

基于卷积神经网络的图像生成方式分类方法

李巧玲^{1,2}, 关晴骁^{1,2}, 赵险峰^{1,2}

(1. 中国科学院信息工程研究所信息安全国家重点实验室, 北京 100093;
2. 中国科学院大学, 北京 100049)

摘要: 提出一种采用卷积神经网络对自然图像和文档扫描图像进行分类的方法, 通过卷积和池化操作提取两类图像具有高区分度的特征, 融合后得到分类判决结果。实验结果表明, 所提出的分类方法在 SKL 图像库上分类精度超过 93%。图像预处理对模型的精度以及模型训练收敛所需时间具有积极效果, 经过图像预处理后训练的卷积神经网络模型对图像文字大小和图像格式稳健。

关键词: 卷积神经网络; 图像生成方式; 内容模式分类; 多媒体安全

中图分类号: TP37

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2016.00096

Image generation classification method based on convolution neural network

LI Qiao-ling^{1,2}, GUAN Qing-xiao^{1,2}, ZHAO Xian-feng^{1,2}

(1. State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China;
2. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Using convolution neural network which through convolution and pooling extracting features of high distinguish ability and then make fusion for classification of natural images and scanned documents. Experimental results show that the classification accuracy of the proposed classification method is more than 93% on the SKL image database. The model is highly robust to font sizes and image formats. Through contrast experiment validated that preprocessing of image has a positive effect on the accuracy of the model and the time cost on training.

Key words: convolution neural network, image generation mode, content pattern classification, multimedia security

1 引言

当前, 网络图像类型日益丰富, 这导致图像安全检测容易出现被测图像和检测模型失配问题, 媒体失配问题使图像安全检测方法的性能大大降低。造成图像类型日益丰富的主要原因是图

像生成方式较多, 这包括拍摄设备拍摄、计算机生成、扫描仪扫描等, 为了使图像安全检测技术适应媒体类型多样性的现状, 技术上需要按照生成方式对图像进行类型分类, 感知图像的类型可以为后续安全检测提供先验知识, 有助于提高后续图像安全检测的精度和效率。

收稿日期: 2016-07-16; 修回日期: 2016-08-09。通信作者: 李巧玲, liqiaoling@iie.ac.cn

基金项目: 国家自然科学基金资助项目 (No.61303259, No.U1536105); 中国科学院战略性先导科技专项课题基金资助项目 (No.XDA06030600); 中国科学院信息工程研究所重点基金资助项目 (No.Y5Z0131201)

Foundation Items: The National Natural Science Foundation of China (No.61303259, No.U1536105), The Strategic Pilot Science and Technology Project of the Chinese Academy of Sciences (No.XDA06030600), The Key Project of Institute of Information Engineering, Chinese Academy of Sciences (No.Y5Z0131201)

随着数字图像处理技术和机器学习领域的快速发展, 存在大量按照生成方式对图像内容模式进行识别的工作, 其中大多数工作集中在区分自然图像和计算机生成图像 (computer graphics) [1-3]。文献[1]提取基于小波直方图的 144 维特征, 输入到 FLD (fisher linear discriminant) 分类器对计算机生成图像和自然图像进行分类。文献[2]通过建立基于一阶和高阶小波统计量的统计模型, 揭示计算机生成图像和自然图像之间微妙的不同。在没有任何人工标注的前提下, 文献[3]通过训练卷积神经网络模型利用图像颜色、光照和内容的协调性分类自然图像和合成图像。在众多的网络传输图像中, 自然拍摄图像和扫描文档图像占到较大的比例, 而这 2 种图像成像方式、内容以及统计特性均有不同。因此, 如果不加区分, 容易造成一些系统的误检测率增高, 如文档扫描图像将极大程度地增加隐写分析系统的虚警率, 而自然图像由于其丰富的内容, 对用于检测文档图像中密级标识的密标检测系统也将带来影响。与自然图像和计算机生成图像的识别方法相比, 针对自然图像和扫描图像的分类手段相对较少。文献[4]是为数不多的检测扫描图像和自然图像的工作, 但该工作与之前大部分区分自然图像和计算机生成图像的工作类似, 采用较为传统的技术路线, 利用隐写分析特征和分类器实现。文献[4]根据图像生成过程的差异性提取不同特征。计算给定图像固定模式噪声的残差, 利用噪声残差的相关统计量构造 15 维的特征向量。使用 SVM 分类器对图像内容模式进行分类, 分类精度达到 89.4%。

传统的用于分类自然图像和文档扫描图像的方法虽然可以达到比较高的准确率, 但仍然存在一定的弊端: 计算单元有限, 无法支持大规模数据集的训练, 对于特征的表达有限。当扫描图像经过 JPEG 压缩之后再提取 15 维特征时, 文献[4]分类的准确率发生明显下降。研究过程中发现, 对于自然图像和文档扫描图像的分类问题具有以下 2 个难点。

1) 文档图像存在字体和字号多样性、版式多样性等问题, 且大量的表格、插图、纸张底纹、文档背景、扫描时的旋转、文档纸张的污损等均

会对分类造成较大的影响。

2) 自然图像中的纹理区域、标牌字符、某些符号等, 也容易对识别准确率造成影响。

传统的分类方法难以完全对这些问题顽健, 由于其特征设计一般依赖于人为经验, 因此难以设计出对以上问题均具有较好顽健性的特征。自然图像和文档图像种类极其丰富, 本文试图使用另一种技术途径解决该问题, 即使用大量多样的训练样本涵盖以上多种情况, 并使用学习能力较强的方法获取对以上多种条件均顽健且更具区分能力的检测模型。

基于上述事实, 本文提出一种高速高精度图像类型识别的方法, 主要针对自然图像和文档扫描图像进行分类。该方法采用深度卷积神经网络 (CNN, convolutional neural network), 利用多层卷积获取对图像内容模式具有高区分度的特征, 并融合得到分类判决信息, 为内容安全性检测提供先验依据, 减少后续不必要的检测, 提高内容安全性检测系统的准确性。图像分类与安全性检测过程如图 1 所示。本文围绕利用卷积神经网络分类自然图像和扫描文档进行探讨, 重点探讨采用多种图像预处理方法、学习方法对检测精度和模型训练收敛速度的影响, 并通过实验验证了合理的预处理对于模型的收敛速度和准确率具有积极作用。本文还对文档扫描图像的字体大小和图像格式的顽健性问题进行了相关实验论证。利用卷积神经网络对自然图像和文档扫描图像进行分类, 具有较好的精度和实时性, 可应用于网络在线媒体数据监控等领域, 具有重要且广泛的应用价值。

2 卷积神经网络简介

2.1 符号系统定义

为保证叙述的严谨性, 首先定义本文所使用的符号系统, 各符号在下文中, 如无特别说明, 则默认为本节所定义。本文涉及的符号系统主要如下。

定义训练样本 $(\mathbf{x}, y)$, $\mathbf{x}$ 为神经网络的输入, 在本实验中 $\mathbf{x}$ 为输入到网络的图像。 y 表示 $\mathbf{x}$ 的类别。 $(\mathbf{x}^i, y^i)$ 为第 i 个训练样本。 $\mathcal{S} = ((\mathbf{x}^1, y^1), (\mathbf{x}^2, y^2) \cdots (\mathbf{x}^n, y^n))$ 为整个训练样本集合。使用 $\mathbf{w}$ 和 $\mathbf{b}$ 对

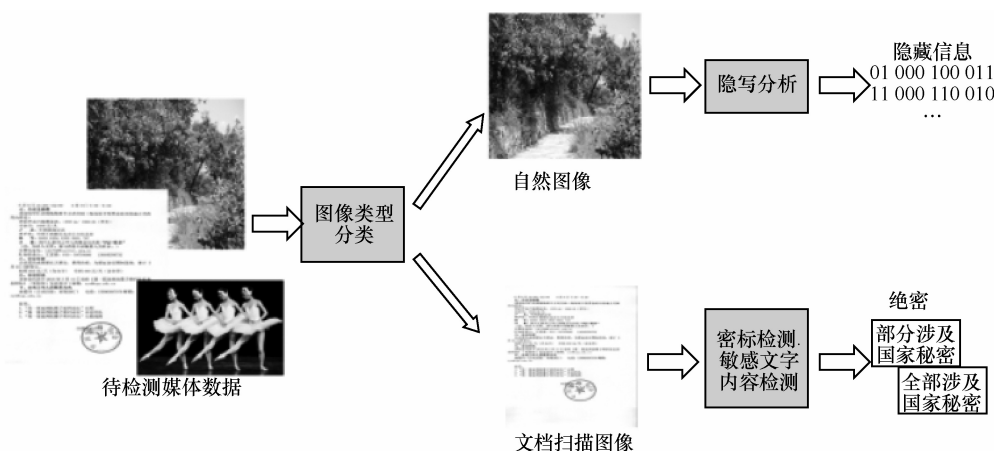


图1 图像分类后进行安全检测

神经网络所有参数进行表示, 在卷积神经网络中 w 代表卷积核, b 表示偏置向量。

2.2 神经网络

神经网络作为机器学习的一门重要技术, 从早期的感知机到目前蓬勃发展的深度学习, 已有了数十年的发展。神经网络的应用涉及各个领域, 如语音识别、机器翻译、人脸识别等。神经网络是通过模仿动物神经元之间传递、处理信息的模式。由简单的处理单元(神经元)相互连接构成一个复杂的网络结构, 整个神经网络是一个复杂的非线性系统。其变换过程可以描述为

$$g(x) = b + wh(c + vx) \quad (1)$$

h 为非线性激活函数, 常见的激活函数有 Tanh、Sigmoid 等。 w 、 v 分别为输出层和隐层的权重矩阵, 尺寸分别为 $n \times m$ 、 $m \times d$, n 为输出的类别数。 b 、 c 分别为输出层和隐层的 n 维和 m 维

偏置向量。如图2左侧所示, 神经网络对输入的 d 维向量, 经过隐层投影成一个 m 维的向量, 再输入到分类器进行分类。

神经元是神经网络的基本组成单元, 每个神经元是一个多输入单输出的信息处理单元, 图2右侧为单个神经元的处理过程, 该过程可以简单表示为

$$\begin{cases} u_i = \sum_{j=1}^d w_{ij} x_j + b_i \\ z_i = h(u_i) \end{cases} \quad (2)$$

将 $(x_1, x_2, \dots, x_d)$ 输入到神经元 i , 每个输入单元都进行加权平均, 权值系数和偏置都是经过训练学习而来。 z_i 为第 i 个神经元的输出, h 为非线性激活函数, 神经网络通过引入非线性的激活函数使网络能够学习出更好的特征表达, 解决线性模型所不能解决的问题。

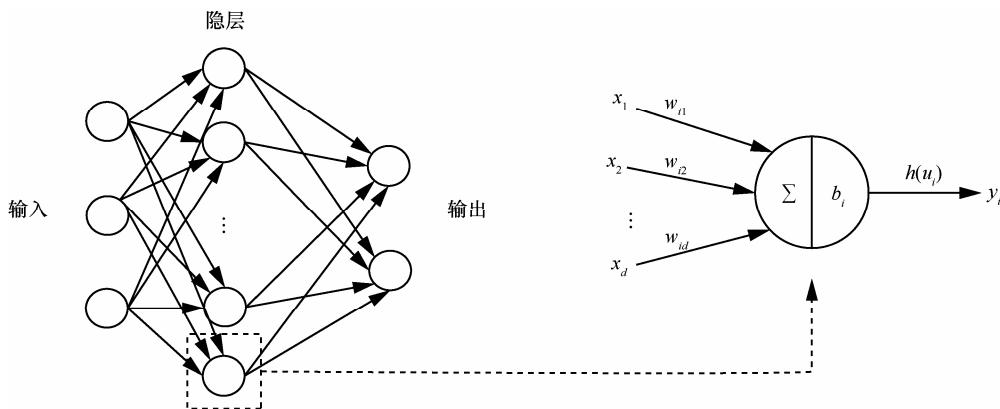


图2 神经网络模型以及单个神经元信息处理过程

神经网络是一个分层的有向图, 同层节点之间没有连接, 节点之间不能越层连接。上层输入经过非线性变换后作为下层神经元的输入。隐层的数目, 每层神经元的个数以及非线性函数的选择是构成神经网络的关键。神经网络使用 BP 算法从大量训练样本中学习出统计规律, 从而对未知事件做预测。只含有较少隐层的神经网络称为浅层模型, 其局限性在于有限的计算单元, 对特征的表示能力有限。当前, 神经网络已发展为深度模型, 与浅层模型相比, 深度模型通过构建多个隐层利用海量的训练数据, 自动地学习更有用的特征, 提升最终分类或预测的准确性。2.3 节介绍的卷积神经网络属于该类深度模型。

2.3 卷积神经网络

深度神经网络通过有监督或者无监督的方式学习层次化的特征表达, 对物体进行从底层到高层的特征描述。卷积神经网络是深度神经网络主流结构之一。最早出现在 20 世纪 80 年代, 最初应用于手写数字识别, 取得了很好的效果。卷积神经网络是在多层神经网络的基础上发展起来的针对图像分类而特别设计的一种深度学习方法。该网络的布局更加接近于生物神经网络。对于图像这种多维向量可以直接输入到网络, 无需进行复杂预处理。

卷积神经网络在传统的神经网络上加入卷积和池化层, 并引入了局部感受野、权值共享的机制, 大大减少了待训练的参数量。卷积层利用卷积核的移动来提取上层输入的局部特征, 然后非线性组合这些特征得到下层的输入, 逐层对图像特征进行抽象。卷积使图像原信号增强, 并且降低噪声, 保持了图像的空间信息, 因而特别适合于对图像进行表达; 池化层利用图像的局部相关性原理, 对卷积后的特征图进行子抽样, 在大大减小数据处理量的同时保留图像的有用信息, 并且保证特征图像对于旋转、平移等变换具有一定的顽健性, 常见的池化方法有 Max Pooling、Mean Pooling。Max Pooling 选择图像区域的最大值作为池化后的值; Mean Pooling 计算图像区域的平均值作为池化后的值。

文献[5]提出了一种特征可视化的方法, 通过

提取各层的特征图像进行可视化。探讨卷积神经网络每层对图像所做的具体操作。卷积神经网络通过逐层迭代, 提取特征。文献[5]认为卷积神经网络中下层的卷积主要提取图像的浅层特征, 如边缘、颜色、纹理等信息。越往上层提取的特征越高级。对特征的可视化, 可以进一步对网络结构进行调优。

卷积神经网络训练的自动地学习针对特定的分类任务最好的卷积核以及这些卷积核的组合方式, 计算出输入图像对于分类任务最好的特征表达。卷积神经网络的训练过程分为前向传播和后向传播 2 个过程, 前向传播为自底向上的学习过程, 输入图像 $\mathbf{x}^i$, 此时的 $\mathbf{x}^i$ 为像素值组成的阵列。经过一系列的卷积、池化和非线性变换得到一个较好的特征表达输入到 Softmax 分类器或 RBF 分类器等。前向传播可以简单表示成一系列的非线性变换: $\bar{\mathbf{y}}^i = G_k[\mathbf{w}_k(\cdots G_1(\mathbf{w}_1\mathbf{x}^i + \mathbf{b}_1)\cdots) + \mathbf{b}_k]$ 。 G_i 为非线性函数, 常用的非线性函数有 Tanh、Sigmoid、Relu 等, $\bar{\mathbf{y}}^i$ 为卷积神经网络对输入图像 $\mathbf{x}^i$ 预测值。反向传播为自顶向下的学习过程, 误差自顶向下传播, 利用不同的优化算法对网络参数进行微调, 达到最小化损失函数的效果。反向传播常使用的优化方法: 梯度下降 (gradient descent)、随机梯度下降 (stochastic gradient descent) 等。反向传播过程可简单地表述为 $\min_{\mathbf{w}, \mathbf{b}} L(\mathbf{w}, \mathbf{b}, \mathbf{S}), L(\mathbf{w}, \mathbf{b}, \mathbf{S})$ 为神经网络的损失函数。神经网络根据不同的任务选定不同的损失函数, 常见的损失函数有绝对损失 $\sum_{i=1}^n |\bar{\mathbf{y}}^i - \mathbf{y}^i|$ 、平方损失 $\sum_{i=1}^n |\bar{\mathbf{y}}^i - \mathbf{y}^i|^2$ 、Softmax 二类概率损失 $\sum_{i=1}^n -\mathbf{y}^i \log h_{\mathbf{w}, \mathbf{b}}(\mathbf{x}^i) - (1 - \mathbf{y}^i) \log [1 - h_{\mathbf{w}, \mathbf{b}}(\mathbf{x}^i)]$ 等, $\bar{\mathbf{y}}^i$ 为网络对第 i 个样本的预测值, $h_{\mathbf{w}, \mathbf{b}}(\mathbf{x}^i)$ 为 Softmax 分类器对样本 $\mathbf{x}^i$ 的预测概率。

卷积神经网络在语音识别和图像处理方面有着独特的优越性, 使其成为当前语音识别和图像识别领域的研究热点。以 Lecun^[6] 提出的 “LeNet-5” 为代表的卷积神经网络, 在手写数字识别任务上取得了不错的效果, Kussl 等^[7]提出的采用排列编码技术的神经网络在人脸识别和小物体识任务上有较好的应用。但目前神经网络在自然图像和文档扫描图像分类任务上的应用还存在空白, 本文以经典的 “LeNet-5” 为原型设计了

“ScanNet”，使用卷积神经网络对数字图像的内容模式进行分类。

3 针对简单分类任务卷积神经网络一般构建及其性质初步分析

随着硬件性能的提升和算法的不断优化，卷积神经网络已从只能完成简单的分类任务发展到能够超越人类识别能力的水平^[8]。针对不同复杂度的分类任务，往往需要构建不同的网络结构。对于简单的分类任务，一般会使用较少的卷积、池化和 Relu(rectified linear unit)^[9]非线性层，每层使用较少的卷积核数目来提取不同类别间的差异性特征。要完成较难的分类任务，神经网络的结构更加复杂，网络参数量更大，需要的操作种类更多。

相对于早期卷积神经网络中使用最多的 Tanh、Sigmoid 非线性函数而言，Relu^[9]： $f(x)=\max(0,x)$ 非线性函数可以增加隐层单元的稀疏性，减少计算量，加速网络收敛，并且 Relu 函数不存在饱和区域，反向传播时，避免了梯度消失的问题。

当训练样本不足，网络参数过多时模型会出现过拟合的现象。构建网络时为了防止过拟合现象的发生，一般采用在损失函数中加入 L1 或 L2 正则化项、early stopping、dropout^[10]等技术。dropout 以一定的概率将隐层的神经元暂时从网络中丢弃，丢弃就是对这些神经元的权重系数暂时不做更新，但是权值仍然保留，以便接下来输入的样本对其进行微调。

GPU 的发展提升了计算机的计算能力。为了充分利用 GPU 强大的计算能力，目前训练神经网络时采用批梯度下降代替原有的梯度下降和单样本的随机梯度下降，一次随机使用一批 (mini-batch) 样本的梯度对参数进行微调。当一小批样本包含的图像数目越多，训练所需的 GPU 显存会相应增加。每一小批图像在选择时引入了随机性，使网络从概率的角度考虑始终可以收敛。与使用单个样本的随机梯度下降法相比，批梯度下降每次可处理的数据量增多，提高了 GPU 的利用率，训练过程的效率得到了很好提升。与一次使用所有样本更新参数的梯度下降相比，计算开销减小，训练所需时间缩短。带有动量^[11] (momentum) 参数的随机梯度下降法使网络的收敛速度变

得更快。

$$\begin{cases} \mathbf{v}^{k+1} = r\mathbf{v}^k + a\nabla_{\mathbf{w}} L(\mathbf{w}, \mathbf{b}, \mathbf{x}) \\ \mathbf{w}^{k+1} = \mathbf{w}^k - \mathbf{v}^{k+1} \end{cases} \quad (3)$$

其中， $\mathbf{v}^k$ 、 $\mathbf{v}^{k+1}$ 分别为第 k 次和第 $k+1$ 次优化时梯度下降的速率， a 为学习速率， r 为动量参数。加入动量参数后下降速率相对于常规的方法要更大，需要相应地减小学习速率。一般动量初始化为 0.5，当模型趋于稳定时逐渐增加动量到 0.9。使用带有动量参数的随机梯度下降法，每一步梯度下降的量都需要参考前一步下降的量，使网络能够更快收敛，并且减小收敛到局部最优点的可能性。

卷积神经网络模型训练的速度和模型的精度受到多种因素的影响。如训练数据间往往存在较大的数值差异，使训练过程中误差下降不稳定，网络学习速率变慢甚至不收敛等。对数据进行减均值、z-score 标准化、白化操作可以消除不同特征分量之间的数值大小差异，改善网络的学习性能。减均值的计算如下。

$$\begin{cases} x_q = \frac{1}{nms} \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^s x_{(j,k)}^{i,q}, & q \in (R, G, B) \\ x_{(j,k)}^{i,q} - x_q, & j \in \{1, 2, \dots, m\} \quad k \in \{1, 2, \dots, s\} \end{cases} \quad (4)$$

n 为训练样本集大小，训练图像原始大小为 $m \times s$ 。 x_q 为训练样本在第 q 通道均值，RGB 三通道的均值分别为整个训练样本的所有像素点在该通道像素值的算术平均。 $x_{(j,k)}^{i,q}$ 为第 i 张图像第 (j, k) 位置像素点在 q 通道的取值。

4 实验环境及网络结构设置

4.1 实验环境

本文实验所使用训练样本和测试样本图像均来自 SKL 图像库。SKL 图像库包含 4 000 张自然拍摄图像和 1 500 张文档扫描图像。表 1 列出了用于建立 SKL 图像库所使用的拍摄设备和扫描设备名称。所有相机拍摄图像均采用 RAW 和 JPEG 这 2 种格式存储。扫描仪分别设置 3 种不同的分辨率：100×100 dpi、300×300 dpi、600×600 dpi，扫描文档存储为 JPEG 格式。图 3 和图 4 分别为 SKL 图像库中典型的自然图像和文档扫描图像示例。

表1 图像生成设备

图像类型	设备名称
自然图像	Canon M2、Canon 450D、Nikon D700、Panasonic GF5
扫描文档	Brother MFC-7860DN



图3 典型自然图像

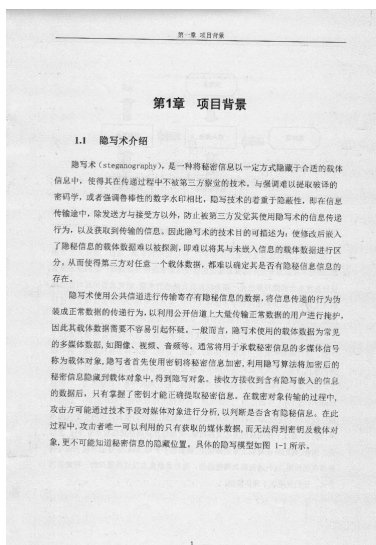


图4 典型扫描文档

训练样本包含1600张图像,其中800张JPEG格式的自然图像,文档扫描图像800张。测试样本包含600张图像,JPEG格式的自然图像300张,扫描文档图像300张,正负样本分布均衡。实验在Ubuntu 14.04上进行,训练过程利用两块NVIDIA GTX TITAN X,采用CUDA和GPU并行计算提升卷积神经网络的训练速度。

4.2 ScanNet 网络结构以及参数设置

适当的卷积神经网络结构对于实现特定的分类任务至关重要。本文涉及的分类任务主要需要表达自然图像与扫描文档的区分信息,从视觉表现而言,两类图像在局部和全局均存在具有区分能力的信息,需要建立从局部达到全局综合的

网络模型。因此,本文分别使用卷积层和全连接层达到以上目的。以经典的LeNet-5为原型设计了ScanNet结构,如表2所示,ScanNet包含3层卷积以及2层全连接。本文利用深度学习框架Caffe(convolutional architecture for fast feature embedding)^[12]搭建ScanNet。Caffe是一款开源的深度学习框架,拥有通用性强、性能高、代码可读性好等特点,支持多种数据类型,并且支持多GPU并行。

表2 ScanNet 网络结构

类型	核尺寸/步长	输出尺寸
Convolution	12×12/5	101×101×96
Max pool	3×3/2	50×50×96
Convolution	5×5/1	50×50×256
Max pool	4×4/2	24×24×256
Convolution	3×3/1	24×24×256
Max pool	4×4/2	11×11×256
Fully connection	—	1×1×512
Fully connection	—	1×1×2
Softmax	—	1×1×2

在ScanNet中,网络每层卷积后连接有Relu非线性层和LRN归一化层,经过归一化后每层的输入更加稳定,网络学习速度更快。经过不断的调优,设定初始学习速率为0.0001。通过迭代学习,损失函数会逐渐接近最小值,与此同时需要减小学习速率。因此,实验每迭代500次学习速率降为原来的 $\frac{1}{10}$ 。实验为一个二类概率分类问题,损失函数为

$$L(\mathbf{w}, \mathbf{b}, \mathbf{x}) = -\frac{1}{m} \sum_{i=1}^m y^i \log h_{\mathbf{w}, \mathbf{b}}(\mathbf{x}^i) + (1 - y^i) \log [1 - h_{\mathbf{w}, \mathbf{b}}(\mathbf{x}^i)] + \frac{\lambda}{2} \sum_{i=1}^k \sum_{j=1}^m w_{ij}^2 \quad (5)$$

其中, y^i 标签取值为0或者1(0标签代表扫描文档,1标签代表自然拍摄图像)。 $h_{\mathbf{w}, \mathbf{b}}(\mathbf{x}^i)$ 为Softmax分类器输出将 $\mathbf{x}^i$ 预测为第 y^i 类的概率。以上损失函数为加入了正则化项的二类概率损失,等式右侧第二项为正则化项。正则化项的加入使上式能够更容易得到全局最优解,防止模型过拟合,实验 $\lambda=0.0005$ 。 m 为一次输入到网络的图像数。

使用带有动量参数的随机梯度下降来学习网络的参数,将动量固定为 0.9,综合考虑网络的学习速度和 GPU 的显存,设置每一小批样本包含 32 张图像。通过对比实验发现对于目前的分类任务网络结构是否使用 dropout 操作对于模型在测试样本上的测试精度没有影响。

4.3 样本归一化对模型的影响及模型顽健性实验论证

本文主要围绕卷积神经网络在识别数字图像的内容模式方面的应用展开了研究。本文就图像减均值归一化操作对于模型的训练速度、测试精度的影响做了进一步的实验论证。实验总共训练 2 个模型,在网络结构和配置参数相同的情况下,模型 A 未对输入的训练样本和测试样本进行减均值中心化处理,模型 B 对训练样本和测试样本的每个像素点都分通道进行第 3 节所述的减均值预处理,经过归一化的训练样本间具有更加相似的分布,网络的训练过程更加高效。由于卷积神经网络只能处理固定尺寸的图像,训练样本和测试样本图像的尺寸不一,还需对图像进行尺寸处理。一般通过缩放、裁剪、扭曲将图像变换到网络要求的尺寸。在需要图像整体特征时通过扭曲可以保留整张原始图像信息。对于本分类任务而言无需图像的整体特征,经过缩放和裁剪之后自然图像和扫描图像之间的差异性特征不会发生改变。因此,本文在训练过程中首先将训练样本的原始图像缩放到 600×600 大小,再使用 512×512 大小的窗口对缩放之后的图像进行中心和四角的裁剪。在保证特征不变的前提下,同时增加两类样本的数量。并对训练样本进行随机置乱以保证训练出的分类模型对类别预测无倾向性。测试样本图像直接通过尺度缩放到 512×512 大小。

图 5 刻画损失函数与训练迭代次数之间的关系,随着迭代次数的增加损失函数逐渐下降并最终达到稳定,当损失函数保持稳定时模型达到收敛。如图 5 所示,训练样本和测试样本未进行归一化操作时,网络需进行 800 次迭代达到收敛,在测试样本上的最优测试精度为 0.97;但进行归一化操作之后,经过 400 次迭代模型便达到稳定,并且此时的测试精度可以达到 0.99,训练的效率要明显高于未经过归一化的数据。通过进一步的实验发现经过归一化操作训练得到的模型对于扫描文件字符大小的顽健性要强于未经过归一化操作训练得到的模型。

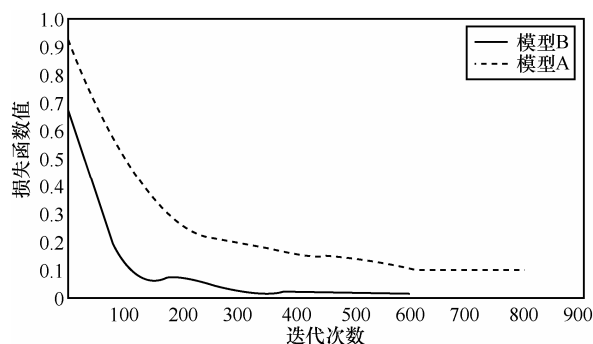


图 5 模型 A 与模型 B 迭代次数和损失函数关系

使用模型 B 对 100 张扫描文档和自然图像进行分类。待分类图像包含扫描文档和自然图像各 50 张,自然图像编号为 1~50,扫描图像编号为 51~100。图 6 为模型 B 对一张图像进行预测的基本流程。将待分类图像输入网络之前先进行减均值、尺寸变换等预处理,然后再输入到神经网络对图像类型进行预测。图 7 为模型 B Softmax 层对 100 张图像输出的分类概率。分类时使用一块 NVIDIA GTX TITAN X,每张图像分类平均耗时 0.7 s,当增加 GPU 的数量,通过多线程可增

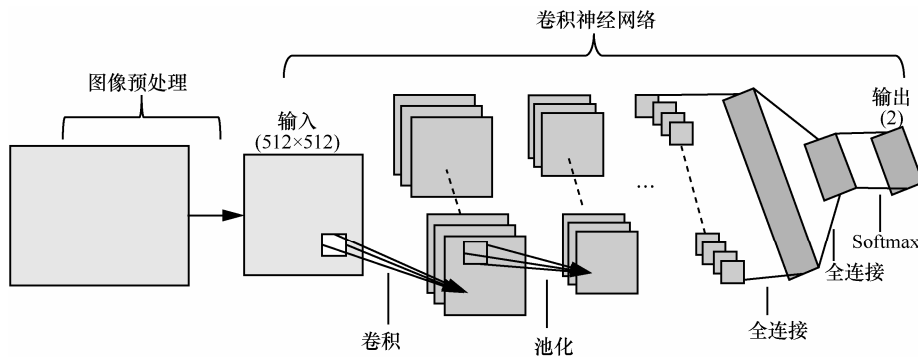


图 6 模型 B 预测过程

加一次性预测分类的图像张数。图中圆形代表被误分的自然图像，菱形表示被错误分类的扫描图像，三角形代表预测类型为自然图像，矩形表示预测类型为文档扫描图像。分类混淆矩阵如表3所示，模型B的平均准确率可达到94.0%。对未能正确分类的扫描文档图像分析发现这类图像具有明显的共性是不含文字。卷积神经网络模型为数据驱动模型，模型的准确率依赖于数据集包含的图像种类和数量。造成模型对此类图像无法正常分类的原因可能是由于训练样本中不包含这类不含文字的文档图像，导致无法正确提取具有高区分度的特征。后续实验可以将此类图像加入到训练样本中，丰富训练样本类型，以达到更高的识别准确率。

表3 模型B混淆矩阵

实际类型	预测类型	
	自然图像	扫描文档
自然图像	49	1
扫描文档	5	45

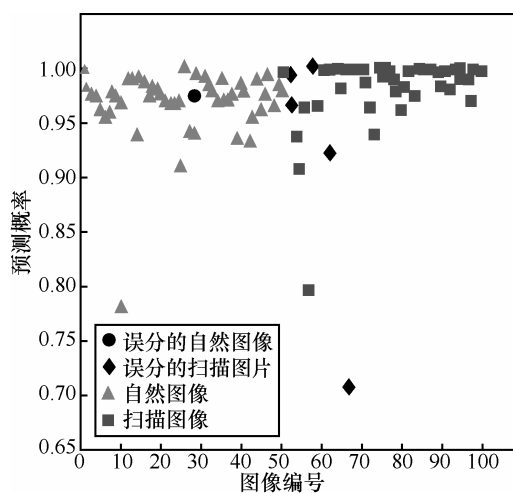


图7 模型B分类概率

文档图像存在字体和字号多样性、版式多样性等问题。本文还从模型对扫描文档的文字大小、图像存储格式的顽健性这2个方面进行了研究。扫描文档文字大小选用八号到初号不同大小的字符，模型B分类结果的准确率可以达到97%。模型A识别的准确率只能达到50%。用模型B对JPEG、TIFF、BMP、PNG格式的图像进行分类，识别的准确率也可达到97%。

以上两组实验表明，经过归一化预处理的模

型对于文档字符的大小，以及图像的格式具有很强的顽健性。这对于后续对图像进行安全检测具有深远意义。

5 结束语

图像类型的日益丰富，对隐写分析、图像内容取证、失泄密检查等图像内容安全检测技术提出了挑战。为了应对图像安全检测技术面临的挑战，适应媒体类型多样性的现状，本文使用卷积神经网络按照图像的生成方式对图像进行类型分类。通过卷积和池化操作提取自然图像和文档扫描图像间具有高区分度的特征，构建高速高精度图像类型识别系统。所提出的分类方法在SKL图像库上的分类精度超过93%。训练的卷积神经网络模型对于图像文字大小和图像格式顽健。本文通过对比实验验证了图像预处理对于模型的精度以及模型训练收敛所需时间具有积极效果。

感知图像的类型有助于提高图像安全检测的精度，对后续的安全检测具有显著意义。除了自然图像和文档扫描图像，计算机合成图像与屏幕截图在进行安全检测前也需要按照生成方式分类。后续实验还会将计算机合成图像和屏幕截图加入到训练样本中，构建更加复杂的网络结构，训练出能对更多内容模式的图像进行准确分类的模型。目前实验使用1600张图像进行训练，后续实验将继续丰富训练样本的数量和类型，通过大样本训练出更加高精度的模型。

参考文献：

- [1] WANG Y, MOULIN P. On discrimination between photorealistic and photographic images[C]//IEEE International Conference on Acoustics Speech and Signal Processing. 2006.
- [2] LYU S, FARID H. How realistic is photorealistic[J]. IEEE International Conference on Signal Processing, 2005, 53(2): 845-850.
- [3] ZHU J Y, KRAHENBUHL P, SHECHTMAN E, et al. Learning a discriminative model for the perception of realism in composite images[C]//IEEE International Conference on Computer Vision. 2015: 3943-3951.
- [4] KHANNA, N, CHIU G T C, ALLEBACH J P, et al. Forensic techniques for classifying scanner, computer generated and digital camera images[C]//IEEE International Conference on Acoustics,

- Speech and Signal Processing. 2008: 1653-1656.
- [5] MAHENDRAN A, VEDALDI A. Understanding deep image representations by inverting them[C]//IEEE Conference on Computer Vision and Pattern Recognition. 2015: 5188-5196.
- [6] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [7] KUSSUL E, BAIDYK T, WUNSCH II D C. Permutation coding technique for image recognition system[M]. Neural Networks and Micromechanics, 2010: 47-73.
- [8] KAIMING H, XIANGYU Z, SHAOQING R, et al. Delving deep into rectifiers: surpassing human-level performance on imagenet classification[C]//IEEE International Conference on Computer Vision. 2015: 1026-1034.
- [9] NAIR V, HINTON G E. Rectified linear units improve restricted boltzmann machines[C]//The 27th International Conference on Machine Learning. 2010: 807-814.
- [10] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. The Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [11] SUTSKEVER I, MARTENS J, DAHL G, et al. On the importance of initialization and momentum in deep learning[C]//The 30th International Conference on Machine Learning. 2013: 1139-1147.
- [12] JIA Y, SHELHAMER E, DONAHUE J, et al. Caffe: convolutional architecture for fast feature embedding[C]//The ACM International Conference on Multimedia. 2014: 675-678.

作者简介:



李巧玲(1992-),女,湖北宜昌人,中国科学院信息工程研究所硕士生,主要研究方向为信息对抗理论与技术。



关晴骁(1984-),男,湖南湘潭人,博士,中国科学院信息工程研究所助理研究员,主要研究方向为多媒体内容安全、通信隐写分析。



赵险峰(1969-),男,安徽淮北人,博士,中国科学院信息工程研究所研究员、博士生导师,主要研究方向为信息安全事件检测分析的理论与技术,包括信息隐藏及其检测、网络安全异常行为检测、大数据安全分析以及相关技术在内容保护、版权保护和系统防护等中的应用。