基于级联全卷积神经网络的显著性检测

张松龙*, 谢林柏**
江南大学物联网工程学院, 江苏 无锡 214122

摘要 提出了一种级联全卷积神经网络的显著性检测方法。网络主要由两层级联的全卷积神经网络组成,第一阶段构建了一个带金字塔池化模块编码-解码架构的全卷积神经网络,金字塔池化模块有效抑制了背景噪声的干扰。第二阶段设计了边缘检测网络,学习显著区域的边缘信息,通过融合两个阶段显著图得到边界精确的显著图。实验结果表明,所提方法在图像显著性检测数据集 ECSSD 和 SED2 上均具有较高的准确率、召回率和较低的平均绝对误差,为目标识别、机器视觉等提供了可靠的预处理结果。

关键词 机器视觉; 显著性检测; 级联全卷积神经网络; 金字塔池化模块; 边缘检测网络
中图分类号 TP391.41 **文献标识码** A **doi**: 10.3788/LOP56.071501

Salient Detection Based on Cascaded Convolutional Neural Network

Zhang Songlong*, Xie Linbo**
School of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu 214122, China

Abstract A saliency detection method is proposed based on a cascaded full convolutional neural network. This network is mainly composed of two full convolutional neural networks. In the first stage, a full-convolutional neural network with a pyramid pooling module encoding and decoding architecture is constructed, and the pyramid pooling module can be used to effectively suppress the interference of background noises. In the second stage, an edge detection network is designed to learn the edge information of a salient region, and the accurate boundary saliency map is obtained by the fusion of two-stage saliency maps. The experimental results show that the proposed method has high accuracy, high recall rate, and low average absolute error in image significance detection dataset ECSSD and SED2, which provides the reliable pretreatment results for target recognition, machine vision and other applications.

Key words machine vision; saliency detection; cascaded full convolution neural network; pyramid pooling module; edge detection network
OCIS codes 150.0155; 330.7326; 110.2970

1 引 言

视觉显著性检测可以快速地寻找图像中的重要区域,已经成为计算机视觉中用于降低计算复杂度的重要预处理步骤。目前,显著性检测算法被广泛运用于图像缩放[1]、图像压缩[2]、目标识别[3]、图像分类[4]等计算机视觉任务中。由于显著性检测受显著对象的尺度、背景和位置等不确定因素的影响,是目前计算机视觉领域中的一个难题。在传统方法[5-7]中研究者们根据观察到的对比度、边缘先验、中心先验等各种先验知识进行显著性检测,生成相应的显著性图。但在复杂的场景中,这些观察经常受限于颜色、对比度等低级特征,而不能准确反映出显著性对象本质的共性,因此传统方法通常无法达到预期效果。近年来,卷积神经网络(CNN)在计算机视觉中得到广泛应用,其中在语义分割[8]、图像分类[9]、边缘检测[10]等领域都获得了重大的突破。与传统方法不同,深度卷积神经网络能够自动地从大

收稿日期: 2018-08-06; **修回日期**: 2018-08-27; **录用日期**: 2018-11-20
基金项目: 国家自然科学基金(61374047,60973095)
* **E-mail**: 6161905052@vip.jiangnan.edu.cn; ** **E-mail**: xielb@126.com

量的样本中学习得到高级特征，从而有效地避免了人工建模和人工设计特征的弊端。因此，推动了卷积神经网络应用在显著性检测领域的研究。文献[11]通过融合丰富的上下文信息学习显著性区域；文献[12]通过多尺度编码上下文信息的深度卷积网络进行显著性对象检测；文献[13]研究了循环全卷积神经网络，将粗糙的预测作为显著先验，逐步优化每个阶段显著值，最终融合生成显著图。虽然这些方法相对传统显著性检测方法效果有了大幅度提高，但是缺少全局上下文信息，因而导致整个显著对象区域混杂背景噪声。同时由于深度网络中大量的池化操作导致图像细节信息丢失，因此这些方法检测的显著区域存在边缘模糊问题。针对这两个问题，本文提出一种级联全卷积神经网络显著性检测方法。该方法通过在编码-解码的网络中引入金字塔模块获取全局信息，有效地抑制了背景的干扰，同时通过边缘优化网络获取局部信息，优化显著区域边缘。

2 基于级联全卷积神经网络显著性对象检测

2.1 网络结构

如图 1 所示，本文的网络结构由两个阶段组成。第一阶段为编码-解码架构的全卷积神经网络，编码器网络是基于 VGG16（Visual Geometry Group 16）[14]舍弃全连接层的 13 层卷积层，同时嵌入一个金字塔池体系结构。解码器是一个可训练的卷积网络，用于对编码器的输出进行上采样，并对显著检测区域的细节进行微调，从而得到一个初步的显著图。第二阶段由卷积层组成的全卷积网络，通过 6 个卷积层提取局部语义特征，从而学习显著区域的边缘。最终，通过跳跃连接的方式融合两个阶段的输出得到最终的显著图。图中 Conv 为卷积，BN 为批量归一化，PPM 为金字塔池化模块，GT 为人工标注图。

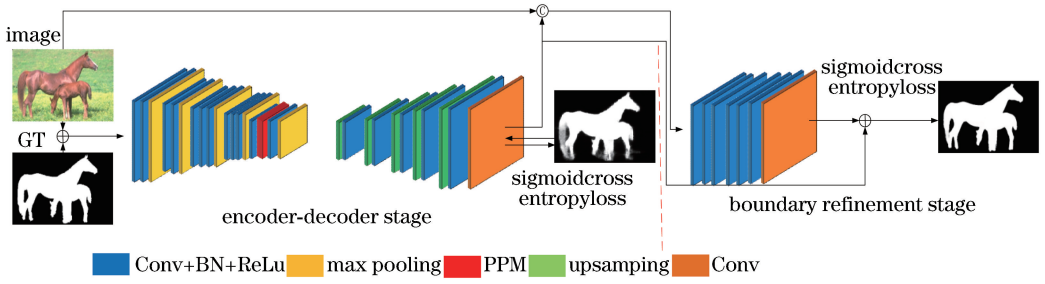

图 1 网络结构图

Fig. 1 Structural diagram of network

2.2 编码-解码的全卷积神经网络

第一阶段的网络由一个编码-解码架构的全卷积神经网络构成。编码器全卷积网络（FCN）从原始图像中提取视觉特征，而解码器 FCN 逐步通过上采样提取编码特征，最终映射到原图大小进行像素级分类。

编码：特征提取是像素级分类的前提和核心。因此，强大的编码器网络非常重要。本文采用 VGG16[14]网络作为编码器特征提取网络，它在图像分类[14]、语义分割[8]等方面均具有突出的效果。VGG16 模型由 13 个卷积层和 3 个全连接层组成，它的卷积层被分为 5 个阶段，在每个阶段的后面连接内核为 2、步长为 2 的最大池化层。同时，在每个卷积层后连接 BN 层和激活函数（ReLu）层。与原始 VGG 网络不同，本文删除全连接层，因为这些层参数量大并且不利于像素级分割。此外，最重要的变化是本文设计了一个金字塔池化模块收集全局上下文信息，有效地去除了显著区域中的背景噪声。

现有的如 FCN[8]和 DeepLab[15]等卷积神经网络只能独立地预测每个像素，而不考虑每个感受野的上下文信息，限制了不同场景理解的能力。这导致在复杂的场景中，检测的显著区域经常混有背景区域。在语义分割中，Zhao 等[16]利用空间金字塔池化层聚合不同区域的上下文信息，从而提高获取全局信息的能力。金字塔池化模块如图 2 所示，本文将 VGG16 中的 conv5_1 的高级卷积特征作为金字塔模块的输入。为了增加金字塔模块的感受野，conv5_1 采用空洞率为 2 的空洞卷积。然后，使用 1×1、2×2 和 4×4 三种不同的金字塔尺度从输入特征图中提取不同的子区域特征，形成不同位置的池化表示。金字塔池化模块不同级别的输出包含不同大小的特征图。在每个金字塔池层之后使用 1×1 卷积运算，以减少相应上下文的维度并保持全局特征的权重。然后构建上采样层，通过双线性插值

获得与 conv5_1 相同尺度的特征映射。最后，不同级别的特征图被连接为混合多尺度上下文全局信息的特征，作为下一卷积层的输入。金字塔池化模块效果如图 3 所示，经过金字塔池化模块有效地获取全局信息，从而去除显著区域的背景噪声。

解码：为了减少网络参数并提高速度，采用相对较小的解码器结构。如图 1 所示，解码器由 5 个卷积层、5 个上采样层和 1 个显著性预测层组成。同样，在每个卷积层后依次连接 BN 层和 ReLu 层。

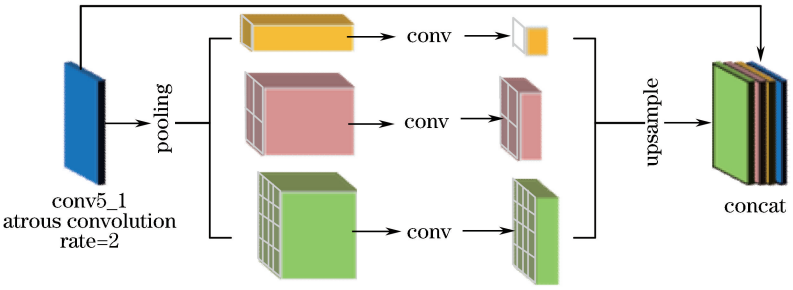

图 2 金字塔池化模块结构

Fig. 2 Structural diagram of pyramid pooling module

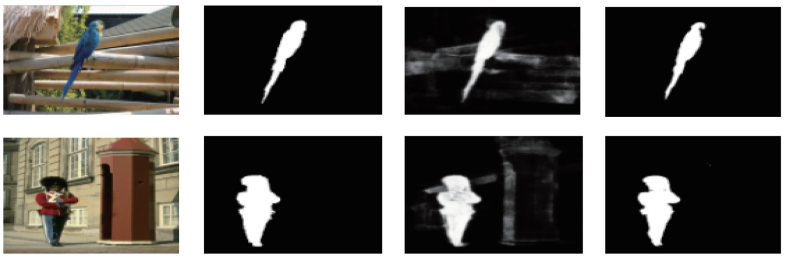

图 3 显著图效果比较。(a)原图；(b) GT；(c)无 PPM 模块；(d)有 PPM 模块

Fig. 3 Effect comparison of saliency maps. (a) Original images; (b) GT; (c) without PPM module; (d) with PPM module

2.3 边界优化网络

通过第一阶段包含空间金字塔池化的编码-解码网络的预测，获得了相对突出的显著区域。但是受深度网络影响，其需要大量运用最大池化操作，导致显著对象的边缘等细节信息丢失，从而显著性检测区域边缘模糊。因此，采用局部边缘优化网络优化显著区域的边缘。边缘优化阶段的网络结构如图 1 中的边界优化阶段(BRS)所示。通过串联原图的 RGB 图和第一阶段的显著图得到四通道数据，将它作为 BRS 网络的输入。BRS 网络是由 6 个卷积层组成的全卷积神经网络，前 3 个卷积层后面各自连接 1 个非线性 ReLu 层。为了在输入和输出特征映射之间保持相同的分辨率和保存细节信息，未在卷积层后引入池化层。但是由于未使用池化层，网络的感受野太小，因此在 BRS 中采用多网格[17]空洞卷积，增加网络的感受野。BRS 各卷积层的参数如表 1 所示，每个卷积核为 3×3，步长为 1。最终，通过“跳跃连接”的模式连接第一、二阶段的输出，经融合得到最终的显著图。

表 1 BRS 的网络参数

Table 1 Network parameters of BRS

Layer	Channel	Kernel size	Dilation rate
1	64	(1+3)×64×3×3	1
2	64	64×64×3×3	1
3	64	64×64×3×3	1
4	128	64×128×3×3	1
5	128	128×128×3×3	2
6	128	128×128×3×3	3
7	1	128×1×3×3	1

3 仿真实验分析

3.1 数据库和对比方法

使用公开数据集 MSRA10K[18]作为训练集。该数据集包含 10000 张高像素图片，并且大多数图片中只有一个显著目标。为了增加训练图片的多样性，通过镜像和旋转 0°、90°、180°、270°产生了 80000 张图片作为训练集。采用两个广泛使用的图像显著性检测数据集 ECSSD[19]和 SED2[20]来验证本文方法的有效性。其中 ECSSD 数据集共 1000 张图像，

含有像数级别的真值标注。该数据集图像包含一个或者多个显著性对象，具有复杂的背景干扰；SED2数据集共100张图片，含有像数级别的真值标注。虽然该数据集图片规模较小，但是背景相对复杂而且均为多显著目标，具有极高的挑战性。

为验证所提方法的优越性，将其与目前主流的9种显著性检测算法进行对比。这9种主流方法分为两类：1）引导学习（BL）[6]和稳健背景检测（wCO）[7]这两个公认的准确度较高的传统方法；2）多尺度深度特征（MDF）[12]、循环全卷积神经网络（RFCN）[13]、深度对比学习（DCL）[21]、深度层级显著网络（DHS）[22]、聚合多层次卷积特征（Amulet）[23]、不确定卷积特征（UCF）[24]、阶段优化模型（SRM）[25]等基于深度学习的显著性检测算法。

实验基于64位的Ubuntu16.04操作系统和英伟达GTX Genforce 1080 GPU，内存为8 G，软件有Matlab2014a、Python2.7，深度学习框架为Caffe[26]。两个网络均使用梯度下降（SGD）方法训练网络，设置动量为0.9，权重衰减为0.0001，设置基础学习率为10^{-8}。本文共训练22 h，单张图片的测试速度为12 frame/s。

3.2 视觉直观比较

视觉直观比较的结果如图4所示，以图4(b)作对比，图4(c)中所提方法生成的显著图更接近GT图，在复杂的场景中仍能生成准确的显著性图。从图4可以看出，所提方法由于整合了全局上下文信息和局部边缘信息，不仅有效地去除了背景区域，而且保留了完整的显著区域边缘信息，因此明显优于其他算法的显著图。

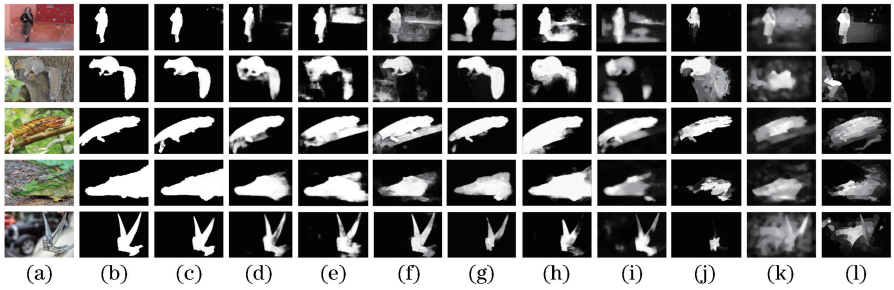

图4 显著图比较。(a)原图；(b) GT；(c)所提方法；(d) SRM算法；(e) Amulet算法；(f) UCF算法；(g) DCL算法；(h) DHS算法；(i) RFCN算法；(j) MDF算法；(k) BL算法；(l) wCO算法

Fig. 4 Comparison of saliency maps. (a) Original images; (b) GT; (c) proposed algorithm; (d) SRM algorithm; (e) Amulet algorithm; (f) UCF algorithm; (g) DCL algorithm; (h) DHS algorithm; (i) RFCN algorithm; (j) MDF algorithm; (k) BL algorithm; (l) wCO algorithm

3.3 准确率-召回率曲线

为进一步验证所提方法的有效性，采用准确率-召回率（*P*-*R*）曲线评价算法性能。如图5所示，所提方法在两个数据集上的准确率-召回率均达到最高。传统方法（包括BL、wCO）在处理复杂场景时，由于采用手工选取的特征进行检测，因而具有较低的准确率和召回率。同时，所提方法包含全局上下文信息和局部边缘信息，在两个数据集上均获得了最高的查准率和查全率。在查全率接近1时，查准率急剧下降，说明该算法能够准确地寻找显著区域边缘。

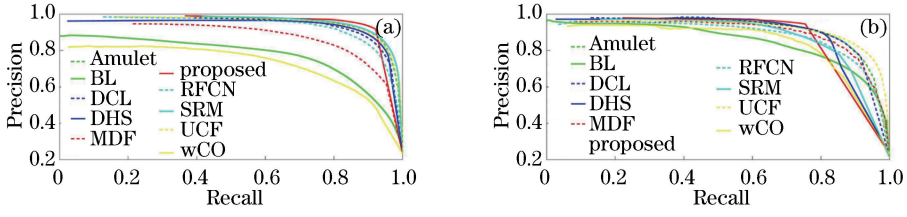

图5 不同方法在不同数据集的*P*-*R*曲线。(a) ECSSD数据集；(b) SED2数据集

Fig. 5 *P*-*R* curves of different methods in different datasets. (a) ECSSD dataset; (b) SED2 dataset

3.4 *F* 值和平均绝对误差

为更进一步验证所提方法的有效性，本文运用*F*值（*F*-measure）和平均绝对误差（MAE）在两个数据集上进行对比。*F*值是通过设定自适应阈值而

得到，其表达式为

$$F=\frac{(1+\beta^2)P\times R}{\beta^2\times P+R},\tag{1}$$

式中 P 和 R 分别代表查准率和查全率，β^2 与文献[14-16]一致取值为 0.3，此时的 F 值更能突出不同算法的优越性。F 值越高，表明算法检测显著区域的准确度越高。MAE 是以像素点为单位计算显著图与 GT 图之间的平均绝对误差，即

$$V_{\mathrm{MAE}}=\frac{1}{M\times N}\sum_{x=1}^{M}\sum_{y=1}^{N}|S(x,y)-G(x,y)|,\tag{2}$$

式中：M，N 为显著图的长和宽；$G(x,y)$ 为 GT 图在 (x,y) 处的值；$S(x,y)$ 为显著图 (x,y) 处的值；MAE 值表明显著图与 GT 图之间的相似程度，MAE 值越小，算法得到的显著图与 GT 图越接近，性能越高。表 2 总结了本文方法与常见显著性检测算法在两个数据集上的 F 值和 MAE 值比较。表 2 中 Base-net 为基础网络，即编码-解码的基础网络，BASE-Net+PPM 为增加了 PPM 模块的第一阶段网络，Base-net+BRS 为含有两个阶段但无 PPM 模块的网络。通过对比发现，PPM 模块和 BRS 对显著图的 MAE 和 F-measure 两个指标均具有明显的提升。从表 2 可以看出，所提方法在两个数据库上的 F 值较高、MAE 值较小，表明了所提算法的优越性。

表 2 两种性能指标的大小对比

Table 2 Size comparison of two performance indicators

Algorithm	ECSSD		SED2	
	MAE	F-measure	MAE	F-measure
wCO	0.171	0.651	0.131	0.694
BL	0.216	0.684	0.189	0.705
MDF	0.105	0.833	0.115	0.801
RFCN	0.107	0.834	0.113	0.767
DHS	0.060	0.871	0.079	0.822
DCL	0.074	0.827	0.101	0.795
UCF	0.078	0.841	0.086	0.810
Amulet	0.059	0.869	0.080	0.871
SRM	0.056	0.892	0.076	0.894
Base-net	0.073	0.817	0.082	0.787
Base-net+PPM	0.064	0.842	0.079	0.834
Base-net+BRS	0.057	0.869	0.075	0.861
Proposed	0.052	0.898	0.070	0.902

4 结 论

提出了级联全卷积神经网络显著性检测方法。一方面，在编码-解码网络中引入金字塔模块，获取不同尺度的全局信息，有效地抑制了背景噪声的干扰。另一方面，通过局部边缘优化网络，学习显著区域的边缘信息，获得精确边界的显著图。实验结果表明，相比于现有的算法，该方法在各项评测指标上均得到明显提高。

参 考 文 献

[1] Fang Y M, Chen Z Z, Lin W S, *et al*. Saliency detection in the compressed domain for adaptive image retargeting[J]. IEEE Transactions on Image Processing, 2012, 21(9): 3888-3901.

[2] Gao R, Tu Q, Xu J, *et al*. Visual saliency detection based on mutual information in compressed domain [C]// Visual Communications and Image Processing, 2015: 1-4.

[3] Ren Z X, Gao S H, Chia L T, *et al*. Region-based saliency detection and its application in object recognition[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2014, 24(5): 769-779.

[4] Sharma G, Jurie F, Schmid C. Discriminative spatial saliency for image classification [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2012: 3506-3513.

[5] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11): 1254-1259.

[6] Tong N, Lu H C, Ruan X, *et al*. Salient object detection via bootstrap learning [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1884-1892.

[7] Zhu W J, Liang S, Wei Y C, *et al*. Saliency optimization from robust background detection[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2014: 2814-2821.

[8] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.

[9] Bi L H, Liu Y C. Plant leaf image recognition based on improved neural network algorithm[J]. Laser & Optoelectronics Progress, 2017, 54(12): 121102.
毕立恒，刘云潺. 基于改进神经网络算法的植物叶片图像识别研究[J]. 激光与光电子学进展, 2017, 54(12): 121102.

[10] Liu Y, Cheng M M, Hu X W, *et al*. Richer convolutional features for edge detection[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2017: 5872-5881.

[11] Zhao R, Ouyang W, Li H S, *et al*. Saliency detection by multi-context deep learning[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1265-1274.

[12] Li G B, Yu Y Z. Visual saliency based on multiscale deep features[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2015: 5455-5463.

[13] Wang L Z, Wang L J, Lu H C, *et al*. Saliency Detection with Recurrent Fully Convolutional Networks[C]// European Conference on Computer Vision, 2016: 825-841.

[14] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-13)[2018-07-25].

[15] Chen L C, Papandreou G, Kokkinos I, *et al*. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.

[16] Zhao H S, Shi J P, Qi X J, *et al*. Pyramid scene parsing network[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6230-6239.

[17] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation[J]. arXiv preprint arXiv:1706.05587, 2017.

[18] Cheng M M, Zhang G X, Mitra N J, *et al*. Global contrast based salient region detection[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2011: 409-416.

[19] Wang L J, Lu H C, Ruan X, *et al*. Deep networks for saliency detection via local estimation and global search[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3183-3192.

[20] Borji A. What is a salient object? A dataset and a baseline model for salient object detection[J]. IEEE Transactions on Image Processing, 2015, 24(2): 742-756.

[21] Li G B, Yu Y Z. Deep contrast learning for salient object detection[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2016: 478-487.

[22] Liu N, Han J W. DHSNet: Deep hierarchical saliency network for salient object detection[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2016: 678-686.

[23] Zhang P P, Wang D, Lu H C, *et al*. Amulet: aggregating multi-level convolutional features for salient object detection[C]// IEEE International Conference on Computer Vision, 2017: 202-211.

[24] Zhang P P, Wang D, Lu H C, *et al*. Learning uncertain convolutional features for accurate saliency detection[C]// IEEE International Conference on Computer Vision, 2017: 212-221.

[25] Wang T T, Borji A, Zhang L H, *et al*. A stagewise refinement model for detecting salient objects in images[C]// IEEE International Conference on Computer Vision, 2017: 4039-4048.

[26] Jia Y, Shelhamer E, Donahue J, *et al*. Caffe: convolutional architecture for fast feature embedding[C]// 22nd ACM International Conference on Multimedia, 2014: 675-678.