

文章编号: 2095-2163(2021)06-0148-05

中图分类号: TP183

文献标志码: A

# 基于神经网络的手绘草图的识别与匹配

杨金凯, 王国中, 范涛

(上海工程技术大学 电子电气工程学院, 上海 201620)

**摘要:** 手绘草图通常是一种由几条简单笔划就能完成的对人们想法或者概念有着明确表达的一种视觉上表现形式。对于人类而言, 识别草图的含义非常容易, 但是对于机器来说是非常困难的。为了实现手绘草图的精准识别与匹配, 本文提出一种特征融合的新思路, 利用神经网络的特点将各级网络特征进行融合来实现草图的精准识别与匹配。将神经网络特征逐层下采样进行融合, 即将浅层特征与深度的语义特征相融合, 而传统网络仅仅利用了深度特征而忽略了浅层特征。实验结果表明该方法的准确率达到 98.3%, 优于传统网络。

**关键词:** 识别草图; 神经网络; 特征融合

## Recognition and matching of hand-drawn sketch based on neural network

YANG Jinkai, WANG Guozhong, FAN Tao

(School of Electronic and Electrical Engineering, Shanghai University of Engineering technology, Shanghai 201620, China)

**[Abstract]** Freehand sketching is usually a visual representation of people's ideas or concepts that can be completed by a few simple strokes. It is very easy for humans to recognize the meaning of sketches, but it is very difficult for machines. In order to realize the accurate recognition and matching of freehand sketches, a new idea of feature fusion is proposed, which USES the features of neural network to fuse the features of all levels of networks to realize the accurate recognition and matching of sketches. First of all, we fuse the features of the neural network layer by layer and merge the shallow features with the semantic features of the depth. However, the traditional network only USES the depth features but ignores the shallow features. Experimental results show that the accuracy of this method is 98.3%, which is better than traditional network.

**[Key words]** recognition sketch; neural network; feature fusion network

## 0 引言

在一些特殊场景下, 如军事行动、工程项目的现场临时指挥和调度, 人们往往需要当场手绘一张草图, 来传达自己的意图。这就要求计算机能够根据绘图者所绘制的草图来匹配标准的符号图, 提高绘图的质量和效率。传统的方法都是采用特征空间的相似度进行匹配, 来实现草图的匹配。如矢量特征描述法即对线条的描述与刻画、图模型匹配法等, 但是实际效果不理想难以实际应用。本文的方法包括 2 个部分, 第一部分是让卷积神经网络识别草图是什么; 第二部分则是根据识别结果匹配标准图, 从而实现草图的匹配。

## 1 卷积神经网络模型

本文提出的基于神经网络的手绘草图的识别与

匹配算法在精准度上可以达到 98% 以上, 比传统的特征匹配方法有非常大的提升, 并且可以达到实际应用的程度, 对于技术的落地应用具有重大意义。

VGGNet 是牛津大学计算机视觉组 (Visual Geometry Group) 和谷歌一起研发的深度卷积神经网络。VGGNet 研究了卷积神经网络的性能和卷积神经网络深度之间的关系, 通过重复使用  $3 \times 3$  卷积核进行卷积以及  $2 \times 2$  的最大池化, 组合成深度为 16~19 层的卷积神经网络, 使得图片分类任务的准确率相较于之前的图片分类方法有很大的提高。不仅如此 VGGNet 的强大特征提取能力使得其迁移能力也很强, 在迁移到其它图像任务上的泛化能力也很强。

VGGNet 特点是所有的卷积层使用的卷积核和池化层都是  $3 \times 3$  的卷积核与  $2 \times 2$  的最大池化, 其性能的提升主要是依赖网络结构的不断加深。VGGNet 的参数数量很大, 但绝大部分的参数都消耗

**基金项目:** 国家重点研发计划(2019YFB1802700); 国家专项发展资金重大项目(ZJ2020-ZD-009)。

**作者简介:** 杨金凯(1993-), 男, 硕士研究生, 主要研究方向: 图像处理、计算机视觉; 王国中(1962-), 男, 博士, 教授, 博士生导师, 主要研究方向: 视频编解码、图像处理、机器学习等; 范涛(1983-), 男, 博士, 讲师, 主要研究方向: 智能视频处理、计算机视觉、人工智能应用。

**通讯作者:** 王国中 Email: wanggz@sues.edu.cn

收稿日期: 2020-12-03

在了最后面的 3 个全连接层,卷积层虽然很深但是消耗的参数量却远远少于全连接层。从图 1 可以看出网络结构从 VGG-A 到 VGG-E 的深度都逐渐变深,但总的网络的参数量并没有增加太多。图 1 中的网络 D 和 E 分别是 VGGNet-16 和 VGGNet-19。

VGGNet 由 5 个块卷积组成,每一个块卷积由 2-3 个卷积层和一个最大池化层构成,其中最大池化层起着保留图片的主要特征以及图片下采样的作用。每一个块卷积中的卷积层的卷积核数量都是相

同的,每个卷积层的激活函数则用的是 ReLU 函数来进行非线性变换<sup>[1]</sup>。卷积块的卷积核数量则是越靠后越多,这是因为随着特征图的缩小,图像的信息会不断减少,因此要通过增加卷积核数量的方法来弥补由于特征图的缩小而带来的信息丢失。在 5 个卷积块的后面加上 2 个全连接层(Fully Connected Layers, FC)用于提取特征图信息。最后再用一个全连接层连接 Softmax 来分类。

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input(224×224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

图 1 VGGNet 各级别网络结构图

Fig. 1 VGGNet level network structure diagram

## 2 VGGNet 模型的改进

### 2.1 批量归一化

卷积神经网络学习的实质是学习图像数据分布规律,网络的泛化能力与训练数据和测试数据分布有着密切的关系。一般情况下都是假定训练数据和预测数据有着相同的分布,但是由于数据都是带有噪声的,为了使得噪声数据对整体数据的分布影响减少,通常对所有输入数据进行归一化处理后再进行训练。这样做是有效的,但是是不够的,因为在训练的过程中,由于每层网络的参数不同,会导致该层输入与输出数据的分布不同,也就是说每层网络都

会改变数据的分布。另外,由于按批次训练数据,对于每个批次来说,其数据分布也会不同,导致网络在每次训练迭代学习方向都是向着新的批次的分布拟合,这大大增加了训练的难度以及过拟合的风险。从网络整体上看,神经网络试图在多个分布中找到相对平均的分布来作为所有分布的一个替代;从每个卷积层上看,因为每次输入到每层网络的数据的分布都不同,所以导致每层网络都在找一个满足所有分布替代分布,神经网络的收敛困难程度就大大增加了。针对上述问题,批量归一化(Batch Normalization)方法则是一个有效的方法,该方法针对每一批数据,在网络的每一层输入之前增加归一

化处理(均值为0,标准差为1),使得每次输入到各个网络层的数据都强制在统一的数据分布下(正态分布)<sup>[2]</sup>。其计算步骤如下:

**Step 1** 将  $m$  个训练样本作为一个批次 (Batch) 输入  $B = \{x_1, x_2, \dots, x_m\}$ ;

**Step 2** 计算一个批次样本的均值,式(1):

$$\mu_B = \frac{1}{m} \sum_{k=1}^m x_k, \quad (1)$$

**Step 3** 计算一个批次样本的方差,式(2):

$$\delta_B^2 = \frac{1}{m} \sum_{i=1}^m (x_k - \mu_B)^2, \quad (2)$$

**Step 4** 归一化,式(3):

$$\overset{\sim}{x}_k = \frac{x_k - \mu_B}{\delta_B}, \quad (3)$$

**Step 5** 尺度变换,式(4):

$$\overset{\sim\sim}{x}_k = g * \overset{\sim}{x}_k + b. \quad (4)$$

其中,  $g$  是再缩放参数(re-shift parameter),  $b$  是再平移参数(re-scale parameter)。

最后将一个批次的数据转换成均值为  $b$  方差为  $g^2$  的正态分布,  $b, g$  是需要训练的参数。

### 2.2 随机失活(dropout)

如果训练样本太少,或者网络模型太复杂(参数多),训练出来的模型往往会产生过拟合的现象。在训练神经网络时的过拟合的问题是经常出现的,而过拟合具体表现在:在训练数据上,网络模型的损失值很小,预测的准确率较高;在测试数据上,损失值很大,预测的准确率较低。为了解决过拟合问题,需要使得神经网络中的神经元在训练时不全都工作,即在模型的全连接层加入 dropout 层<sup>[3]</sup>。其工作过程如下:

(1) 随机(临时)删掉网络中一定数量的隐藏神经元,输入输出神经元保持不变。如图2(右)所示;

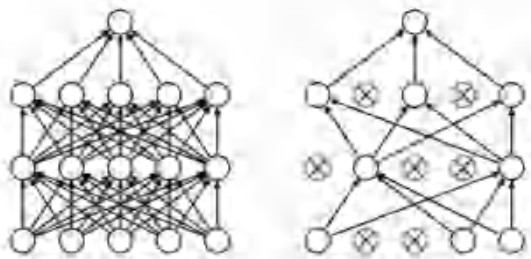


图2 标准的神经网络(左)和应用 Dropout 后的神经网络(右)

Fig. 2 The standard neural network (left) and the post-Dropout neural network (right)

(2) 固定修改后的网络,让一个批次的训练数

据输入到网络训练,对没删除的神经元参数进行梯度下降更新;

(3) 恢复被删除的神经元,重复(1)(2)操作;

(4) 不断重复这一过程,直至网络模型收敛。

Dropout 的过程就是在前向传播时,让某个神经元的激活值以一定的概率  $p$  停止工作,这样可以明显地减少过拟合现象,使模型泛化性更强,因为其不会太依赖某些局部的特征。

### 2.3 特征融合

神经网络提取特征越是靠前的特征其浅层次的细节特征(颜色、纹理等特征)就越丰富,越靠后的特征其深层次的语义信息就越丰富。依据这个特点,可以将深层次的语义特征和浅层次的细节特征融合到一起。这样融合在一起的信息会减少原图的信息丢失,使得图像特征在表达图像时更准确。本文提出采用浅层特征下采样的方式,将浅层特征跟深层特征融合到一起,网络结构如图3所示。

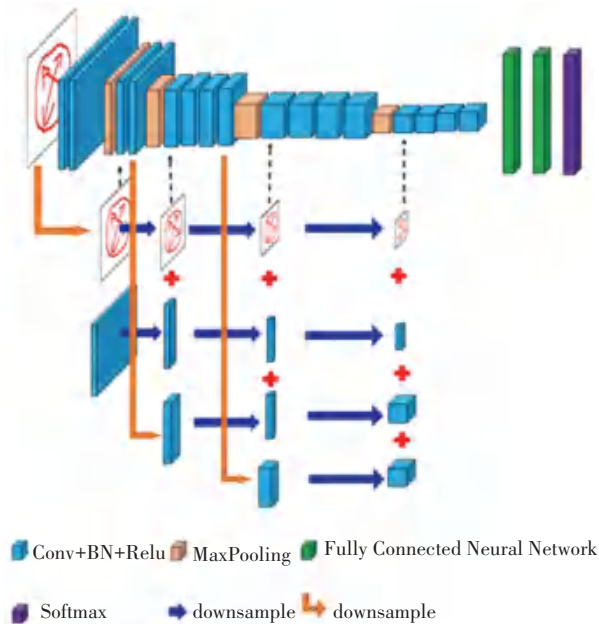


图3 网络结构图

Fig. 3 Network structure diagram

### 3 算法流程

在手绘标识草图送入神经网络的某一层进行处理之前,对手绘标识草图做归一化。按照训练样本的批量进行处理,先减掉这批样本的均值,除以标准差,然后进行缩放和平移,缩放和平移参数由训练得到。预测时使用训练时确定的这些值来计算。手绘标识草图经过神经网络进行分类处理得到图片所属类别,根据所属类别匹配标准的标识图,从而完成手

绘草图的匹配任务,流程图如图 4 所示。

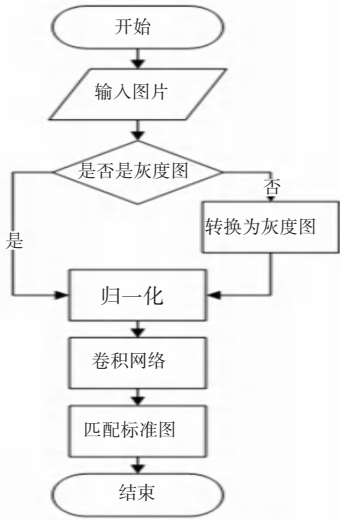


图 4 算法流程图

Fig. 4 Algorithm flow chart

### 4 实验及结果分析

#### 4.1 实验平台

实验是在 PC 机上完成的,配置为 Windows10 系统, CPU 为英特尔 Xeon (至强) W - 2123 @ 3.60 GHz四核, GPU 为 Nvidia GeForce GTX 2080Ti, 内存为 12 G。

#### 4.2 实验数据

实验使用的数据是徒手绘制的 7 054 张符号图,包括 67 种不同的符号。每一种符号手绘图在 250 张左右,大小为 64×64×3。为了保证数据的平衡,在实验中分别取 67 个手绘符号草图数据的 80% 作为训练集,20%作为测试集。部分数据集展示如图 5 所示。

#### 4.3 评价标准

##### 4.3.1 准确率 (Accuracy)

准确率是分类匹配问题中最为原始的评价指标。准确率是预测正确的结果占总样本的百分比,公式(1)如下:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

其中:

真正例 (True Positive, TP):被模型预测为正的正样本;

假正例 (False Positive, FP):被模型预测为正的负样本;

假负例 (False Negative, FN):被模型预测为负

的正样本;

真负例 (True Negative, TN):被模型预测为负的负样本;

$$样本总数 = TP + TN + FP + FN$$

$$正样本数 = TP + FN$$

$$负样本数 = TN + FP$$

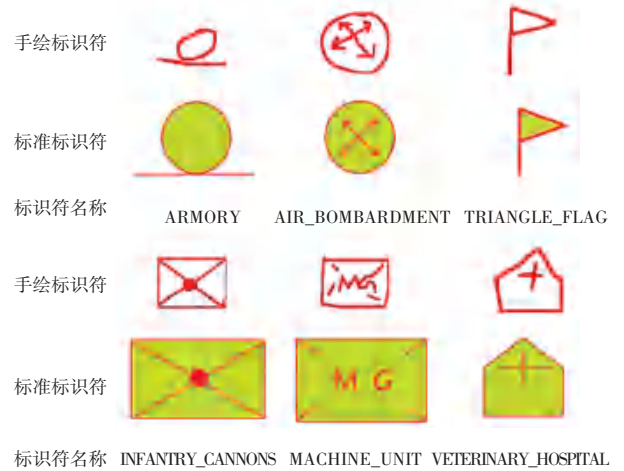


图 5 部分数据集展示图

Fig. 5 The partial data set

##### 4.3.2 精确率 (Precision)

精确率又叫查准率,是针对预测结果而言的,是在所有被预测为正的样本中实际为正的样本的概率,其公式(6)如下:

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

精确率和准确率是完全不同的两个概念,精确率代表对正样本结果中的预测准确程度,而准确率则代表整体的预测准确程度,既包括正样本,也包括负样本。

##### 4.3.3 召回率 (Recall)

召回率 (Recall) 又叫查全率,是在实际为正的样本中被预测为正样本的概率,公式(7)如下:

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

本次实验采用准确率来做为模型的评判标准。

#### 4.4 实验结果

为了验证本文的算法有效性,徒手绘数据集上使用准确率进行了评估,结果见表 1。

由实验结果可知本方法在准确率方面表现的很优秀,精准度很高,完全可以用于实时的匹配,比传统的直接用特征搜索的方式更具有使用价值。

表1 不同方法在本数据集上的准确率的比较

Tab. 1 Compare the accuracy of different algorithms in this data set

算法选择	准确率 top-1/%
HOG+SVM	58.01
MKL-SVM <sup>[4]</sup>	66.81
FV-SP <sup>[5]</sup>	69.92
AlexNet	70.58
SN1.0 <sup>[6]</sup>	75.93
Inception V3	76.07
SN2.0 <sup>[7]</sup>	78.95
ResNet-50	79.88
VGG	88.02
Ours	98.32

## 5 结束语

将 VGG 算法应用到手写图像的匹配中,取得了非常不错的效果。本方法采用的改进后的 VGG 进行特征融合的语义提取,根据语义进行手写图像的匹配。实验表明在数据集很小的情况下,经过细致的训练,匹配准确度高达 98% 以上,综合表现十分优秀。此外本方法是两阶段方法,通过手写图像的

(上接第 147 页)

合方法,得到云存储空间多维资源数据聚类的随机概率密度分布集。基于决策边界的多数类样本分析方法,充分利用数据空间的类间指数分布,采用差分进化方法,遍历云存储空间多维资源数据聚类区域的候选目标集,实现对云存储空间多维资源数据的并行关联规则聚类 and 可靠性挖掘。研究得知,本文方法进行云存储空间多维资源数据聚类的收敛性较好,并行关联规则聚类性较强,提高了数据的检测识别率。

## 参考文献

- [1] 李艳霞, 柴毅, 胡友强, 等. 不平衡数据分类方法综述[J]. 控制与决策, 2019, 34(4): 673-688.
- [2] 吴园园, 申立勇. 基于类重叠度欠采样的不平衡模糊多类支持向量机[J]. 中国科学院大学学报, 2018, 35(4): 536-543.
- [3] 罗俊, 陈黎飞. 基于 BERT 的不完全数据情感分类[J]. 计算机应用, 2021, 41(1): 139-144.
- [4] 于爽, 李淑梅. 图书电子资源信息分布式多模块检索仿真[J]. 计算机仿真, 2020, 37(6): 439-442.
- [5] 徐玲玲, 迟冬祥. 面向不平衡数据集的机器学习分类策略[J]. 计算机工程与应用, 2020, 56(24): 12-27.

语义信息进行匹配,并不是直接利用特征匹配,因此本方法只能用于特定场景下的手写图像的匹配,寻找一个可以通用的手写图像的匹配仍有很大的研究空间。

## 参考文献

- [1] GLOROT X, BORDES A, BENGIO Y. Deep Sparse Rectifier Neural Networks [J]. Journal of Machine Learning Research, 2011, 15: 315-323.
- [2] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift [J]. arXiv preprint arXiv:1502.03167, 2015.
- [3] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting [J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [4] LI Y, HOSPEDALES T M, SONG Y Z, et al. Free-hand sketch recognition by multi-kernel feature learning [J]. Computer Vision and Image Understanding, 2015, 137: 1-11.
- [5] SCHNEIDER R G, TUYTELAARSY T. Sketch classification and classification-driven analysis using Fisher vectors [J]. ACM Transactions on Graphics, 2014, 33(6CD): 174.1-174.9.
- [6] YU Q, YANG Y, SONG Y Z, et al. Sketch-a-net that beats humans [J]. arXiv preprint arXiv:1501.07873, 2015.
- [7] YU Q, YANG Y, LIU F, et al. Sketch-a-Net: A Deep Neural Network that Beats Humans [J]. International Journal of Computer Vision, 2017, 122(3): 411-425.
- [8] 农艺, 唐忠. 综合用户属性和相似度的协同过滤推荐算法 [J]. 微型电脑应用, 2019, 35(11): 27-29.
- [9] 李景龙. 基于数据挖掘的电子商务卖家分层模型研究 [D]. 北京: 北京邮电大学, 2019.
- [10] 高海英. 针对计算机网络发展关系数据库技术 [J]. 微型电脑应用, 2020, 36(8): 31-33.
- [11] Jiang J, Ng K M, Poh K L, et al. Vehicle routing problem with a heterogeneous fleet and time windows [J]. Expert Systems with Applications, 2014, 41(8): 3748-3760.
- [12] AZADEH A, FARROKHI-ASL H. The close-open mixed multi depot vehicle routing problem considering internal and external fleet of vehicles [J]. Transportation Letters - The International Journal of Transportation Research, 2017, 11(2): 78-92.
- [13] XIAO Y, KONAK A. The heterogeneous green vehicle routing and scheduling problem with time-varying traffic congestion [J]. Transportation Research, Part E: Logistics and Transportation Review, 2016, 88: 146-166.
- [14] 王浩炎, 王玉玫. 两类认知用户及信道聚合的认知无线网络研究 [J]. 计算机工程与应用, 2020, 56(22): 100-108.
- [15] 程磊, 高茂庭. 结合时间加权和 LDA 聚类的混合推荐算法 [J]. 计算机工程与应用, 2019, 55(11): 160-166.
- [16] 刘生建, 杨艳, 周永权. 一种群体智能算法-狮群算法 [J]. 模式识别与人工智能, 2018, 31(5): 431-441.