

基于深度卷积生成对抗网络的图像识别算法

刘恋秋

引用本文:

刘恋秋. 基于深度卷积生成对抗网络的图像识别算法[J]. 液晶与显示, 2020, 35(4): 383–388.

LIU Lian-qiu. Image recognition algorithms based on deep convolution generative adversarial network[J]. *Chinese Journal of Liquid Crystals and Displays*, 2020, 35(4): 383–388.

在线阅读 View online: <https://doi.org/10.3788/YJYXS20203504.0383>

您可能感兴趣的其他文章

Articles you may be interested in

基于密集连接生成对抗网络的图像颜色迁移

Image color transfer with dense connections generative adversarial networks

液晶与显示. 2020, 35(3): 272–280 <https://doi.org/10.3788/YJYXS20203503.0272>

基于深度卷积神经网络的多特征融合的手势识别

Multi-feature fusion gesture recognition based on deep convolutional neural network

液晶与显示. 2019, 34(4): 417–422 <https://doi.org/10.3788/YJYXS20193404.0417>

基于深度可分离卷积的交通标志识别算法

Traffic sign recognition algorithm based on depthwise separable convolutions

液晶与显示. 2019, 34(12): 1191–1201 <https://doi.org/10.3788/YJYXS20193412.1191>

自由立体显示中基于深度卷积神经网络的虚拟视点生成方法

Virtual viewpoint image generation method based on deep convolutional neural network in autostereoscopic display

液晶与显示. 2019, 34(11): 1031–1036 <https://doi.org/10.3788/YJYXS20193411.1031>

基于深度学习的航空对地小目标检测

Detection of small target in aerial photography based on deep learning

液晶与显示. 2018, 33(9): 793–800 <https://doi.org/10.3788/YJYXS20183309.0793>

文章编号:1007-2780(2020)04-0383-06

基于深度卷积生成对抗网络的图像识别算法

刘恋秋*

(重庆财经职业学院,重庆 402160)

摘要:针对传统深度卷积生成网络收敛速度慢、稳定性较差的问题,本文在传统深度卷积生成对抗网络的基础上,提出了深度卷积生成对抗网络的优化算法。首先在预处理部分,融合了 Canny 算子和 Prewitt 算子的多个方向的卷积核来初始化输入图片参数,同时训练模块。为了减少训练时间,将训练分为 3 个阶段,每个阶段都采用不同的损失函数,从而提升网络的收敛速度及识别效果。最后再将训练后的判别网络中的卷积神经网络用来提取图像特征。LFW 和 CIFAR-100 的实验证明,本文提出的算法具有很高的可行性和有效性,比传统生成对抗网络、CNN 等图像识别具有更高的识别成功率,达到 89.5%,为生成对抗网络在计算机视觉领域的应用提供了有益的参考。

关键词:生成对抗网络;深度卷积;特征提取;计算机视觉

中图分类号:TP391 **文献标识码:**A **doi:**10.3788/YJYXS20203504.0383

Image recognition algorithms based on deep convolution generative adversarial network

LIU Lian-qiu*

(Chongqing College of Finance and Economics, Chongqing 402160, China)

Abstract: In view of the slow convergence and poor stability of traditional deep convolutional generation networks, this paper combines the Connaught kernels of multiple directions of Canny operator and Prewitt operator on the basis of traditional deep convolution generation confrontation network. To initialize the input picture parameters and reduce the training time, the training is divided into three stages, each stage adopts a different loss function, thereby improving the convergence speed and recognition effect of the network and then discriminating the volume in the network after training. The neural network is used to extract image features. The experiments of LFW and CIFAR-100 prove that the proposed algorithm is highly feasible and effective, and provides a useful reference for generating anti-network applications in the field of image recognition.

Key words: generate confrontation network; deep convolution; feature extraction; computer vision

1 引言

生成对抗网络(GAN)是近年来最受关注的

无监督式神经网络之一,该模型功能强大且应用场景广泛,最常见的应用是图像生成。在 GAN 的框架中^[1],结果的生成是通过对抗过程不断进化,在 GAN 中同时有两个模型训练:捕获数据分

收稿日期:2019-07-03;修订日期:2019-10-21.

* 通信联系人, E-mail:liulianqiu0627@hotmail.com

布的生成模型 G 以及估计数据分布的判别模型 D 。生成器 G 的训练目标是骗过判别器 D , 换言之最大化 D 犯错误的概率; D 的目标则是最大化自己的正确率。该框架本质上是一个博弈过程, 最终收敛于纳什均衡。在空间上, 对于生成器 G 和判别器 D , 如果使用零和博弈的 loss 函数, 当 D 训练完美达到 $1/2$ 时, D 就无法再分辨出真实数据和生成数据, 也就无法再给 G 提供梯度。

为了进一步提升 GAN 的稳定性和识别效率, 有很多学者都提出了自己的优化方案, 提出了虚拟批量标准化的算法, 利用批量训练数据的方式显著改善网络性能。Radford 提出了深度卷积生成对抗网络 (DCGAN) 的算法^[2], 该算法将 GAN 的概念扩展到卷积神经网络中, 可以生成更高质量的图片。在此基础上, Sailimans 等在近红外光谱 (NIR) 图像的场景下提出了着色 DCGAN 模型^[3], 该方案的核心是针对 NIR 图像将其划分为 RGB 3 个通道, 再利用 DCGAN 的模型分别进行训练, 进一步提升了识别的准确率。但是针对大部分的图像处理而言, 基于深度卷积生成对抗网络的图形识别算法仍存在收敛速度慢、训练过程不够稳定的缺点。本文提出了融合加权 Canny 算子和 Prewitt 算子的深度卷积生成对抗网络算法^[4], 该加权算子对多个方向进行卷积, 从而初始化输入图片参数, 有效减少了噪声的干扰。

2 深度卷积神经网络算法

卷积神经网络又简称卷积网络 (CNN), 该模型主要用于处理网格状结构数据的特殊网络结构^[5-8]。该模型是受到猫的视觉皮层细胞研究的启发, 模仿其神经结构然后提出了感受野 (Receptive Field) 的概念。卷积神经网络将时序信息等单一变量的信息作为一维的数据格式, 而图片、位置等信息则被认为是二维的数据格式。卷积神经网络在推出后, 取得了巨大的成功。吕永标基于深度学习理论, 将图像去噪过程看成神经网络的拟合过程, 构造简洁高效的复合卷积神经网络, 提出基于复合卷积神经网络的图像去噪算法, 该算法由 2 个 2 层的卷积网络构成, 分别训练 3 层卷积网络中的部分初始卷积核, 缩短阶段网络的训练时间和增强算法的鲁棒性, 最后运用卷积网络

对新的噪声图像进行有效去噪。实验表明, 文中算法在峰值信噪比、结构相识度及均方根误差指数上与当前较好的图像去噪算法相当, 尤其当噪声加强时效果更佳且训练时间较短^[9]。王秀席等针对现有车型识别算法耗时长、特征提取复杂、识别率低等问题, 引入了基于深度学习的卷积神经网络方法。此方法具有鲁棒性好、泛化能力强、识别度高等优点, 因而被广泛使用于图像识别领域。在对公路中的 4 种主要车型 (大巴车、面包车、轿车、卡车) 的分类实验中, 改进后的卷积神经网络 LeNet-5 使车型训练、测试结果均达到了 98% 以上, 优于传统的 SIFT+SVM 算法, 其算法在减少检测时间和提高识别率等方面都有了显著提高, 在车型识别上具有明显优势^[10]。随着大数据时代来临以及 GPU 并行计算速度的飞速发展, 卷积神经网络本身不断优化 (ReLU 激活函数取代 Sigmoid 函数, Dropout 思想的提出), 其计算效率得到了大幅度提升。许赞杰等针对常用的激活函数在反向传播神经网络中具有收敛速度较慢、存在局部极小或梯度消失的问题, 将 Sigmoid 系和 ReLU 系激活函数进行了对比, 分别讨论了其性能, 详细分析了几类常用激活函数的优点及不足, 并通过研究 Arctan 函数在神经网络中应用的可能性, 结合 ReLU 函数, 提出了一种新型的激活函数 ArcReLU, 既能显著加快反向传播神经网络的训练速度, 又能有效降低训练误差并避免梯度消失的问题^[11]。

深度卷积神经网络包含以下几个核心部分:

(1) 局部感知。图像的空间联系中局部的像素联系比较紧密, 而距离较远的像素相关性则较弱。因此, 每个神经元其实只需对局部区域进行感知, 而不需要对全局图像进行感知。

(2) 权值共享。在上述的局部连接中, 每个神经元都有对应的参数, 再进行卷积完成特征提取。例如, 假设神经元都对应 50 个参数, 共 2 000 个神经元, 如果这 10 000 个神经元的 25 个参数都是相等的, 则参数量就变为 25 个。把这 25 个参数对应卷积操作, 完成了特征提取。在卷积神经网络中相同的卷积核的权值和偏置值是一样的。同一种卷积核按照固定对图像进行卷积操作, 卷积后得到的所有神经元都使用同一个卷积核区卷积图像, 都是共享连接参数。因此, 权值共享减少了卷积神经网络的参数数量。

(3)卷积。该步骤利用卷积核对图像进行特征提取。卷积过程本质上是一个去除无关信息,留下有用信息的过程。其核心就是卷积核的大小步长设计和数量的选取。个数越多提取的特征越多,但网络的复杂度也在增加,如果特征数量太少又不足以描述特征。在该步骤中的卷积核的大小影响网络结构的识别能力,步长决定了采取图像的大小和特征个数。

(4)池化。在卷积神经网络中,池化层一般在卷积层后,通过池化来降低卷积层输出的特征向量维数。池化过程最大程度地降低了图像的分辨率,同时降低了图像的处理维度,但又保留了图像的有效信息,降低了后面卷积层处理复杂度,大大降低了网络对图像旋转和平移的敏感性。一般采用的池化方法有两种:平均池化(Mean pooling)和最大池化(Max pooling)。平均池化是指对图像目标局部区域的平均值进行计算,将其作为池化后该区域的值;最大池化则是选取图像目标区域的最大值作为池化后的值。

3 基于深度卷积生成对抗网络的图像识别

3.1 图像预处理

为了能够提升图像训练效率,对图像进行预处理,流程如下:

(1)假设给出图像 I ,首先对其进行归一化处理,假设有像素为 128×128 的图像 I ,将其归一化为 $I_n \in [0, 1]^{128, 128, 3}$;

(2)利用 Canny 算子和 Prewitt 算子的加权综合,对图像进行进一步的卷积预处理,提取核心特征。

3.2 训练流程

采用 DCGAN 的 (G, D) 架构,生成器 G 是一个编码解码的 CNN 结构,判别器 D 是一个步长卷积方案,不断重复进行下采样来完成二分分类^[12]。在每次训练的地带中,我们都随机采样一批训练数据,对每个训练图像 I 我们运行生成器,接着用判别器来进行分类,然后计算损失和更新的参数。

为了减少训练过程,本文将训练过程分为 3 个阶段,每个阶段定义一个损失函数:

$$L_{\text{MSE}}(I) = \|M \odot G(I)\|, \quad (1)$$

$$L_D(I) = -[\ln D(I) + \ln(1 - D(G(I)))] \quad , \quad (2)$$

$$L_G(I) = L_{\text{MSE}}(I) - \alpha \ln D(G(I)) \quad , \quad (3)$$

第一个训练阶段通过损失函数 $L_{\text{MSE}}(I)$ 不断升级调整生成器的权重,第二阶段则使用 $L_D(I)$ 来调整生成器的权重,第三阶段过程类似,但是 α 是作为经验值,根据不同的场景而变化。

3.3 网络结构

在生成器 G 的网络结构中,我们采用传统的编码器-解码器结构,以及扩张的卷积来增加神经元的感受野。对于判别器 D ,我们将图像沿垂直方向分为左右两部分,分别为 I_l, I_r ,为了产生出预测结果 I_d ,判别器需要分别计算 $D_g(I_d), D_l(I_l), D_r(I_r)$,最后生成判别器的概率 p ,生成器和判别器的网络参数如表 1~5 所示。

表 1 卷积层 1

Tab.1 Convolution layer 1

Type	f	η	s	n
CONV	7	1	1	64

表 2 卷积层 2

Tab.2 Convolution layer 2

Type	f	η	s	n
CONV	3	1	2	128
CONV	3	1	1	256

表 3 卷积层 3

Tab.3 Convolution layer 3

Type	f	η	s	n
CONV	3	2	1	128
CONV	3	4	1	256
CONV	3	8	1	256

表 4 输出

Tab.4 Output

Type	f	η	s	n
DCONV	4	2	1	128
CONV	3	1	1	64
OUT	3	1	1	3

表 5 判别器参数表
Tab.5 Discriminator parameter table

Type	f	s	n
CONV	5	2	64
CONV	5	2	64
CONV	5	64	64
CONV	5	2	64
FC	—	—	512

4 结果与讨论

考虑到深度卷积网络的计算成本,本文主要利用 GPU 的并行计算能力,采用 5 台 Inter(R) I7-8700K,64 G 内存,NVIDIA RTX 24 G GPU 计算机,利用 TensorFlow^[13] 平台实现本文提出的算法并运行。

4.1 CIFAR-100 数据集

本文采用的 CIFAR-100 数据集包含大量的测试图片^[14]。与 CIFAR-10 不同的是,CIFAR-100 有 100 个分类,其中每个类都包含 500 个训练图像和 100 个测试图像,每个图像都带有两个标签。使用本文 2.3 节提出的模型实现生成器和

判别器,首先对数据集中的图像进行预处理,将图片进行归一化,再利用 Canny 算子和 Prewitt 算子的加权,最后作为训练模型的输入。在生成器的全连接层中加入 Dropout,在预处理的 3 个阶段,由于 CIFAR-100 的前景背景相对复杂多变,因此 T_1 设置为 20 000, T_2 为 3 000, T_3 为 10 000, α 为 0.006,实验总共运行了 25 h,进行了 50 个 epoch。在训练过程中,判别器的损失函数逐渐下降,而生成器的损失函数逐渐上升,最终判别器以微弱优势战胜了生成器。最终检测效果如图 1 所示。在与其他经典方法的对比中,其收敛的速度更快,检测率也更优,如表 6 所示。

表 6 CIFAR-100 图像集实验准确率
Tab.6 Experimental accuracy of CIFAR-100 image set

识别模型	准确率/%
Gabor+wavelet	79.42
Inception-ResNet	95.24
AlexNet	92.41
FAST-RCNN	91.22
DCGAN+ Canny+Prewitt	96.25
WGAN	94.32

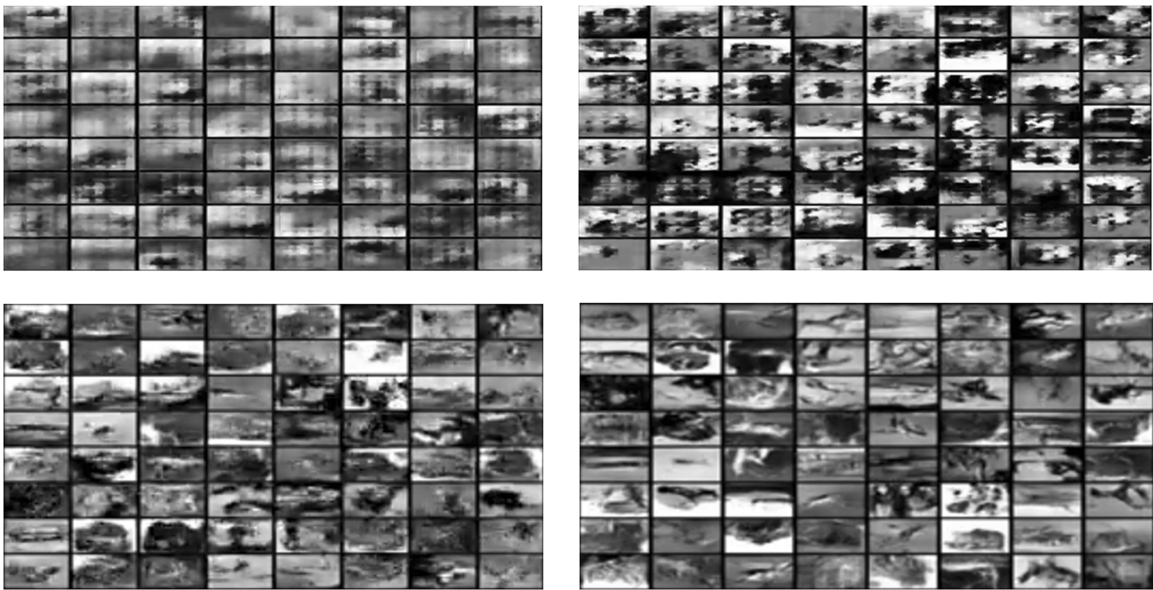


图 1 CIFAR-100 生成器迭代结果
Fig.1 Iterative results of CIFAR-100 generator

4.2 LFW 实验

LFW 是一个人脸识别专用的图像训练集^[15],有 2 845 张图片,每张图片中包含多个人脸,共有 5 171 个人脸作为测试集。测试集范围包括不同姿势、不同分辨率、旋转和遮挡等图片,基本能呈现人脸表情的各种状态,同时包括灰度图和彩色图。本文采用灰度图进行实验。由于整个数据集相对较小,同时前景相对简单,因此预处理时间消耗较小, T_1 设置为 1 000, T_2 为 2 000, T_3 为 10 000, α 为 0.005,实验共运行 4 h,进行了 50 个 epoch。在迭代过程中随着 epoch 的增大,生成器产生的图片越加清晰精准,质量不断提升,结果如图 2 所示。

从表 7 中可以看到,本文采用的 DCGAN+softmax+centermax 的模型优于传统的 DCGAN



图 2 LFW 生成器头像迭代结果

Fig.2 Iterative results of LFW generator

表 7 LFW 图像集实验识别准确率

Tab. 7 Experimental recognition accuracy of LFW image set

识别模型	准确率/%
AlexNet	87.52
Inception-ResNet	88.74
Gabor+wavelet	79.71
Fast-RCNN	83.31
DCGAN	82.56
DCGAN+Canny+Prewitt	89.54

模型,在相对较小的数据集中,具备更强的特征提取能力;本文提出的模型的检测率优于其他有监督和无监督算法,证明了本文方法的可行性。

5 结 论

提出了一个基于深度卷积对抗网络的模型,利用该模型对训练完成后的判别器进行特征提取并用于图像识别。利用 Canny 算子和 Prewitt 算子进行图像预处理,同时将训练划分为 3 个阶段设置不同的训练参数,提高了分类算法的内聚性。在 WILDERFACE 和 LFW 数据集上进行实验,结果表明相对于其他传统检测算法,本文提出的模型在 LFW 识别准确率达到 89.54%,CIFAR-100 上达到 96.25%,证明了本文提出的模型在图像检测领域的可行性。

参 考 文 献:

- [1] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, *et al.* Generative adversarial networks [J]. *Advances in Neural Information Processing Systems*, 2014, 3:2672-2680.
- [2] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks [J]. *Computer Science*, 2015, 12(36):1280-1287.
- [3] SALIMANS T, GOODFELLOW I, ZAREMBA W, *et al.* Improved techniques for training GANs [J]. *ArXiv*, 2016, 606.03498v1.
- [4] 丁鹏, 张叶, 刘让, 等. 结合形态学和 Canny 算法的红外弱小目标检测[J]. *液晶与显示*, 2016, 31(8):793-800.
DING P, ZHANG Y, LIU R, *et al.* Infrared weak target detection based on morphology and Canny algorithm [J]. *Chinese Journal of Liquid Crystal and Displays*, 2016, 31(8): 793-800. (in Chinese)
- [5] KIM Y. Convolutional neural networks for sentence classification [J]. *ArXiv*, 2014, 1408.5882.
- [6] 金志刚 李静昆. 基于对象性和多层线性模型的协同显著性检测 [J]. *光学 精密工程*, 2019, 27(8): 1845-1853.
JIN Z G, LI J K. Co-saliency detection based on objectness and multi-layer linear model [J]. *Optics and Precision Engineering*, 2019, 27(8): 1845-1853. (in Chinese)

- [7] 潘仙张 张石清 郭文平. 多模深度卷积神经网络应用于视频表情识别 [J]. 光学 精密工程, 2019, 27(4): 963-970.
PAN X Z, ZHANG S Q, GUO W P. Video-based facial expression recognition using multimodal deepconvolutional neural networks [J]. *Optics and Precision Engineering*, 2019, 27(4): 963-970. (in Chinese)
- [8] 张坤华, 谭志恒, 李斌. 结合粒子群优化和综合评价的脉冲耦合神经网络图像自动分割[J]. 光学 精密工程, 2018, 26(4): 962-970.
ZHANG K H, TAN Z H, LI B. Automated image segmentation based on pulse coupled neural network with particle swarm optimization and comprehensive evaluation [J]. *Optics and Precision Engineering*, 2018, 26(4): 962-970. (in Chinese)
- [9] 吕永标, 赵建伟, 曹飞龙. 基于复合卷积神经网络的图像去噪算法 [J]. 模式识别与人工智能, 2017, 30(2): 97-105.
LYU Y B, ZHAO J W, CAO F L. Image denoising algorithm based on compound convolution neural network [J]. *Pattern Recognition and Artificial Intelligence*, 2017, 30 (2): 97-105. (in Chinese)
- [10] 王秀席, 王茂宁, 张建伟, 等. 基于改进的卷积神经网络 lenet-5 的车型识别方法 [J]. 计算机应用研究, 35(7): 301-304.
WANG X X, WANG M N, ZHANG J W, *et al.* Vehicle recognition method based on improved convolutional neural network lenet-5 [J]. *Computer Application Research*, 35 (7): 301-304. (in Chinese)
- [11] 许赞杰, 徐菲菲. 基于 arcrelu 函数的神经网络激活函数优化研究 [J]. 数据采集与处理, 2019, 34(3): 517-529.
XU Y J, XU F F. Optimization of activation function in neural network based on arcrelu function [J]. *Data Acquisition and Processing*, 2019, 34(3): 517-529. (in Chinese)
- [12] ZHENG Z, ZHENG L, YANG Y. Unlabeled samples generated by GAN improve the person re-identification baseline *in vitro* [C]. IEEE International Conference on Computer Vision, USA: Honolulu, IEEE, 2017.
- [13] ABADI M, AGARWAL A, BARHAM P, *et al.* TensorFlow: large-scale machine learning on heterogeneous distributed systems [J]. *ArXiv*, 2016, 1603.04467.
- [14] GOODFELLOW I J, WARDE-FARLEY D, MIRZA M, *et al.* Maxout networks [J]. *Computer Science*, 2013: 1319-1327.
- [15] TAIGMAN Y, MING Y, RANZATO M, *et al.* DeepFace: closing the gap to human-level performance in face verification [C]. IEEE Conference on Computer Vision & Pattern Recognition, USA: Columbus, Ohio, IEEE, 2014.

作者简介:



刘恋秋(1982—),男,重庆市人,硕士,讲师,2007 年于重庆邮电大学计算机学院获硕士学位,主要从事计算机仿真方面的研究。E-mail:liulianqiu0627@hotmail.com