采样层，从而不降低图像的分辨率. 本文使用图 1 中的两层卷积层构成的网络作为特征提取器，两层的卷积核个数分别为 8 和 15，则由第二个卷积层输出的像素点 i 的多尺度特征向量 $\boldsymbol{f}_{i\_\mathrm{ms}}$ 为 15 维. 记全连接层输入像素点 i 的特征向量为 $\boldsymbol{f}_{i\_\mathrm{fc}} = [\boldsymbol{f}_{i\_\mathrm{ms}}, f_{i\_y}]$，其中 f_{i_y} 为相应像素在图像中垂直方向（y 轴）的归一化位置信息，$f_{i\_y} = Y_i/h$，Y_i 为像素点 i 的垂直方向坐标，h 为输入图像的高度. 然后，通过全连接层进行分类推理，给出每个像素点 i 属于三类几何类别的概率，记为 $p_{ij}\{j = 1,2,3\}$，分别对应属于天空、立体物和地面的概率.

为了获取图像区域的几何标记，本文利用文献[3]提供的超像素分割结果，统计各个超像素区域内所有像素点属于 3 个几何类别的概率均值，记超像素区域 s 的几何类别概率均值为 $\boldsymbol{\mu}_s = [\mu_{s1}, \mu_{s2}, \mu_{s3}]$，其中 μ_{sj} 表示超像素区域 s 中所有像素点属于第 $j\{j = 1,2,3\}$ 类几何类别的概率值的均值，计算公式如下：

$$\mu_{sj} = \frac{1}{M}\sum_{i=1}^{M} p_{ij} . \tag{8}$$

其中，M 为超像素区域 s 包含的像素点个数.

根据式(8)计算 $\boldsymbol{\mu}_s$ 的表达，本文将其中最大值 μ_{sj} 的下标索引对应的几何类别作为超像素区域 s 的几何类别标记.

2.2　基于多特征融合的图像区域几何标记模型

为利用图像上下文信息提高图像区域几何标记的准确度，受文献[11]利用条件随机场模型对卷积神经网络输出标记概率进行平滑思想的启发，本文提出将 MSKCN 与 CRF 相结合的图像区域几何标记模型，如图 2 所示.

首先，为文献[3]提供的超像素分割所得的超像素区域 s 建立特征表达式：$\boldsymbol{f}_s = \{\boldsymbol{f}_{\mathrm{tra}}, \boldsymbol{f}_{s\_\mathrm{ms}}\}$，其中 $\boldsymbol{f}_{\mathrm{tra}}$ 表示文献[3]中所用的传统特征，$\boldsymbol{f}_{s\_\mathrm{ms}} = \{\boldsymbol{\mu}_{\mathrm{ms}}, \boldsymbol{\sigma}_{\mathrm{ms}}\}$ 为对 MSKCN 网络中 2.1 节所描述的特征提取器所提特征处理所得的超像素区域的特征信息，$\boldsymbol{\mu}_{\mathrm{ms}}$ 和 $\boldsymbol{\sigma}_{\mathrm{ms}}$ 分别表示超像素区域 s 中所有像素点由 2.1 节中特征提取器获取的 15 维特征向量 $\boldsymbol{f}_{i\_\mathrm{ms}}$ 在每个维度上的均值和标准差，计算公式如下：

$$\boldsymbol{\mu}_{\mathrm{ms}} = \frac{1}{M}\sum_{i=1}^{M}\boldsymbol{f}_{i\_\mathrm{ms}} , \tag{9}$$

$$\boldsymbol{\sigma}_{\mathrm{ms}} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(\boldsymbol{f}_{i\_\mathrm{ms}} - \boldsymbol{\mu}_{\mathrm{ms}})^2} . \tag{10}$$

本文利用文献[3]提供的超像素分割结果生成超像素图像,并结合上述建立的超像素特征表达,对超像素图像建立条件随机场模型,从而推断超像素区域的几何标签. 本文建立的条件随机场模型中一元和二元势函数分别为

$$\varphi_i(\boldsymbol{\alpha}, y_i, \boldsymbol{x}_i) = \boldsymbol{\alpha}\boldsymbol{x}_i,$$
$$\varphi_{i,j}(\boldsymbol{\beta}, y_i, y_j, \boldsymbol{x}_{i,j}) = \boldsymbol{\beta}\boldsymbol{x}_{i,j}. \quad (11)$$

式中:$\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 分别为一元势函数、二元势函数中的权重系数矩阵;$\boldsymbol{x}_i = [\boldsymbol{f}_{s\_ms}, 1]^{\mathrm{T}}$ 表示第 i 个超像素区域的观测向量;$\boldsymbol{x}_{i,j}$ 表示相邻两超像素区域间的二元差异性表达向量;y_i 表示第 i 个超像素区域的类别标号;$\boldsymbol{x}_{i,j}$ 计算公式为

$$\boldsymbol{x}_{i,j} = \begin{bmatrix} |\boldsymbol{x}_i - \boldsymbol{x}_j| \\ 1 \end{bmatrix}. \quad (12)$$

其中,$|\boldsymbol{x}_i - \boldsymbol{x}_j|$ 表示相邻两个超像素区域观测向量之间的距离,本文采用卡方距离度量该值.

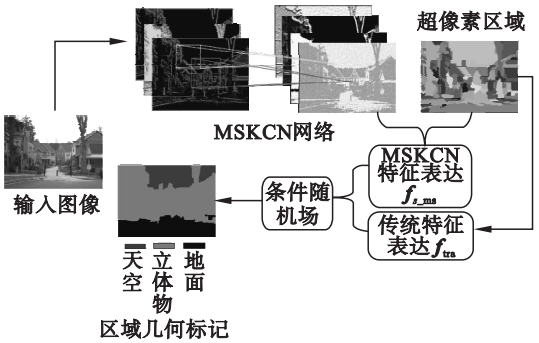

图 2 基于多特征融合的模型

Fig. 2 Model combined with multiple features

3 实验结果与分析

3.1 实验设置及数据集

为进行图像区域的几何标记,并与领域内现有方法进行对比,本文使用 GC 公开数据集[3]进行实验. 该数据集包括 300 张来自不同自然场景的图像,本文按照文献[3]的实验数据集划分方式进行 5 次交叉验证实验.

3.2 实验结果与分析

为证明本文所提出的 Hoiem[3] 特征 + MSKCN + CRF 模型(2.2 节所描述的模型)在图像区域几何标记问题上的有效性,本文与现有的方法 Hoiem[3]、Gould[4]和 Lazebnik[6]进行对比. 为使对比实验公平,所有对比方法均使用 GC 数据集进行实验,5 次交叉验证实验结果平均值如表 1 所示.

从表 1 的实验结果中可以看出,使用本文所提 MSKCN 获取的特征结合传统特征[3]作为推断线索,为超像素图像建立 CRF 模型实现图像区域几何类别的标记,较已有方法提高了标记准确率,说明本文模型对图像区域几何标记的有效性. 同时,Hoiem[3] 特征 + CRF 算法与 MSKCN + CRF 算法的实验结果对比,说明由本文提出的基于 MSKCN 的特征提取器所提取的多尺度特征,对图像区域几何标记具有有效性.

表 1 不同图像区域几何标记方法识别效果对比

Table 1 Comparison of different geometric labeling methods of image regions

方法	平均准确率/%
Hoiem[3]	88.1
Gould[4]	86.9
Labznik[6]	84.0
Hoiem[3] 特征 + CRF	85.0
MSKCN	86.7
MSKCN + CRF	87.3
Hoiem[3] 特征 + MSKCN + CRF	89.4

本文提出的 Hoiem[3] 特征 + MSKCN + CRF 模型以及文献[3]算法的混淆矩阵如图 3 所示. 从图 3 中可以看出,本文所提出的模型与文献[3]相比提高了图像区域几何标记中各类别的识别准确率,降低了误识别率,本文模型在 GC 数据集上的效果示例如图 4 所示.

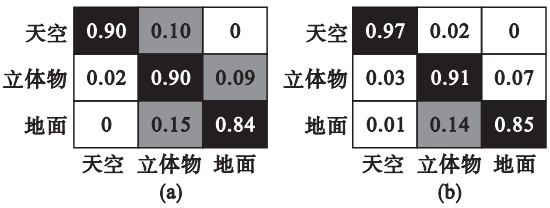

图 3 GC 数据集混淆矩阵结果

Fig. 3 Confusion matrix of GC database

(a)—Hoiem 算法[3];

(b)—本文 Hoiem[3] 特征 + MSKCN + CRF 模型.

本文所提出的多尺度核卷积网络结构简单,计算性能较高. 对于一张 640 × 480 的输入图像,使用未优化的 MATLAB 代码在配备 Intel Core2 Duo E7500 CPU 和 2 GB RAM 的 PC 平台上单核运行大约需要 1.2 s,结合条件随机场模型对图像区域进行几何标记所需时间大约为 10 s,而 Hoiem 等[3]在相应尺寸图像上运行时间则为 11.5 s. 同时,所提出的网络结构易于进行并行处理或分布式处理,计算性能的加速空间很大,根据参考文献[12]的经验,加速比可高达 10 倍以上.

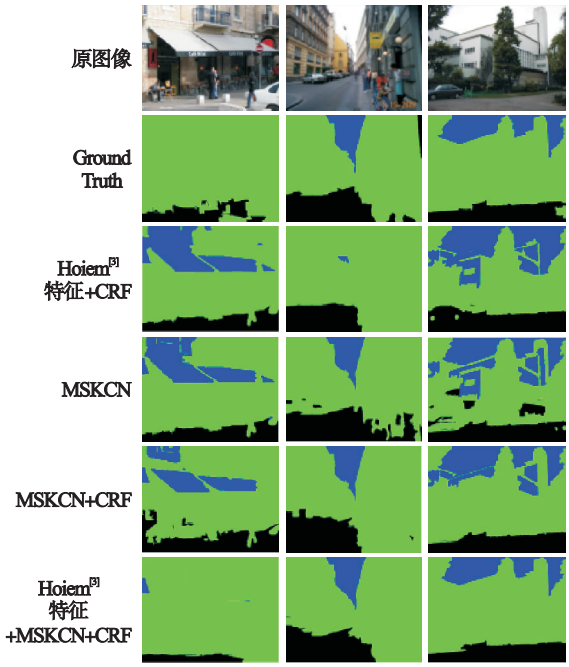

图 4　各算法图像区域几何标记结果(蓝色为天空，绿色为立体物，黑色为地面)

Fig. 4　Geometric labeling results of image regions with multiple methods(sky is blue, vertical is green, support is black)

4　结　　论

本文提出一种基于多特征融合的图像区域几何标记方法. 该方法利用提出的多尺度核卷积网络提取像素多尺度特征信息，推断像素点几何类别，并结合超像素分割获得图像超像素区域的几何标记. 随后，通过结合传统特征，建立超像素区域的特征表达，并为超像素图像建立条件随机场模型，结合图像上下文信息对超像素区域的几何类别进行推断. 在公共数据集上的实验结果表明了本文方法的有效性.

参考文献：

[1] Alvarez J M, Gevers T, Lopez A M. 3D scene priors for road detection[C]// IEEE Conference on Computer Vision and Pattern Recognition (CVPR). San Francisco, 2010: 57 – 64.

[2] Alvarez J M, Lopez A M, Gevers T. Combining priors, appearance, and context for road detection [J]. *IEEE Transaction Intelligent Transportation Systems*, 2014, 15(3): 1168 – 1178.

[3] Hoiem D, Efros A A, Hebert M. Recovering surface layout from an image [J]. *International Journal of Computer Vision*, 2007, 75(1): 151 – 172.

[4] Gould S, Fulton R, Koller D. Decomposing a scene into geometric and semantically consistent regions [C]// Intelligent Conference on Computer Vision. Kyoto: IEEE, 2009: 1 – 8.

[5] Alvarez J, Gevers T, Lopez A. Road scene segmentation from a single image [J]. *Lecture Notes in Computer Science*, 2012, 75(1): 376 – 389.

[6] Lazebnik S, Raginsky M. An empirical Bayes approach to contextual region classification [C]// Computer Vision and Pattern Recognition. Miami: IEEE, 2009: 2380 – 2387.

[7] Tighe J, Lazebnik S. SuperParsing: scalable nonparametric image parsing with superpixels [C]// European Conference on Computer Vision. Crete: IEEE, 2010: 352 – 365.

[8] Lafferty J, McCallum A, Pereira F. Conditional random fields: probabilistic models for segmentation and labeling sequence data [C]// Proceedings International Conference on Machine Learning. San Francisco, 2001: 282 – 289.

[9] Lecun Y, Bengio Y, Hinton G. Deep learning [J]. *Nature*, 2015, 521(7553): 436 – 444.

[10] Sermanet P, LeCun Y. Traffic sign recognition with multi-scale convolutional networks [C]// International Joint Conference on Neural Networks. San Jose: IEEE, 2011: 2809 – 2813.

[11] Farabet C, Couprie C, Najman L, et al. Learning hierarchical features for scene labeling [J]. *IEEE Transaction Pattern Analysis and Machine Intelligence*, 2013, 35(8): 1919 – 1929.

[12] Krizhevsky A, Sutskever I, Hinton G. ImageNet classification with deep convolutional neural networks [C]// International Conference on Neural Information Processing Systems. South Lake Tahoe: IEEE, 2012: 1106 – 1114.