

## 基于深度学习的手绘草图识别

赵鹏<sup>1,2</sup>,王斐<sup>2</sup>,刘慧婷<sup>1,2</sup>,姚晟<sup>1,2</sup>

(1. 安徽大学 计算智能与信号处理教育部重点实验室,安徽 合肥 230039;

2. 安徽大学 计算机科学与技术学院,安徽 合肥 230601)

**摘要:** 现有的手绘草图识别方法严重依赖于费时费力的手工特征提取,而经典的深度学习模型主要是为彩色多纹理自然图像设计,用于识别手绘草图时效果不甚理想。提出一种基于深度学习的手绘草图识别方法(Deep-Sketch),该算法根据手绘草图缺失颜色、纹理信息的特点,使用大尺寸的首层卷积核取代自然图像识别中常使用的小尺寸首层卷积核,获得更多的空间结构信息。利用训练浅层模型获得的模型参数来初始化深层模型对应层的模型参数,以加快收敛,减少训练时长。加入不改变特征大小的卷积层来加深网络深度等方法以减小错误率。实验结果表明,所提出的方法较之其它几种主流的手绘草图识别方法具有良好的正确率,对250类手绘草图识别正确率达到69.2%。

**关键词:** 手绘草图识别;深度学习;卷积神经网络

**中图分类号:** TP391.41

**文献标志码:** A

### Sketch Recognition Using Deep Learning

ZHAO Peng<sup>1,2</sup>, WANG Fei<sup>2</sup>, LIU Huiting<sup>1,2</sup>, YAO Sheng<sup>1,2</sup>

(1. Key Lab. of Intelligent Computing and Signal Processing of Ministry of Education, Anhui Univ., Hefei 230039, China;

2. School of Computer Sci. and Technol., Anhui Univ., Hefei 230601, China)

**Abstract:** In order to solve the existing problem of the sketch recognition heavily relying on the manual feature extraction which is very time-consuming, a method of sketch recognition based on deep learning, called Deep-Sketch, was proposed. The classical deep learning models were mainly designed for natural color image recognition which failed on the sketch recognition. Deep-Sketch aimed to obtain more spatial structure information by using the large-size convolution kernel instead of the small-size convolution kernel in the first convolution layer. In addition, a shallow model was trained to obtain parameters which were used to initialize the corresponding layer parameters of the Deep-Sketch to reduce the model training time. Deep-Sketch was deepened with the convolution layers which kept the feature size to reduce the error rate. The results showed that the Deep-Sketch is superior to other state-of-the-art sketch recognition methods and achieves 69.2% accuracy on the sketch dataset including 250 classes.

**Key words:** sketch recognition; deep learning; convolution neural network

手绘草图是人类最直观且最原始的沟通工具。随着智能触摸屏技术的发展和相关设备的普及,人们可以方便地绘制草图。近年来,手绘草图的相关研究,包括手绘草图识别<sup>[1-2]</sup>,基于手绘草图的图像检索<sup>[3]</sup>和基于手绘草图的3D模型检索<sup>[4]</sup>等。其

中,手绘草图识别就是利用分类模型对人类绘制的草图自动赋予一个类别关键词的过程。与自然图片相比,手绘草图没有颜色和纹理信息,一般为二值图像或灰度图像,具有高度的抽象性和夸张性,并且存在由于用户绘制过程中的停顿和不连贯导致草图轮

收稿日期:2015-07-20

基金项目:国家自然科学基金资助项目(61472001;61202227);安徽省自然科学基金项目(1408085MF122;1508085MF127);安徽省高校自然科学基金研究重点项目(KJ2016A041);安徽大学信息保障技术协同创新中心公开招标课题(ADXXBZ2014-5;ADXXBZ2014-6)

作者简介:赵鹏(1976—),女,副教授,博士。研究方向:智能信息处理;机器学习。E-mail:zhaopeng\_ad@163.com

网络出版时间:2016-4-22 11:39:44 网络出版地址: <http://www.cnki.net/kcms/detail/51.1596.T.20160422.1139.007.html>

<http://jsuese.scu.edu.cn>

廓不完整等问题,使得手绘草图的识别成为一个非常具有挑战性的问题。

目前的研究主要是利用传统的图像分类模式进行手绘草图的识别,即手工提取特征并进行分类标注。首先将图像表达成一组描述符,如 SIFT<sup>[5]</sup> 和 HOG<sup>[6]</sup> 用于特征提取。然后使用无监督算法,主要是通过聚类学习出字典,建立基于字典的图像特征描述符。最后,利用有监督的学习和词算法对训练集中带标签的图像特征集进行训练学习获得分类器。例如,Eitz 等<sup>[1]</sup> 使用 SIFT 袋模型。Li 等<sup>[7]</sup> 对每幅草图整合多种低层特征并建立星图。Schneider 等<sup>[2]</sup> 使用 SIFT 和基于高斯混合的 fisher vector。另外,还有基于轮廓特征的手绘草图识别方法<sup>[8]</sup>、基于区域特征的手绘草图识别方法、基于骨架特征的手绘草图识别方法和基于图像熵的手绘草图识别方法<sup>[9]</sup> 等,这些方法的普遍问题在于需要手工去处理和提取特征,费时费力且依靠经验。与手工规则构造特征的方法相比,利用深度学习模型直接从大数据中学习特征,更有利于描述数据本身的丰富内涵信息。

深度学习中的卷积神经网络(ConvNets)最近在语音分析<sup>[10]</sup>、语音识别<sup>[11]</sup>、大规模的图像和视频识别中取得了巨大的成功<sup>[12-15]</sup>,这归功于:1)大量带标签训练集的出现;2)高性能计算系统,如 GPU<sup>[16]</sup> 和大规模分布式集群的出现;3)更好模型调优策略,如使用更小的卷积核<sup>[13,15]</sup>,使用稠密连接取代稀疏连接<sup>[14]</sup>,增加网络深度来减小模型误差<sup>[17-19]</sup> 和 Dropout<sup>[20]</sup> 策略等。

目前,深度学习主要应用于识别彩色多纹理自然图像<sup>[14,21]</sup>,由于手绘草图缺少颜色、纹理信息,且

存在轮廓不完整等问题,这些模型在手绘草图识别中的效果并不理想。作者提出了基于深度学习的手绘草图识别方法(Deep-Sketch):1)使用大尺寸的首层卷积核取代自然图像识别中常使用的小尺寸首层卷积核。由于自然图像的纹理较丰富,使用小卷积核能够提取到细节的变化,而手绘草图缺少这些纹理信息,当使用小卷积核时,会导致大部分提取到的特征为空,使得传递给第2层卷积核的信息较少,信息丢失过快,不能很好地获得草图的结构信息。2)训练浅层模型作为深层模型对应层的初始值,并加入不改变特征大小的卷积层,其大小为  $3 \times 3$ ,步长为1,填充边界为1,不仅加深网络深度、减小模型误差,而且减少训练时长,加快收敛。

### 1 卷积神经网络

从 LeNet-5<sup>[22]</sup> 开始,典型的卷积神经网络通过一系列隐层,得到长度为  $C$  的1维概率向量  $\hat{y}_i$ ,向量中的每个概率分布对应  $C$  种不同的类。每层主要包含以下4部分:1)卷积层,每个卷积图都由前一层网络的输出结果与学习获得的卷积核进行卷积运算产生(对于第一层来说,前一层输出结果就是输入图片);2)矫正层(rectified linear units, ReLUs),对每个卷积结果都进行矫正运算  $\text{ReLU}(x) = \max(0, x)$ ;3)最大池化层(Max-Pooling),对矫正运算结果选取最大值,此层为可选层;4)局部采样标准化层(local response normalization, LRN)<sup>[12]</sup>,对降采样图进行对比度归一化操作,使得输出特征平稳,此层为可选层。最后是全连接层,输出层是一个 softmax 分类器。

图1为该模型的网络结构示意图。

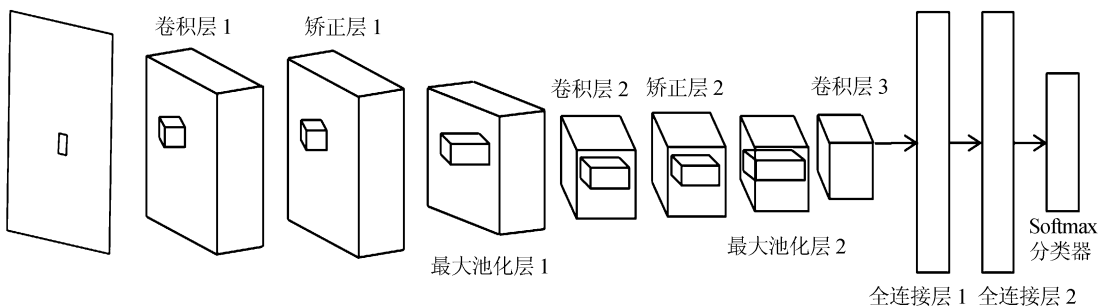


图1 卷积神经网络网络结构示意图

Fig.1 Standard structure of convolutional neural networks

#### 1.1 卷积层

卷积层中,上一层的特征图会和当前卷积层的卷积核进行卷积,卷积的结果加权求和后,经过非线性函数处理得到这一层的特征图。这样每个神经元

可以提取输入图像的结构特征,如方向、线段、端点、角点等。卷积操作计算如式(1)所示。

$$y^{i(r)} = \max(0, b^{i(r)} + \sum_i k^{ij(r)} * x^{i(r)}) \quad (1)$$

式中,  $x^i$  和  $y^j$  表示第  $i$  个输入特征和第  $j$  个输出特征,  $k^{ij}$  表示  $x^i$  和  $y^j$  之间的卷积核,  $*$  表示卷积操作,  $b^j$  表示第  $j$  个输入特征的偏置系数,  $r$  表示卷积区域。采用非线性函数 ReLU 作为激活函数, ReLU 被证明比 sigmoid 函数的拟合效果要好<sup>[10]</sup>。

设当前层包含  $M$  个大小为  $(M_x, M_y)$  的特征图, 卷积核大小为  $(K_x, K_y)$ , 卷积核在  $x, y$  方向上的步长分别为  $S_x$  和  $S_y$ , 为了得到整数结果, 可对当前层添加大小为  $p$  的边界, 卷积的结果取卷积核能完全落在图片内的部分, 则卷积后特征图  $(M_x, M_y)$  大小如式(2) 计算所得。

$$\begin{cases} M_x^n = \frac{M_x^{n-1} + 2p - K_x^n}{S_x^n} + 1, \\ M_y^n = \frac{M_y^{n-1} + 2p - K_y^n}{S_y^n} + 1 \end{cases} \quad (2)$$

式中,  $n$  表示当前层数,  $n - 1$  表示前一层。

## 1.2 最大池化层

池化层中的采样操作无重叠地采集上一层特征图同一大小的子区域, 取最大值作为输出结果。通过降低特征图的分辨率并抽样选取优秀的特征, 使得模型具有更好的抗噪能力, 同时降低了特征维度。最大池化层的一般形式为如式(3) 所示。

$$y_{j,k}^i = \max_{0 \leq m, n < s} \{x_{j+s+m, k+s+n}^i\} \quad (3)$$

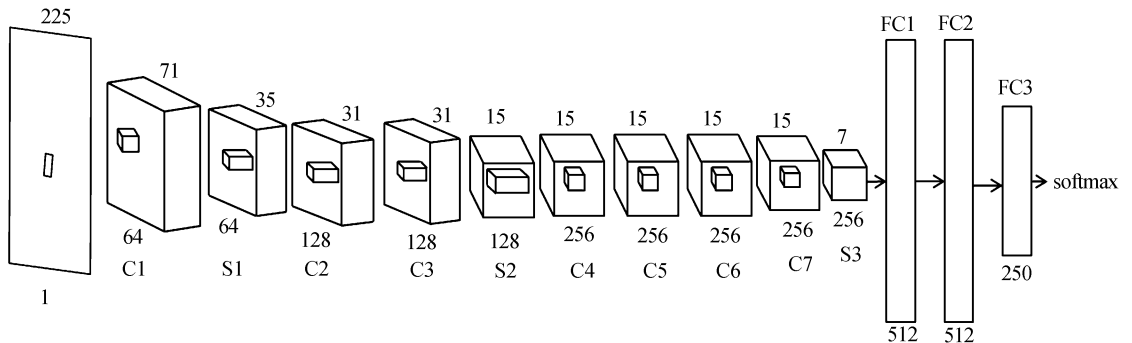


图2 Deep-Sketch 网络结构示意图

Fig.2 Illustration of the architecture of Deep-Sketch

## 2.2 更大的第1层卷积核

由于第一卷积层距离原始图像最近, 参数也最敏感, 后续的操作都依赖于其输出。现有的方法倾向于使用更小的首层卷积核, 如  $11 \times 11$ <sup>[12]</sup>、 $7 \times 7$ <sup>[14]</sup>, 乃至  $3 \times 3$ 。由于自然图像的纹理较丰富, 使用小卷积核能够提取到细节的变化, 更适合提取诸如眼睛、纽扣等纹理信息<sup>[14]</sup>。但是手绘草图缺少这些纹理信息, 当使用小卷积核时, 会导致大部分提取到的特征为空, 使得传递给第2层卷积核的信息较少, 信息丢失过快, 不能很好地获得草图的结构信

息, 第  $i$  个输出特征  $y^i$  中的每个神经元在一个  $s \times s$  不重合的区域进行采样操作。输出特征大小计算方法与卷积层相同。

## 2 Deep-Sketch 模型

将分别从全模型网络结构、更大的第1层卷积核、减少过拟合、重叠采样层、加深网络深度、取消LRN、初始化权值和使用更大的 dropout 几个方面介绍所提出的 Deep-Sketch 模型。

### 2.1 全模型

Deep-Sketch 首先使用7层卷积层(C1 ~ C7), 每层后加上一个矫正层, 同时第C1、C3、C7层后加上最大池化层(S1 ~ S3)。第8个卷积核大小为  $7 \times 7$ , 与前一 pooling 层的大小相同, 因此认为是全连接层。Deep-Sketch 使用3个全连接层。最后一层输出单元数目为类别的数目, 由于实验采用的草图集包括250类, 因此最后一层输出250个单元, 依次对应相应的类别, 采用 softmax loss 函数作为损失函数。卷积核数目从64开始, 每经过一次最大池化层, 卷积核数目翻倍, 直到512为止。偏置值初始化为0.1。整个网络结构如图2所示。

图2中, C表示卷积层, S表示最大池化层, FC表示全连接层。

息。而使用较大尺寸的卷积核可以得到更多的结构信息。对比使用不同尺寸大小的卷积核, 发现较大的卷积核可以提高识别效果。

### 2.3 减少过拟合

为了减少过拟合的影响, 对数据集进行人工扩展, 以期弥补由于标签数据缺少带来的过拟合问题。数据集扩展可通过对草图进行水平翻转和旋转来实现。先将训练集中的原始草图  $1111 \times 1111$  像素变换为  $256 \times 256$  像素, 然后对原始草图和水平翻转草图, 分别抽取5个  $225 \times 225$  切片(中间和四

角),将训练草图集扩展为原来的10倍;最后对原始的草图进行旋转(从 $-5^\circ$ 到 $+5^\circ$ ,每次增加 $1^\circ$ ),将草图集继续扩展10倍,因此训练集共扩展了20倍。

### 2.4 重叠采样层

标准的卷积神经网络的Max-Pooling层是不会从上一层重复采样的。若设步长为 $s$ ,池化区域大小为 $z$ ,则 $z = s$ ,标准的卷积神经网络使用 $z = 2, s = 2$ 。在AlexNet<sup>[12]</sup>中采用重叠采样,使用 $z = 3, s = 2$ 。实验显示,相对于标准的卷积神经网络,AlexNet可以减少过拟合。Deep-Sketch采用了这个设置,发现在减小层数的同时,带来一些空间不变性。

### 2.5 加深网络深度

实验结果显示网络越深得到的结果越好<sup>[18-19]</sup>。Deep-Sketch为了不改变2个Max-Pooling层间特征图的大小,对图片添加大小为1的边距,并使用大小为3的感受野(这是能够获取上/下、左/右和中间的最小的尺寸),步长为1。

### 2.6 取消LRN

LRN主要用于提供明亮度矫正,而手绘草图中没有明亮度信息,因此添加它不仅不会带来效果的提升反而会增加计算时间和内存花费。所以Deep-Sketch取消了LRN。

### 2.7 初始化权值

首先通过训练浅层模型作为深层模型对应层的

初始权值。即先设计了一个浅层模型A(表1、2的第1项),使用3层卷积层,每层加上一个矫正层和Max-Pooling层。第1个全连接层(FC)卷积核大小为 $7 \times 7$ ,与前一pooling层的大小相同。在前2个全连接层后加上Dropout层<sup>[16]</sup>。第3层为250通道,依次对应相应的类别。最后1层采用softmax loss函数作为损失函数。卷积层卷积核数目从64开始,每经过一次Max-Pooling层,值翻倍,直到512为止。根据高斯分布初始化权值,其中, $\mu = 0, \sigma = 0.01$ 。

通过实验对比模型A和添加LRN的模型A(即A-LRN),把效果更好的模型的参数作为构造Deep-Sketch的相应层的参数,如第1~3、5、6、10~13层。Deep-Sketch中的第4、7~9层根据高斯分布随机初始化卷积权值,其中, $\mu = 0, \sigma = 0.01$ 。3种网络结构具体设置如表1(0~5层)和表2(6~14层)所示。其中,项目栏中的数字表示为第几层;表身中卷积层的参数按照“conv<卷积核大小>-<步长>-<拓展边界>-<数量>”来定义,例如Deep-Sketch中第1个卷积层参数“conv15-3-0-64”,表示卷积核尺寸大小为 $15 \times 15$ ,步长为3,填充边界为0,卷积核的个数为64。每个卷积层后跟1层ReLU。类似地,最大池化层(Max-Pooling层)的参数按照“MP<池化层大小>-<步长>-<拓展边界>-<数量>”来定义。

表1 3种模型网络结构设置(0~5层)

Tab.1 Architecture settings of three modes(Layer 0~5)

模型	0 输入层	1 卷积层	2Max-Pooling	3 卷积层	4 卷积层	5Max-Pooling
A	225 × 225	conv15-3-0-64	MP3-2-0-64	conv5-1-0-128	—	MP3-2-0-128
A-LRN	225 × 225	conv15-3-0-64	MP3-2-0-64	conv5-1-0-128	—	MP3-2-0-128
Deep-Sketch	225 × 225	conv15-3-0-64	MP3-2-0-64	conv5-1-0-128	conv3-1-1-128	MP3-2-0-128

表2 3种模型网络结构设置(6~14层)

Tab.2 Architecture settings of three modes(Layer 6~14)

模型	6 卷积层	7 卷积层	8 卷积层	9 卷积层	10Max-Pooling	11FC	12FC	13FC	14 分类器
A	conv3-1-1-256	—	—	—	MP3-2-0-256	512	512	250	softmax
A-LRN	conv3-1-1-256	—	—	—	MP3-2-0-256	512	512	250	softmax
Deep-Sketch	conv3-1-1-256	conv3-1-1-256	conv3-1-1-256	conv3-1-1-256	MP3-2-0-256	512	512	250	softmax

### 2.8 使用更大的Dropout

Dropout<sup>[12,20]</sup>是指在训练模型时随机让网络某些隐含层节点的权值暂时不工作,不工作的那些节点可以暂时认为不是网络结构的一部分,但是它的权值需要保留下来,只是不对它进行更新,而下次样本输入时通过随机选择,它可能又需要工作。只是

不对它进行更新,而下次样本输入时通过随机选择,它可能又需要工作。所以对每一次输入,相当于神经网络会随机选一个不同的网络结构,但这些不同的结构共享权值。因此,Dropout被认为是一种模型组合的替代方法。

这一机制减少了神经元之间的相互适应,一个

神经元不能依赖另一个神经元。由于选取了神经网络的一个随机子集,这样模型会强制学习更好的鲁棒的特征。由于本文数据集小于 ImageNet,所以采用更大的 Dropout 进行补偿,Dropout 取值 0.55。

### 3 实验结果与分析

使用 Eits 等<sup>[1]</sup>组织和收集的目前最大的手绘草图集,该数据集含有 250 类手绘草图,其中每类包含 80 幅灰度手绘草图。该手绘草图集的部分草图如图 3 所示。原始图像大小为  $1\ 111 \times 1\ 111$ 。使用 4 折交叉验证,其中 3 份用于训练,1 份用于测试。分别设计了 3 组对比实验。第 1 组实验在原始数据集上分别对第 1 层卷积核设置不同尺寸,大小为  $7 \times 7$ 、 $11 \times 11$  和  $15 \times 15$ ,进行手绘草图识别实验;第 2 组实验分别在原始数据集和扩展后的数据集上利用 Deep-Sketch 进行手绘草图识别实验;第 3 组实验将 Deep-Sketch 和其它主流手绘草图识别方法进行对比。实验评价指标采用手绘草图识别正确率。

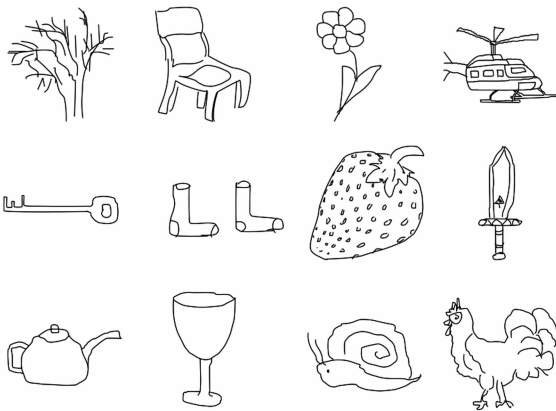


图 3 手绘草图库示例

Fig.3 Examples in sketch base

第 1 组实验使用原始数据集。为了减少层数对结果的影响,层数设置与 Deep-Sketch 一致,除第一卷积层外,其余层与 Deep-Sketch 设置相同。卷积核步长均为 3,数量为 64。表 3 显示不同大小的第 1 层卷积核对手绘草图识别正确率的影响,更大的卷积核获得了更高的识别准确率。

表 3 第 1 层卷积核尺寸对识别正确率的影响

Tab.3 Influence of different sizes of first convolution kernels

第 1 层卷积核大小	正确率/%
$7 \times 7$ <sup>[14]</sup>	62.5
$11 \times 11$ <sup>[12]</sup>	64.5
$15 \times 15$	65.8

第 2 组实验分别在原始数据集和利用第 2.3 节

的方法进行扩展后的数据集上利用 Deep-Sketch 进行手绘草图识别实验。

表 4 显示训练数据集的扩展能够减少过拟合,有效提高模型训练的有效性和识别的正确率。

表 4 数据集扩展对识别正确率的影响

Tab.4 Influence of data augment

是否使用数据扩充	正确率/%
使用	69.2
不使用	65.8

第 3 组实验将提出的 Deep-Sketch(包括未用浅层模型训练初始化参数的 Deep-Sketch 和使用浅层模型训练初始化参数的 Deep-Sketch)分别和 HOG-SVM、SIFT-Fisher Vector、Alex-Net、A 以及 A-LRN 进行对比。

实验结果如表 5 所示,模型 A-LRN(加 LRN)的结果(63.4%)要略低于没有使用 LRN 的 A 模型(63.8%)。无初始化的模型(65.8%)要低于有初始化的情况(69.2%),表明利用浅层模型训练得到的参数初始化 Deep-Sketch 参数,可以获得更好的识别率。模型 A(63.8%)的正确率要远低于 Deep-Sketch(69.2%),表明更深更复杂的模型能够更好地处理手绘草图的类内多样性和类间相似性问题。与传统的图像处理方法 HOG-SVM 和 SIFT-Fisher Vector 相比,提出的方法正确率提升比较明显,分别为 13.2% 和 7.7%。与经典深度卷积网络模型 Alex-Net<sup>[12]</sup>相比,Deep-Sketch 的正确率提高了 3%,但是 Deep-Sketch 使用 925 万训练参数,远小于 Alex-Net 的 6 000 万参数(相当于本文方法的 6.5 倍)。

表 5 与现有方法的对比结果

Tab.5 Comparison with other state-of-the-art recognition methods on sketch

方法	正确率/%
HOG-SVM <sup>[1]</sup>	56
SIFT-Fisher Vector <sup>[2]</sup>	61.5
Alex-Net <sup>[12]</sup>	66.2
A	63.8
A-LRN	63.4
Deep-Sketch(无初始化)	65.8
Deep-Sketch	69.2

### 4 结论

提出一个基于深度学习的手绘草图识别方法,

讨论了数据集扩展、第一层卷积核大小和优化初始权值等不同因素对识别效果的影响。实验结果表明,与现有的主流方法相比,提出的 Deep-Sketch 获得了较高的识别正确率。下一步将研究基于分布式计算的深度学习的手绘草图识别,以减少运行时间。

### 参考文献:

- [1] Eitz M, Hays J, Alexa M. How do humans sketch objects? [J]. *ACM Transactions on Graph*, 2012, 31(4): 44–54.
- [2] Schneider R G, Tuytelaars T. Sketchclassification and classification-driven analysis using fisher vectors [J]. *ACM Transactions on Graphics*, 2014, 33(6): 174–183.
- [3] Sun Zhenbang, Wang Changhu, Zhang Liqing, et al. Free hand-drawn sketch segmentation [C]//*Proceedings of the 12th European Conference on Computer Vision*. Florence: Springer, 2012: 626–639.
- [4] Hu Rui, Collomosse J. A performance evaluation of gradient field hog descriptor for sketch based image retrieval [J]. *Computer Vision and Image Understanding*, 2013, 117(7): 790–806.
- [5] Yasseen Z, Verroust-Blondet A, Nasri A. Sketch-based 3D object retrieval using two views and a visual part alignment [J]. *The Eurographics Association*, 2015, 3(15): 39–46.
- [6] Lowe D G. Distinctive image features from scale-invariant keypoints [J]. *International Journal of Computer Vision*, 2004, 60(2): 91–110.
- [7] Li Yi, Hospedales T M, Song Yizhe, et al. Free-hand sketch recognition by multi-kernel feature learning [J]. *Computer Vision and Image Understanding*, 2015, 137: 1–11.
- [8] Amanatiadis A, Kaburlasos V, Gasteratos A, et al. Evaluation of shape descriptors for shape-based image retrieval [J]. *IET Image Processing Journal*, 2011, 5(5): 493–499.
- [9] Zhao Peng, Zhu Weiwei, Wu Xianwen, et al. Feature description method for freehand sketch by combining entropy and potential energy [J]. *Journal of South China University of Technology: Natural Science Edition*, 2014, 42(5): 122–127. [赵鹏, 朱伟伟, 吴献文, 等. 融合熵和势能的手绘草图特征描述方法 [J]. *华南理工大学学报: 自然科学版*, 2014, 42(5): 122–127.]
- [10] Abdel-Hamid O, Mohamed A-R, Jiang Hui, et al. Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition [C]//*Proceedings of 2012 IEEE International Conference on Acoustics, Speech and Signal Processing*. Kyoto: IEEE, 2012: 4277–4280.
- [11] Chen Mengzhe, Zhang Qingqing, Pan Jieli, et al. Training-target optimization of deep neural network in speech recognition [J]. *Journal of Sichuan University: Engineering Science Edition*, 2016, 48(1): 166–172. [陈梦喆, 张晴晴, 潘接林, 等. 语音识别中神经网络目标值优化 [J]. *四川大学学报: 工程科学版*, 2016, 48(1): 166–172.]
- [12] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks [C]//*Advances in Neural Information Processing Systems*. Nevada: MIT Press, 2012: 1097–1105.
- [13] Simonyan K, Zisserman A. Two-stream convolutional networks for action recognition in videos [C]//*Advances in Neural Information Processing Systems*. Montreal: MIT Press, 2014: 568–576.
- [14] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks [C]//*Proceedings of the 13th European Conference on Computer Vision*. Zurich: Springer, 2014: 818–833.
- [15] Howard A G. Some improvements on deep convolutional neural network based image classification [EB/OL]. (2013–12–19) [2015–05–09]. <http://arxiv.org/abs/1312.5402>.
- [16] Ciresan D C, Meier U, Masci J, et al. Flexible, high performance convolutional neural networks for image classification [C]//*Proceedings of the 22th International Joint Conference on Artificial Intelligence*. Barcelona: MorganKaufmann, 2011, 22(1): 1237–1243.
- [17] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [EB/OL]. (2015–04–10) [2015–05–04]. <http://arxiv.org/abs/1409.1556>.
- [18] Szegedy C, Liu Wei, Jia Yangqing, et al. Going deeper with convolutions [EB/OL]. (2014–09–17) [2015–05–02]. <http://arxiv.org/abs/1409.4842>.
- [19] Lin Min, Chen Qiang, Yan Shuicheng. Network in network [EB/OL]. (2014–03–04) [2015–05–04]. <http://arxiv.org/abs/1312.4400>.
- [20] Hinton G E, Srivastava N, Krizhevsky A, et al. Improving neural networks by preventing co-adaptation of feature detectors [EB/OL]. (2012–07–03) [2015–05–03]. <http://arxiv.org/abs/1207.0580>.
- [21] Cui Zhen, Chang Hong, Shan Shiguang, et al. Deep network cascade for image super-resolution [C]//*Proceedings of the 13th European Conference on Computer Vision*. Zurich: Springer, 2014: 49–64.
- [22] Lecun Y, Boser B, Denker J S, et al. Backpropagation applied to handwritten zip code recognition [J]. *Neural Computation*, 1989, 1(4): 541–551.