

DOI:10.13718/j.cnki.xsxb.2019.05.018

# 基于 CNN 的计算机生成图像识别方法<sup>①</sup>

秦毅<sup>1</sup>, 吴蔚<sup>2</sup>

1. 重庆电子工程职业学院 人工智能与大数据学院, 重庆 401331;

2. 重庆电子工程职业学院 通识教育与国际学院, 重庆 401331

**摘要:** 针对计算机生成图像(Computer Generated images, CG)与真实照片(Photograph, PG)识别率不高的问题, 该文提出了一种改进的卷积神经网络方法来实现 CG 与 PG 的识别. 该方法首先对识别问题进行卷积神经网络二分类建模, 并选择 VGG-19 网络结构作为基础, 建立不同的模型. 该方法创新性地引入迁移学习, 节省训练时间和大量计算资源, 最后使用 softmax 分类器进行分类. 实验结果表明, 该文方法对 PG 图像的识别准确率达到 92%, 与其他方法比较, 该文方法识别准确率最高, 说明该文方法具有可行性与有效性.

**关键词:** 计算机生成图像; 迁移学习; 卷积神经网络; 图像识别

中图分类号: TP391

文献标志码: A

文章编号: 1000-5471(2019)05-0109-06

随着计算机图形学等技术的发展, 计算机生成图像(Computer Generated images, CG)的逼真程度已经能让肉眼无法区分<sup>[1-2]</sup>. 这对那些致力于虚拟生成真实图片的行业很有利, 如游戏、影视行业, 但同时也会带来危害. 例如, 一个影视人能够修改一幅图像欺骗众人, 这使得它变成了一个具有欺骗性的武器. 因此, 辨别一幅逼真的生成图像和一幅真实照片(Photograph, PG)具有重要的意义. 图 1 显示了真实(PG)人脸图像和生成(CG)人脸图像是很难辨别的.

近期的一些研究<sup>[3-6]</sup>已经涵盖了 CG 图像和视频检测. 数字取证方法已经被提出. 尽管如此, 识别 CG 图像的技术能力仍然有很大的提升空间. 文献[1]提出一种基于局部二进制计数模式的 CG 图像盲鉴别算法, 通过颜色空间的转换、归一化直方图和支持向量分类, 实现 CG 图像识别. 文献[7]提出一种新的用于区分视频中计算机生成的人脸和人脸的取证技术. 这种技术可以识别血流变化引起面部外观的微小波动, 由于这些变化是由人体脉搏引起的, 因此它们不太可能在计算机生成的图像中找到. 该算法使用这种生理信号的缺席或存在来区分真实人脸图像和计算机产生的人脸图像.



图 1 PG 图像(左)和 CG 图像(右)

文献[8]中提出通过局部二进制计数模式实现对计算机生成图像的识别, 该算法对原始图像的颜色空间提取局部二进制计数模式矩阵, 求取矩阵归一化直方图, 最后采用支持向量机(SVM)进行分类. 该方法

① 收稿日期: 2018-03-23

基金项目: 重庆市教委科学技术研究项目(KJ1602906;KJ1729408).

作者简介: 秦毅(1980-), 男, 副教授, 主要从事计算机应用技术研究.

能有效识别自然图像和计算机生成图像,但缺点是提取的特征维度过高.文献[9]提出一种基于局部三值计数的新颖算法来分类照片级逼真的计算机图形和照片图像.实验结果表明,该算法有效地降低了分类特征的维数,保持了良好的分类性能,缺点是当图像数据集过大时,用于训练的时间过长.

针对上述图像识别方法的问题,本文提出了一种改进的卷积神经网络方法实现计算机生成图像的识别,并在数据训练时引入迁移学习策略,能够降低特征维度和提高识别精度.本文将这个区分 CG 和 PG 图像问题建模为一个分类问题,对于给定的一幅图像,只需对这幅图像预测出一个对应的“CG”或者“PG”的标签.

## 1 本文方法

本文使用 Caffe 深度学习框架来实现提出的方法,同时还在训练过程中进行数据增广以提高模型的泛化能力.数据增广包括 2 个主要的操作:输入图像水平翻转和随机平移.随机平移包括水平和竖直 2 个维度.

### 1.1 卷积神经网络框架

受猫视觉皮层电生理研究启发,有研究者提出卷积神经网络(convolutional neural networks, CNN)与普通神经网络的区别在于,卷积神经网络包含了 1 个由卷积层和子采样层构成的特征抽取器.卷积神经网络由 3 个部分构成:①输入层,②由  $n$  个卷积层和池化层组合组成,③由一个全连结的多层感知机分类器构成.

CNN 由输入层、卷积层、激活函数、池化层、全连接层组成,主要通过卷积层和池化层对输入图像进行处理,其结构见图 2.

卷积层用来提取图像特征,在这一层,卷积神经网络通过 2 个特有方式降低参数数目:①局部感受野,②权值共享.卷积层越多,特征的表达力越强,第  $l$  层卷积层的表达式为

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} \times k_{ij}^l + b_j^l\right) \quad (1)$$

其中  $x_j^l (j = 1, 2, \dots, N^l)$  表示该层的特征图,  $M_j$  为输入层感受野,  $k_{ij}^l$  为卷积核,  $b_j^l$  为偏置,  $f(\cdot)$  表示激活函数,一般激活函数有 sigmoid 函数、tanh 函数和 RELU 函数,本文使用 RELU 函数作为激活函数,表达式为

$$h^{(i)} = \max(w^{(i)T}x, 0) = \begin{cases} w^{(i)T}x, & w^{(i)T}x > 0 \\ 0, & \text{else} \end{cases} \quad (2)$$

其中  $i$  表示隐藏单元个数,  $w^{(i)}$  表示权值.

池化层连接在卷积层后面,池化层的对象是特征图的局部区域,这样做能够使特征具有一定的空间不变性.具体过程为:对于卷积层得到的输入图像的邻域特征图(feature map, FM),使用池化技术,对 FM 降采样得到新的特征,这一步减少参数,降低特征维度,并且保证了特征具有空间不变性的能力.池化操作中最常使用的方法有随机池化(stochastic-pooling)、最大池化(max-pooling)和均匀池化(mean-pooling).

mean-pooling,即对邻域内特征点只求平均;max-pooling,即对邻域内特征点取最大;stochastic-pooling 则介于两者之间,通过对像素点按照数值大小赋予概率,再按照概率进行降采样.本文采用最大池化操作,表示形式为

$$x_j^l = f(\mu_j^l p(x_i^{l-1}) + b_j^l) \quad (3)$$

其中,  $f(\cdot)$  表示激活函数,  $\mu_j^l$  表示权重系数,  $p(\cdot)$  表示池化操作,  $b_j^l$  表示偏置.最终将输出特征作为

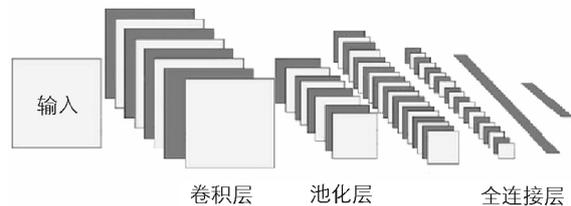


图 2 卷积神经网络结构

输出层 Softmax 回归的输入, 通过损失函数实现对 CGs 和 PGs 的识别.

对于卷积神经网络架构, 本文选用在 ImageNet 中取得成功的 VGG19 网络结构为基础网络, 整体的模型框架如图 3 所示. VGG19 网络结构主要的贡献是采用了一个非常小的 3x3 的卷积核, 通过将深度 CNN 网络增大到 16 ~ 19 层, 大大提高了现存的性能.

### 1.2 迁移学习

基于卷积神经网络方法需要大量的训练数据, 这导致训练需要很长的时间, 因此迁移学习<sup>[10-11]</sup>得到了关注. 这项工作说明: 任何训练在自然图像上的深度神经网络都表现出一种奇怪的现象: 从第一层学习到的特征与 Gabor 滤波器相似. 在许多类型的数据集和任务中都是通用的. 如果最后一层由特定的数据集训练, 这些特征可以从一般转换到特定情况.

迁移学习就是利用之前学到的知识来帮助完成新环境下的学习任务, 之所以在卷积神经网络中引入迁移学习是基于以下原因: 目前大多数成功的模型都是依赖于大量的有标签数据, 而很多学习任务很难获得大量的有标签数据, 对于每一个任务都从头开始训练, 成本非常高. 迁移学习定义:

给定源域(source domain)  $D_s$  和对应的任务  $T_s$ , 给定目标域(target domain)  $D_t$  和对应任务  $T_t$ . 迁移学习即是在  $D_s \neq D_t$  或  $T_s \neq T_t$  时, 利用  $D_s$  和  $T_s$  中的知识来帮助学习  $D_t$  上的预测函数  $f_t(\cdot)$ .

对于 Domain(域), 其定义为

$$D = \{\chi, P(X)\} \tag{4}$$

其中  $\chi$  为特征空间,  $P(X)$  为边际概率分布,  $X = (x_1, x_2, \dots, x_n) \in \chi$ . 对于 Task(学习任务) 定义为

$$T = \{y, f(\cdot)\} \tag{5}$$

其中  $y$  为标签空间,  $f(\cdot)$  表示通过训练集  $\{x_i, y_i\}$  来训练, 当用来预测  $X$  的标签时,  $f(\cdot)$  可以表示为  $P(y | X)$ .

迁移学习代表了转移从一个问题上学习到的知识到另一个问题的可能性. 一般来说, 迁移学习过程包括传递(源)神经网络的参数, 根据特定数据集上的特定任务, 将这个网络预训练为另外一个(目标)网络, 这个目标网络有一个不同的数据集来解决另一个不同的问题. 迁移学习简单且成本效益很高, 一方面迁移学习节省了很多训练时间和大量的计算资源; 另一方面如果从零开始训练, 就需要大量的训练图像, 然而一些特定任务的数据集图像通常不是很大. 迁移学习很好地解决了这两个问题. 因此, 本文通过利用预训练的 VGG-19 网络进行迁移学习.

迁移学习策略取决于多种因素, 但最重要的 2 个是数据集的大小以及新数据与原数据集的相似度. 谨记网络前面几层的 CNN 特征更加泛型(generic), 在后面几层中更加具有数据集特定性(dataset-specific). 通过持续的反向传播来微调预训练 CNN 的权重, 达到 CNN 最佳性能.

### 1.3 损失函数

在本文的工作中, 区分 CGs 和 PGs 的问题被建模为一个分类问题. Softmax 损失函数是在基于卷积神

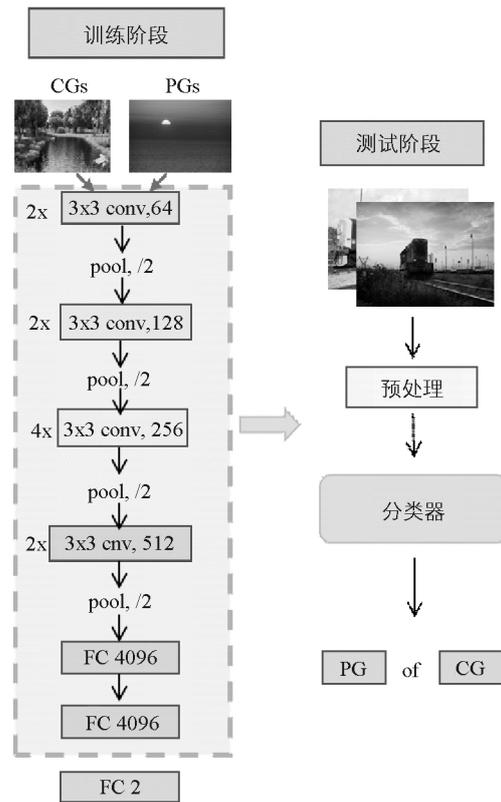


图 3 修改后的 VGG-19 网络结构

神经网络的方法中最常用的损失函数,它是一个二元逻辑回归到多元的推广,Softmax 损失函数的形式如公式(6)或公式(7)所示.其中  $f_j$  表示第  $j$  个元素的分数量,Softmax 函数如公式(8).

$$L_i = -\log \left( \frac{e^{f_{y_i}}}{\sum_j e^{f_j}} \right) \quad (6)$$

$$L_i = -f_{y_i} + \log \left( \sum_j e^{f_j} \right) \quad (7)$$

$$f_j(z) = \frac{e^{z_j}}{\sum_k e^{z_k}} \quad (8)$$

Softmax 损失函数源于交叉熵损失函数,用于评价一个真实类别的分布  $p$  和一个估计分布  $q$  之间的关系,如公式(9)所示.

$$H(p, q) = -\sum_x p(x) \log q(x) \quad (9)$$

## 2 实验结果与分析

在 2 个图像数据库上对本文方法进行实验,数据库 1 是自行建立的数据库,数据库中 2 000 张计算机生成图像和 2 000 张真实图像,这些图像都以 JPEG 图像格式压缩,大小介于 12 kb 到 1.8 Mb 之间,图像库内容分类范围广,包括人物、风景、动物和建筑等. CG 图像中的 400 幅来自 Columbia University CG 库,1 600 幅来自国内外的 CG 网站; PG 图像中的 1 600 幅来自 Columbia University PG 图像库,400 幅来自其他 PG 图像库.数据 2 是 DSTok 数据集,包括 4 850 张 CG 图像和 4 850 张 PG 图像,这些图像涵盖了汽车、动物以及户外等多个领域,都以 JPEG 图像格式压缩,大小介于 12 kb 到 1.8 Mb 之间.一些图像如图 4 所示,数据集里的图像都被缩放为  $224 \times 224$ ,第一行为 PG 图像,第二行是 CG 图像.

使用五折交叉验证协议,展示本文所有实验的平均精度.本文设置初始学习率为 0.001,衰减系数为每迭代 2 万次下降 0.1 倍.本文设置训练批次大小为每批次 64 张图像,并且权重衰减系数设为 0.0005,优化方法为 SGD (Stochastic Gradient Descent),动量因子设为 0.9.

实验中训练样本和测试样本的比例是 4 : 1.最后一层全连接 softmax 网络层会影响整个随机初始化网络的训练,为了减少影响,在进行微调整个网络之前,先对最后一层进行了预训练.在 DSTok 数据集上对本文方法进行实验,实验结果见表 1.



图 4 DSTok 数据集例子

表 1 本文方法验证结果

折	VGG19(无迁移)	VGG19(迁移)	VGG19(迁移+微调)
1	0.74	0.81	0.92
2	0.67	0.77	0.91
3	0.75	0.8	0.92
4	0.69	0.72	0.93
5	0.69	0.75	0.92
平均	0.71	0.77	0.92

由表 1 中数据可知,使用迁移学习和微调来进行训练的模型取得了最佳性能.而迁移学习模型的性能也优于不使用迁移学习训练的模型,由此可以说明迁移学习和微调策略具有有效性和鲁棒性.表 2 为本文

模型与现有最先进的方法在图像数据库 1 上的对比结果。

表 2 本文方法与其他方法结果对比

方法	CG	PG	平均准确率
CON	0.918	0.887	0.902
CUR	0.806	0.805	0.805
GLC	0.64	0.63	0.635
HOG	0.754	0.72	0.74
HSC	0.818	0.787	0.802
LBP	0.904	0.838	0.871
Li	0.948	0.911	0.93
LSB	0.672	0.651	0.662
LYU	0.942	0.899	0.92
POP	0.57	0.575	0.573
SHE	0.748	0.677	0.713
SOB	0.554	0.552	0.553
VGG19	0.922	0.918	0.92

表 2 表明本文所建立的模型在最先进的模型中具有竞争力。在自建图像数据集上, 本文的平均精度达到了 92%, 这些结果证明本文的迁移学习技术和微调策略比已有方法性能更好。

### 3 结 论

本文为了解决计算机生成图像与真实照片识别率不高的问题, 提出了一种微调迁移学习的卷积神经网络方法来实现计算机生成图像检测, 该方法采用了 VGG-19 网络架构, 并在数据训练时引入迁移学习策略, 节省很多训练时间和大量计算资源, 提高工作效率和识别精度。实验结果表明, 在自建图像数据库和 DSTok 数据集上, 相比已有方法, 本文提出的模型取得了更好的识别精度, 说明本文方法的可行性与有效性。

#### 参考文献:

- [1] 申铨京, 李梦臻, 吕颖达, 等. 基于 LBC 的计算机生成图像盲鉴别算法 [J]. 计算机科学, 2015, 42(6): 135-138, 161.
- [2] 敖道敢. 无监督特征学习结合神经网络应用于图像识别 [D]. 广州: 华南理工大学, 2014.
- [3] DE REZENDE E R S, RUPPERT G C S, CARVALHO T. Detecting Computer Generated Images with Deep Convolutional Neural Networks [C]//2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). Niteroi: IEEE, 2017.
- [4] 刘内美. 高识别率的鲁棒 ART 快速图像识别算法 [J]. 西南师范大学学报(自然科学版), 2016, 41(8): 106-113.
- [5] PENG F, ZHOU D L, LONG M, et al. Discrimination of Natural Images and Computer Generated Graphics Based on Multi-Fractal and Regression Analysis [J]. AEU-International Journal of Electronics and Communications, 2017, 71: 72-81.
- [6] TOKUDA E, PEDRINI H, ROCHA A. Computer Generated Images vs. Digital Photographs: A Synergetic Feature and Classifier Combination Approach [J]. Journal of Visual Communication and Image Representation, 2013, 24(8): 1276-1292.
- [7] CONOTTER V, BODNARI E, BOATO G, et al. Physiologically-Based Detection of Computer Generated Faces in Video [C]//2014 IEEE International Conference on Image Processing (ICIP). Paris: IEEE, 2014.
- [8] 张 震, 杨宇豪. 基于 Benford 模型的自然图像与计算机生成图像的鉴别 [J]. 北京工业大学学报, 2013, 39(6): 930-935.
- [9] TAN D Q, SHEN X J, QIN J, et al. Detecting Computer Generated Images Based on Local Ternary Count [J]. Pattern Recognition and Image Analysis, 2016, 26(4): 720-725.
- [10] 魏晓聪, 林鸿飞. 面向迁移学习的文本特征对齐算法 [J]. 计算机工程, 2017, 43(2): 215-219, 226.

- [11] SHIN H C, ROTH H R, GAO M C, et al. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning [J]. IEEE Transactions on Medical Imaging, 2016, 35(5): 1285-1298.

## On Computer Generated Image Recognition Method Based on CNN

QIN Yi<sup>1</sup>, WU Wei<sup>2</sup>

1. School of Artificial Intelligence and Big Data, Chongqing College of Electronic Engineering, Chongqing 401331, China;

2. School of General Education and International Academy, Chongqing College of Electronic Engineering, Chongqing 401331, China

**Abstract:** In order to solve the problem low recognition rate for Computer Generated images (CG) and Photographs (PG), an improved convolution neural network method is proposed to realize the recognition of CG and PG. This method first set up the two-classification model of the convolution neural network for the recognition problem and the VGG-19 network structure is selected as the basis to establish different models. This method innovatively introduces migration learning and saves training time and massive computing resources. Finally, softmax classifier is used to classify. The experimental results show that the accuracy of the proposed method for PG image recognition is up to 92%, and the recognition speed is faster. Compared with other methods, the method has the highest recognition accuracy and demonstrates the feasibility and effectiveness of the proposed method.

**Key words:** computer generated images; transfer learning; convolutional neural network; image identification

责任编辑 夏娟