

基于贝叶斯网络和卷积神经网络的手绘草图识别方法^①

李鸿雁¹, 苏庭波²

1. 商丘学院 应用科技学院, 河南 商丘 476000; 2. 商丘学院 计算机工程学院, 河南 商丘 476000

摘要: 为实现笔画的分组和识别, 现有的草图识别算法通常会采用限制用户的绘图习惯来达到目的. 该文提出了利用贝叶斯网络和卷积神经网络(CNN)的草图识别方法解决此问题. 首先, 使用高斯低通滤波器处理输入草图, 得到更平滑的图像. 然后将连续输入的笔划分为两部分, 分别使用贝叶斯网络和卷积神经网络对分割后的笔画进行识别, 当笔画的可靠性大于阈值时, 以贝叶斯网络的识别结果为准, 反之采用 CNN 的识别结果. 实验结果表明, 该文算法在电路符号绘制过程中的识别率和绘制完成后的识别率均取得了较好的结果. 该文算法具有良好的应用前景.

关键词: 贝叶斯网络; 卷积神经网络; 笔画分组; 草图识别

中图分类号: TP393

文献标志码: A

文章编号: 1000-5471(2019)09-0096-07

一般情况下, 笔画识别算法要求用户按照固定的规范实现绘图, 以便提高该算法识别笔画的精确度. 草图识别包括 2 个子问题: 笔画分组和符号识别.

训练手势识别器在文献[1-3]中提出了限制, 要求用户必须单笔完成语义符号的绘制. 文献[4]假设场景仅包含一个对象, 使用单个笔画绘制出语义符号. 在文献[5]中, 明确要求用户在画完一个语义符号后必须按下按钮. 文献[6]允许具有由多个笔划组成的多个对象和对象的场景, 但是假设没有对象共享笔画. 文献[7]采用时间阈值法, 要求用户在绘制语义符号后应有明显的停顿, 文献[8]采用特殊的图形符号(如箭头)进行分组. 这些算法有很多限制, 使手绘界面不但失去了自然绘制的特点, 而且用户在手绘方面也不能更好地进行快速表达和自由绘制.

通常, 利用结构匹配方法和统计识别算法实现手绘草图的识别^[9-10]. 结构匹配方法通过提取手绘草图的基本图元及其图元之间的空间位置关系, 对语义符号加以识别^[11]. 这种方法方便对相似的语义符号加以区分, 但其抗干扰能力较弱, 容易受噪声影响. 而统计识别算法将草图中所有语义符号作为统一整体, 并计算多维特征向量. 这种算法抗干扰能力强, 但相应的分支能力很弱. 近期, 人们在绘制草图的过程中针对素描识别使用了增量贝叶斯方法^[12], 并通过推理得出用户在绘制草图的过程中采用此方法会更加自由.

在机器学习领域中, 深度学习是一个崭新的研究方向. 在一个包含多个隐藏层的非线性网络模型中, 通过对大量原始数据的训练进行预测或对样本进行分类. 在图像识别和计算机视觉领域, CNN 取得了最显著的成果^[13]. 此外, 深度学习被广泛应用于自然语言处理、行人检测、手势识别、数据挖掘和语音识

① 收稿日期: 2018-06-01

基金项目: 河南省科技厅项目(182102210511).

作者简介: 李鸿雁(1984-), 女, 讲师, 主要从事人工智能及图像处理研究.

别^[14-15]. 与其他深层神经网络相比, CNN 可以直接处理二维图像, S 层自动编码^[16]. 当二维图像被转换为一个时, 输入数据的空间结构特征将丢失.

为了让用户在绘图过程中更自由, 并进一步提高手绘示意图的识别率, 本文提出了一种基于贝叶斯网络和 CNN 的识别算法.

1 算法模型

1.1 预处理和笔画分割

针对原有笔画采样过程的不稳定性, 先对每个笔画的采样值进行预处理. 使用高斯低通滤波器对手绘图进行滤波, 从而获得平滑的手绘图.

手绘笔画可以分为基本的图像元素, 比如直线和圆弧. 分割图像元素可以分为采集笔画数据和检测特征点 2 个步骤. 任何笔画特征均由 3 组特征值的方向、曲率和速度组成. 例如任一笔画细分为 N 个点, 第 N 个点的坐标用 (x_n, y_n) 表示, 则采样点运动方向向量的角度偏移 d_n 为

$$d_n = (\text{Arctan2}((y_{n+1} - y_{n-1}), (x_{n+1} - x_{n-1})) - \text{Arctan2}((y_{n+1} - y_{n-1}), (x_{n+1} - x_{n-1}))) \cdot 180/\pi \quad (1)$$

式(1)中, 圆弧正切函数 $\text{Arctan2}(x, y)$ 可以处理特殊情况 $x = 0$. 曲率 c_n 是采样点单位弧度的方向偏差, 用它描述采样点曲率信息, 其表达式为

$$c_n = \frac{|\sum_{i=n-k}^{n+k-1} \varphi(d_{i+1} - d_i)|}{D(n-k, n+k)} \quad (2)$$

式(2)中, $D(n-k, n+k)$ 是 $n-k$ 点到 $n+k$ 点之间的弦长之和. 采样点运动速度信息一般采用速度 $Speed_i$ 进行描述

$$Speed_i = \frac{dl(\text{point}_i, \text{point}_{i-1})}{(T_i - T_{i-1})} \quad (3)$$

其中, $dl(\text{point}_i, \text{point}_{i-1})$ 是点 point_i 到 point_{i-1} 的折线距离, T_i 为 i 点的绘图时间.

分析手绘草图的特点, 特征点(笔画的显著拐点)通常具备波峰和波谷 2 个特点, 分别是曲率变化迅速和速度变化迅速. 通过扫描草图的曲率和速度曲线, 可以检测出草图的特征点, 并将监测结果作为候选关键点. 关于像“S”这样的图形符号, 当它们在拐点处速度变化不明显时, 可以利用滑动窗口法计算这些拐点曲率的正负值. 通常正负值变化的点也是特征点, 根据扫描草图和计算曲率监测的特征点, 划分出线段和圆弧等基本几何图元.

1.2 贝叶斯网络模型

假设草图在整个识别过程中的所有语义符号用 C_1, C_2, \dots, C_i 表示. 图形语义符号可以看作是由圆弧、线段和其他基本元素, 即 $C_i = (x_1, x_2, \dots, x_n, r_1, r_2, \dots, r_m)$ 组成的几何图形, 依据一定的空间约束关系. 其中 $x_j \in (\text{line}, \text{arc})$, $r_j \in (\text{parallel}, \text{interest}, \text{adjacent}, \text{vertical}, \text{lengthEqual}, \text{closed})$ 是空间约束关系的集合. 本文采用的贝叶斯网络结构如图 1 所示. 在图 1 中, S_1, S_2, \dots, S_k 是最初的笔画, 为贝叶斯网络提供了依据, 但它们不是贝叶斯网络的一部分. 由于同一语义符号存在多种不同的绘制方式, 并没有完整的集合加以描述, 因此本文的贝叶斯网络节点不直接与原始笔画对应, 而具体将其分为语义符号类 C 、原始 x 和空间约束关系 r . 为了降低计算量, 假定这些特性与基本元素 x_j 和 r_j 之间没有直接联系, 抽象在图 1 中就是它们之间没有边直接相连, 彼此之间是相互独立的.

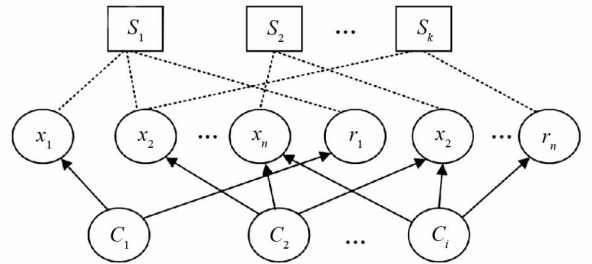


图 1 贝叶斯网络结构

网络节点的 C_i 值为 $(\text{true}, \text{false})$. x_j 和 r_j 的值分别是语义符号中包含的基本元素和空间关系的数量.

确定的笔画集合通过贝叶斯识别器获得后,被划分为基本的图像元素.利用贝叶斯网络计算 8 维特征向量后验概率公式的最大值,将 8 维特征向量划分到最紧密的一类 C_i 中.其中,后验概率公式为

$$P(C_i | x_1, r_1, x_2, r_2, \dots, x_n, r_n) = \frac{P(x_1, r_1, x_2, r_2, \dots, x_n, r_n | C_i)P(C_i)}{P(x_1, r_1, x_2, r_2, \dots, x_n, r_n)} \quad (4)$$

在 $P(x_1, r_1, x_2, r_2, \dots, x_n, r_n)$ 不变的条件下,后验概率公式的最大值为

$$C = \operatorname{argmax} P(x_1, r_1, x_2, r_2, \dots, x_n, r_n | C_i)P(C_i) \quad (5)$$

根据朴素贝叶斯的假设,联合概率分布值的获得是每个属性概率分布的乘积,因为每个特征向量都是独立的,所以在候选符号 C_1, C_2, \dots, C_i 中最匹配的 C' 类定义为

$$C' = \operatorname{argmax} P(C_i) \prod_n P(x_n | C_i) \prod_m P(r_m | C_i) \quad (6)$$

由于草图的模糊性,因此针对候选语义符号,利用贝叶斯分类器计算其信度值.通过研究,本文将信度阈值设为 0.85.如果其高于手绘符号的信度值,则根据信度值的降序来提取 k 个语义符号.反之,则会将高于这个阈值的信度值赋给拥有最高后验概率的笔画集.

1.3 卷积神经网络

由于局部像素的图像一般比较相似,因此 CNN 对图像的理解是由局部到全局的.首先, CNN 感知图像的局部特征,然后在更高的层次上结合这些局部特征.通过这种方法,可以获得图像的全局特征及其拓扑结构,进而确定图像的属性 and 类别.因此,卷积神经网络对缩放、平移、倾斜或其他变形形式具有高度的不变性.

CNN 有 2 个典型的特点:① 2 个神经元层之间局部连接,而不是完全连接.因此,连接到输入图像的卷积层是为像素块构建的本地链接,而不是传统的基于像素点的全连接.② 卷积核的权重参数在同一层中共享.这 2 个特性大大降低了深网参数的数量,降低模型的复杂度,加快训练速度.它使得 CNN 在处理单元的像素值上有很大优势. CNN 的主要组件包括卷积层、池化层、激活函数、全连接层和分类器,如图 2 所示.

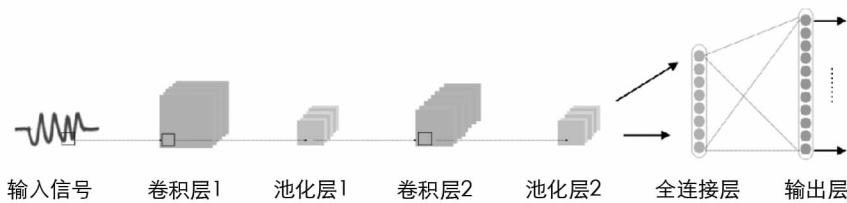


图 2 卷积神经网络结构

1.3.1 卷积层

卷积神经网络的核心组成部分是卷积层,其主要功能是提取输入的局部特征,并通过卷积核的固定步长移动.特征图中每个元素的输出是一个神经元的输出,神经元连接的输入是前一个输出特征图的局部区域.感觉场的输入是通过一组突触权重来计算的,神经元的输出通过激活函数获得.在生成特征图的过程中,通过共享这组突触权重,可以大大减少参数的数量.卷积运算的数学表达式为

$$a_n^l = f\left(\sum_{\forall m} (a_m^{l-1} * k_{m,n}^l) + b_n^l\right) \quad (7)$$

其中, a_n^l 和 a_m^{l-1} 是当前层和之前层的特征图. $k_{m,n}^l$ 是从上一层的第 m 个特征图到当前层的第 n 个特征图的卷积核. $f(x) = 1/[1 + \exp(-x)]$ 是神经元的激活功能. b_n^l 是神经元的偏移,表示不同的卷积核可以提取不同的特征,同样也是与上一层特征映射卷积核的响应.

1.3.2 池化层

通过在输出特征图上非重叠矩形区域的函数变换,池化层可以获得较高层次的不变特征,它的功能是

改变卷积层的特征图并降低特征图的维度, 同时降低倾斜和位移的输出灵敏度, 改进模型的泛化能力. 在正向计算过程中, 通过池化操作, 逐步扩展输入图像和特征图, 并将其集成到全局特征中.

常用的池化方法有平均池和最大池, 池层可以在保存原始特征信息的同时减少功能维度. 池层节点的输出可以表示为

$$a_n^l = f(k_n^l \times \frac{1}{s^2} \sum_{s \times s} a_n^{l-1} + b_n^l) \quad (8)$$

式中, $s \times s$ 是较低的采样模板维度, k_n^l 为模板权值.

1.3.3 全连接层

由于卷积层通常使用多个卷积核模板, 因此输出也是同样大小的特征图. 需要一个或多个全连接层, 以便将特征映射用于分类. 整个连接层通常连接到池层和分类器的最后一层, 用于融合多个特征图表示的不同特征. 卷积神经网络的强大特征提取能力来源于多卷积核模板的卷积运算, 每个卷积核模板的输出代表一个不同角度的特征表达式, 因此将这些特性合并在一起是一个很重要的过程. 全连接层中的每个神经元都连接到上一层输出特性的所有神经元, 最终的输出结果 a_n^{out} 可以表示为

$$a_n^{out} = f(\sum_{\forall m} (a_m^{out-1} * k_{m,n}^{out}) + b_n^{out}) \quad (9)$$

式中, a_m^{out-1} 为上一层的输出, $k_{m,n}^{out}$ 为卷积核的权值, b_n^{out} 为偏移量. 完整的连接层结合了前面特征图的所有特性, 然后将其输入到 softmax 分类器中. 常用的激活函数为 Sigmoid, Tanh 和 Relu(Rectified unit).

如图 2 所示, 本文设计了一个 6 层网络. 每一个池化层均紧随在卷积层之后.

1.4 样本训练

在利用贝叶斯网络进行推理的过程中, 根据所有符号逐一收集的笔画样本, 将其进行样本训练得到 $P(x | H)$. 其中, x 表示基本几何图元的数量, 包括直线段、圆弧、平行线、由 8 维特征向量组成的闭合线等; H 表示语义符号的范围. 在确定训练样本特征的基础上, 得到 $P(x_k = n | C_i)$ 和 $P(r_k = m | C_i)$, 利用的计算公式为

$$P(x_k = n | C_i) = \frac{C(C_i, x_k, n)}{C(C_i)} \quad (10)$$

$$P(r_k = m | C_i) = \frac{C(C_i, r_k, m)}{C(C_i)} \quad (11)$$

其中, $C(C_i, x_k, n)$ 为样本中语义符号 C_i 基元 x_k 数目为 n 的数量; $C(C_i, r_k, m)$ 为样本中语义符号 C_i 空间约束关系 r_k 数目为 m 的数量; $C(C_i)$ 为样本中出现 C_i 的总次数. 利用训练样本的特征向量和概率构造出高斯分布函数, 并利用其计算样本分类时的条件概率.

$$P(X | C_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp - \frac{(x - \mu)^2}{2\sigma^2} \quad (12)$$

训练样本由选择样本中的单位符号(CNN)进行训练, 而卷积神经网络中使用的代价函数 $J(\omega, b)$ 和 $J(\omega, b)$ 可以通过以下公式得到.

$$J(\omega, b) = \frac{1}{m} \sum_{i=1}^m J(\omega, b; x_i, y_i) = \frac{1}{m} \sum_{i=1}^m \left(\frac{1}{2} \| h_{\omega, b}(x_i) - y_i \|^2 \right) \quad (13)$$

其中, 单元与单元之间的连接参数 ω, b 为偏移量, m 为训练样本数量. $h_{\omega, b}(x_i)$ 为卷积神经网络最后一层的输出值, y_i 是第 i 个训练样本所对应的标签. 通过不断更新参数 ω 和 b , 求出代价函数 $J(\omega, b)$ 的最小值, 从而达到网络训练的目标. 其中, 在 1 层参数 ω 和 b 的计算方法为

$$\omega_{ij}^l = \omega_{ij}^l - \alpha \frac{\partial}{\partial \omega_{ij}^l} J(\omega, b) \quad (14)$$

$$b_i^l = b_i^l - \alpha \frac{\partial}{\partial b_i^l} J(\omega, b) \quad (15)$$

其中, α 是学习速率. 在逆向传导的应用算法中, 提出前向计算, 可以计算出卷积神经网络最后一层的输出值 $h\omega, b(x_i)$.

1.5 本文算法

该算法的流程如图 3 所示. ①利用高低通滤波器预处理每个笔画的采样值, 获得相应的数据. ②通过检测笔画的特征点, 把所有笔画分为若干基本图元, 并计算图元之间的空间约束关系. 根据给定的笔画样本, 利用训练模块计算出神经网络的条件概率和训练参数. ③利用识别模块根据笔画采集的特征进行匹配. ④根据可信度确定相应的语义符号, 计算相应的后验概率. 如果可信度值小于某一阈值, 则使用卷积神经网络识别的结果来获得相应的语义符号. 也就是说, 当笔画的可信度值大于阈值时, 将使用贝叶斯识别结果, 反之则使用 CNN 识别结果.

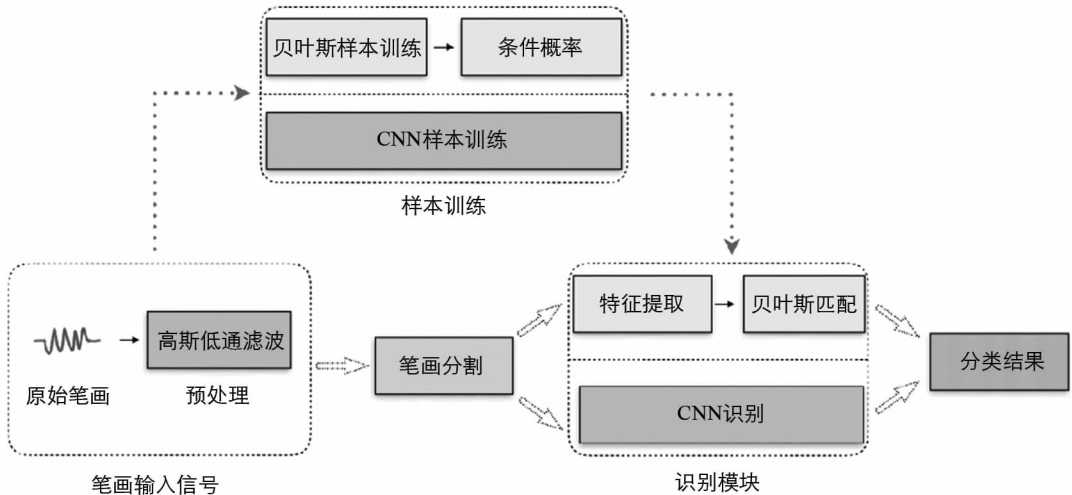


图 3 算法流程图

2 实验结果与分析

选取电路图符号的示意图作为实验数据. 有 15 人参与了这个实验, 每个人被要求画 10 幅草图. 每个草图包含图 4 中的所有图形符号. 然后得到 150 个草图和 750 个符号, 并取出 20 个符号作为训练样本. 在贝叶斯网络中, 可信度阈值设置为 0.85.

根据图 5 所示的手工绘制电路图, 在绘图过程中识别率如表 1 所示. 当电路符号绘制 75% 时, 识别率的平均值为 81.27%, 说明通过用户输入残缺的草图, 可以推算用户绘制草图的用意. 基于贝叶斯网络分类器可以处理不完整的数据. 用户在绘制出全部完整的符号后, 针对电路图中常用的 5 个电路符号的平均识别率达到 92.16%, 其中识别率较高的为复杂符号, 最高为 95.8%, 识别率较低的如形状和拓扑信息的符号, 最低为 88.9%.

表 1 绘制过程中的识别率

电路符号	绘制 75% 时的识别率	绘制完整后的识别率
电容	75.5	89.9
二极管	80.8	92.4
三极管	88.2	95.8
电阻	82.0	94.3
电源	79.85	88.4
平均值	81.27	92.16

%

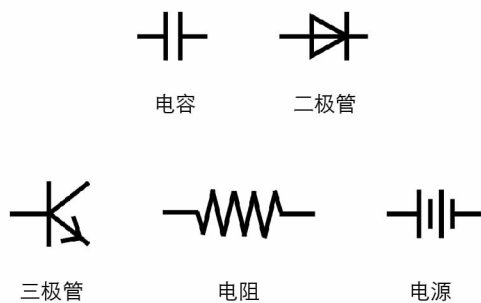


图4 电路图符号

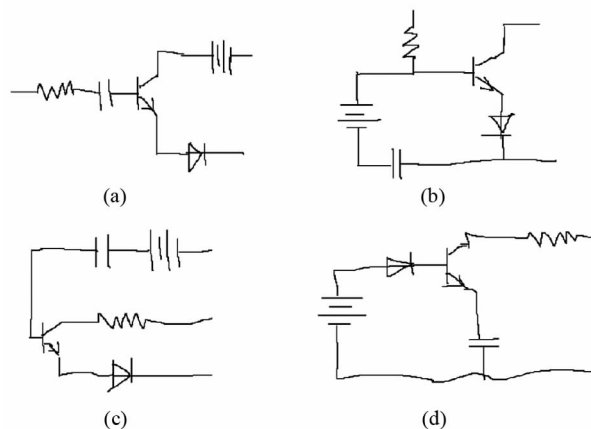


图5 手绘电路图

3 结 论

本文提出的基于贝叶斯网络和 CNN 的手绘草图识别算法, 不仅为了让用户在绘制过程中更加自由, 同时也有助于进一步提高手绘草图的识别率. 基于草图结构贝叶斯网络的最大后验概率, 算法对连续输入笔画进行动态最优分组, 再利用贝叶斯网络和 CNN 对分割后的笔画进行分类. 通过电路符号测试, 表明该算法在线性渐进式针对手绘草图的识别过程中非常有效, 对于当今以笔式交互为代表的人机交互具有重要的现实意义. 本文目前的实验对象比较简单, 若在此基础上进一步提高算法在不同领域中手绘草图的泛化识别能力, 将逐渐转向更复杂图形的识别和基于本体的概念识别.

参考文献:

- [1] 谭台哲, 韩亚伟, 邵 阳. 基于 RGB-D 图像的手势识别方法 [J]. 计算机工程与设计, 2018, 39(2): 511-515.
- [2] FANG G, HE L, KONG F. Research on Technologies of Computer Aided Sketching Design [J]. Computer Engineering, 2006, 32(18): 1-2, 5.
- [3] 吴晓凤, 张江鑫, 徐欣晨. 基于 Faster R-CNN 的手势识别算法 [J]. 计算机辅助设计与图形学学报, 2018, 30(3): 468-476.
- [4] ANDERSON D, BAILEY C, SKUBIC M. Hidden Markov Model Symbol Recognition for Sketch-Based Interfaces [C]// AAAI Fall Symposium Series Making Pen-Based Interaction Intelligent and Natural. Palo Alto: AAAI Press, 2013.
- [5] KURTOGLU T, STAHOVICH T F. Interpreting Schematic Sketches Using Physical Reasoning [C]// Proceedings of AAAI Spring Symposium on Sketch Understanding. Palo Alto: AAAI Press, 2002.
- [6] SEZGIN T M, DAVIS R. HMM-Based Efficient Sketch Recognition [C]. San Diego: Proceedings of the 10th International Conference on Intelligent User Interfaces, 2005.
- [7] FONSECA M, PIMENTEL C, JORGE J. CALI: An Online Scribble Recognizer for Calligraphic Interfaces [C]// Proceedings of AAAI Spring Symposium on Sketch Understanding. Palo Alto: AAAI Press, 2002.
- [8] LESLIE M G, LEVENT B K, THOMAS F S, et al. Combining Geometry and Domain Knowledge to Interpret Hand-drawn Diagrams [J]. Computers and Graphics, 2005, 29(4): 547-562.
- [9] SUN Z X, FENG G, ZHOU R. Techniques for Sketch-Based User Interface: Review and Research [J]. Journal of Computer-Aided Design & Computer Graphics, 2005, 17(9): 1889-1899.
- [10] SUN X Z. Sketch-Based Graphic Input Tool for Conceptual Design [J]. Journal of Computer Aided Design & Computer Graphics, 2003, 15(9): 1145-1152.
- [11] 宋保华, 叶 军, 于明玫, 等. 笔输入草图的分层识别 [J]. 计算机辅助设计与图形学学报, 2004, 16(6): 753-758.
- [12] SHAO Y X, WANG C H, XIAO B H, et al. Self-Generation Voting Based for Handwritten Chinese Character Recogni-

tion [J]. *Acta Automatica Sinica*, 2013, 39(4): 450-454.

- [13] RAZAVIAN A S, AZIZPOUR H, SULLIVAN J, et al. CNN Features off-the-shelf: an Astounding Baseline for Recognition [C]//2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Columbus: IEEE, 2014.
- [14] GOLDBERG Y. Neural Network Methods for Natural Language Processing [J]. *Synthesis Lectures on Human Language Technologies*, 2017, 10(1): 1-309.
- [15] JOHN V, MITA S, LIU Z, et al. Pedestrian Detection in Thermal Images Using Adaptive Fuzzy C-means Clustering and Convolutional Neural Networks [C]//Proceedings of the 14th IAPR International Conference on Machine Vision Applications. Tokyo: IEEE, 2015.
- [16] CAI J, CAI J Y, LIAO X D, et al. Preliminary Study on Hand Gesture Recognition Based on Convolutional Neural Network [J]. *Computer Systems & Applications*, 2015, 24(4): 113-117.

Sketch Recognition with Bayesian Network and Convolution Neural Network

LI Hong-yan¹, SU Ting-bo²

1. *Shangqiu university applied science and technology college, Shangqiu Henan 476000, China;*

2. *School of Computer Engineering, Shangqiu University, Shangqiu Henan 476000, China*

Abstract: Most of the existing sketch recognition algorithms have been used to restrict the user's drawing habits to achieve the stroke grouping and recognition. In this paper, a sketch recognition method based on Bayesian network and convolutional neural network (CNN) is proposed to solve this problem. First, the input sketch is processed by gaussian low-pass filter and a smoother stroke can be obtained. The stroke of continuous input is divided, then the bayesian network and CNN are performed on stroke recognition respectively. The recognition result of bayesian network is adopted when the reliability of stroke is larger than the threshold, otherwise recognition result of CNN will be adopted. The experiment result shows that the proposed algorithm is effective in circuit symbol recognition. The recognition rate was achieved 81.27% in the drawing process, and the final recognition rate was achieved 92.16%.

Key words: Bayesian network; convolution neural network; stroke grouping; sketch recognition

责任编辑 夏娟