

基于条件生成对抗网络的手写数字识别

王爱丽, 薛冬, 吴海滨, 王敏慧

引用本文:

王爱丽, 薛冬, 吴海滨, 等. 基于条件生成对抗网络的手写数字识别[J]. 液晶与显示, 2020, 35(12): 1284–1290.

WANG Ai-li, XUE Dong, WU Hai-bin, et al. Handwritten digit recognition based on conditional generative adversarial network[J]. *Chinese Journal of Liquid Crystals and Displays*, 2020, 35(12): 1284–1290.

在线阅读 View online: <https://doi.org/10.37188/YJYXS20203512.1284>

您可能感兴趣的其他文章

Articles you may be interested in

基于深度卷积生成对抗网络的图像识别算法

Image recognition algorithms based on deep convolution generative adversarial network

液晶与显示. 2020, 35(4): 383–388 <https://doi.org/10.3788/YJYXS20203504.0383>

基于残差整流增强卷积神经网络的表情识别

Expression recognition based on residual rectifier enhanced convolution neural network

液晶与显示. 2020, 35(12): 1299–1308 <https://doi.org/10.37188/YJYXS20203512.1299>

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AttentionGAN和形态学重建的TFT阵列缺陷检测

TFT array defect detection based on AttentionGAN and morphological reconstruction

液晶与显示. 2020, 35(12): 1270–1277 <https://doi.org/10.37188/YJYXS20203512.1270>

多尺度特征融合与极限学习机结合的交通标志识别

Traffic sign recognition based on multi-scale feature fusion and extreme learning machine

液晶与显示. 2020, 35(6): 572–582 <https://doi.org/10.3788/YJYXS20203506.0572>

文章编号:1007-2780(2020)12-1284-07

基于条件生成对抗网络的手写数字识别

王爱丽, 薛 冬, 吴海滨*, 王敏慧

(哈尔滨理工大学 测控技术与通信工程学院, 黑龙江 哈尔滨 150080)

摘要:针对当训练样本不足时,传统深度学习算法在手写数字识别中会出现训练不稳定、识别精度较低等问题,提出了基于条件生成对抗网络的识别方法。首先,在条件生成对抗网络的基础上,利用生成器使用类别标签控制图像生成的优点,将生成器产生的图像样本作为训练数据,扩充数据集。同时,利用反卷积网络和卷积网络分别构成生成器和判别器的网络结构,去掉全连接层以提升模型稳定性。然后,引入条件批量归一化,利用它使用类别标签的优点,使网络学习更多的特征。最后,改进判别器为分类器,并提出新的损失函数,加快收敛速度,提高识别精度。实验结果表明,本文所提出的手写数字识别方法生成的图像质量更好,识别准确率更高,达到 99.43%,为生成对抗网络及其变体在图像识别领域中的应用提供了参考。

关键词:手写数字识别;条件生成对抗网络;条件批量归一化;图像生成

中图分类号:TP391 **文献标识码:**A **doi:**10.37188/YJYXS20203512.1284

Handwritten digit recognition based on conditional generative adversarial network

WANG Ai-li, XUE Dong, WU Hai-bin*, WANG Min-hui

(College of Measuring and Control Technology and Communication Engineering,
Harbin University of Science and Technology, Harbin 150080, China)

Abstract: In order to solve the problem of unstable training and low recognition accuracy of traditional deep learning algorithm in handwritten digit recognition when training samples are insufficient, a recognition method based on conditional generative adversarial network is proposed. Firstly, on the basis of the conditional generative adversarial network, the generator of category labels is used to control image generation. The image samples generated by the generator are used to expand the training data set. At the same time, the deconvolution network and the convolutional network are used to construct the network structure of generator and discriminator respectively, and the full connection layer is removed to improve the stability of the model. Then, the conditional batch normalization is introduced. Taking advantage of its use of category labels to make the network learn more features. Finally, the discriminator is improved to be a classifier, and a novel loss function is proposed to speed up the convergence rate and improve the recognition accuracy. The experimental results show that the handwrit-

收稿日期:2020-05-27;修订日期:2020-08-14.

基金项目:国家自然科学基金(No.61671190)

Supported by National Natural Science Foundation of China(No.61671190)

* 通信联系人, E-mail: woo@hrbust.edu.cn

ten digit recognition method proposed in this paper has better image quality and higher recognition accuracy of 99.43%. This paper provides a reference for the application of generative adversarial network and its variants in the field of image recognition.

Key words: handwritten digit recognition; conditional generative adversarial network; conditional batch normalization; image generation

1 引言

图像识别是计算机视觉领域中常见的任务之一,通过提取数字图像中有效的特征信息,来赋予图像各自的标签类别,从而完成识别过程。目前,图像识别已经广泛应用于医疗诊断^[1]、交通标志识别^[2]等领域中。

近年来,深度学习的快速发展使得其在图像识别领域取得了很多成果。相比于传统的人工识别方法,使用深度学习方法可以提取到图像更多的深层次特征,从而提高识别准确率。然而,深度学习方法需要大量的训练数据才能取得好的识别效果,所以对于数据量较小的数据集,识别效果并不理想,而生成对抗网络(GAN)的提出很大程度上缓解了这种问题^[3]。

GAN的核心思想来源于博弈论的纳什均衡^[4],它含有一个生成器和一个判别器,生成器在训练过程中将生成的假样本送入判别器来扩充训练集的数据量,这使得判别器学习到更多的图像特征,从而提高分类准确率。因此,基于GAN的这种优点,许多学者将其应用到图像分类、识别领域。Zhu等人^[5]将GAN应用到高光谱数据的分类中,提出了利用空间特征和光谱特征的1D-GAN和3D-GAN,显著提高了高光谱数据的分类精度。刘坤等人^[6]使用半监督GAN实现了X光图像的分类,充分利用有标签的数据提高分类效果。杨旺功等人^[7]利用深度卷积生成对抗网络(DCGAN)实现了花朵图像的分类,实验结果表明所提方法的分类准确率高且稳定性好。Kuang等人^[8]首次将GAN用于肺结节恶性肿瘤的无监督分类中,实验表明此方法使用少量标记样本就能取得较好的分类结果。

虽然GAN在图像处理领域具有较好的表现,但是GAN存在训练太过自由,生成的图像不可控等问题,因此Mirza等人^[9]提出了条件生成对抗网络(CGAN),在生成器和判别器中加入图

像的类别标签,使得生成的图片可以人为控制。因此,本文将CGAN和条件批处理归一化^[10-11](CBN)结合,提出CBN-CGAN网络。CBN利用类别标签对每一类数据进行批处理归一化,使网络学习到更多的特征,提高模型鲁棒性。

2 条件生成对抗网络理论

2.1 条件生成对抗网络

条件生成对抗网络是在GAN的基础上发展而来的,通过把附加信息 y 加入到生成器 G 和判别器 D 的输入中,使得GAN的无监督学习变为有监督学习。训练阶段,随机噪声 z 与条件变量 y 同时输入到 G 中,得到尽量服从真实数据分布 P_{data} 的生成数据 $G(z|y)$;然后将真实数据 x 、条件变量 y 和 G 生成的数据 $G(z|y)$ 同时输入到 D 中,最后输出一个标量来估计输入数据来自真实数据的概率。其目标函数为:

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x|y)] + E_{z \sim p_z(z)} [\log(1 - D(G(z|y)))]. \quad (1)$$

D 的目标是实现数据来源的二分类判别, G 的目标是最大化 D 判断输入数据错误的概率。因此,这两个相互对抗并迭代优化的过程使得 G 和 D 的性能不断提升。当最终 D 无法正确判别数据真假时,可以认为 G 已经学到了真实的数据分布,二者达到平衡状态。

2.2 条件批量归一化

深度学习算法的一个缺点是网络在训练阶段容易出现无法收敛的问题。因此,谷歌公司通过对输入数据进行批量归一化(BN)处理,提高了网络的训练速度,解决梯度消失的问题^[12]。但是在条件生成模型中存在一个问题,不同类别的训练数据放在一起做批处理归一化不妥当。因为在一个批量的数据中含有不同类别的数据,而每个类别的数据在计算后得到的均值和方差的数值也是不同的,因此在还原数据阶段,应该采用不同的标准化、平移和缩放处理的方式来还原每一类数据。

条件批量归一化^[10-11] (CBN) 处理数据时不使用整个批量的统计数据,而是在各类数据特征图内部做归一化,依赖类别标签还原出原始的每一类数据。CBN 对数据的批处理如下:

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

$$\sigma_B^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2, \quad (3)$$

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}, \quad (4)$$

$$y_i = \gamma(c) \hat{x}_i + \beta(c). \quad (5)$$

式(2)和(3)分别计算了数据的均值和方差;式(4)和(5)对数据进行标准化、平移和缩放的处理。

3 基于 CGAN 的图像生成及识别

3.1 网络模型结构

本文网络模型将 CGAN 与 CBN 相结合,称为 CBN-CGAN 网络,网络模型结构如图 1 所示,主要包括 3 方面的改进。

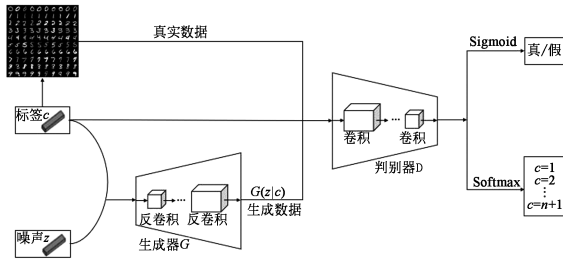


图 1 CBN-CGAN 网络结构图

Fig.1 Structure diagram of CBN-CGAN

(1) 将生成器和判别器的网络结构改为全卷积的网络结构,借助卷积神经网络(CNN)强大的提取特征的能力,更好地学习图像的特征。在判别器输出末端加入 Softmax 分类器,使模型适用于多分类任务。

(2) 在生成器和判别器的网络层中添加 CBN,充分利用类别标签来对每一类数据进行批量化处理,使得网络充分学习到特征图中每个类别的特征,提高生成图像的质量和识别精度。同时,提升模型稳定性,加快收敛速度,缓解梯度消失的问题。

(3) 提出新的目标损失函数,使模型最终的输出既包括对图像来源的真假判别,即将真实图像

判断为真,将生成器生成的图像判断为假,又包括对图像类别标签的多分类结果。

3.2 改进的生成器网络结构

本文网络模型的生成器是一个全卷积的网络结构,使用反卷积层代替池化层,反卷积核的大小为 4×4 ,并且去掉全连接层来增加模型稳定性。首先,输入端是两个反卷积层,它们分别将 100 维的噪声 z 和 n 维的类别标签(假设数据集有 n 个类别)进行维度转换,转换成 $(256, 4, 4)$ 的三维张量,然后将二者连接成 $(512, 4, 4)$ 的三维张量作为下一个反卷积层的输入。最后,3 个反卷积层将这个 3 维张量转换成 $(1, 32, 32)$ 的图像。本文在中间两个反卷积层后添加 CBN, CBN 对数据进行归一化处理得到均值和方差,利用类别标签控制偏置因子和缩放因子的取值,从而在映射原始数据时充分还原出每一类数据,使网络充分学习到各类数据的特征分布。

CBN 还可以解决初始化差和模式崩塌的问题,同时确保梯度传播到模型的深层。此外,为了不降低模型稳定性,在输入端和输出端的反卷积层后不添加 CBN。其次,本文在中间两个反卷积层后添加 ReLU 激活函数层,提高模型的学习速度,在最后一个反卷积层后添加 Tanh 激活函数层。CBN-CGAN 的生成器网络结构如图 2 所示。

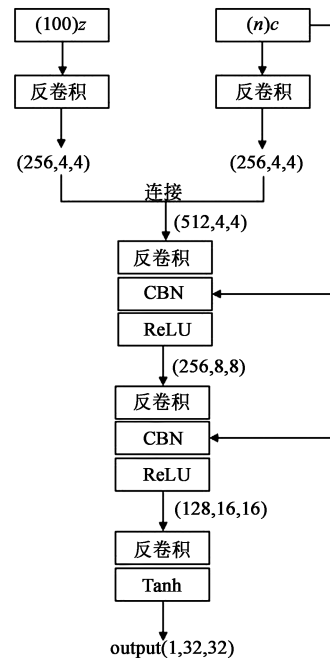


图 2 CBN-CGAN 生成器网络结构图

Fig.2 Structure diagram of CBN-CGAN generator

3.3 改进的判别器网络结构

本文判别器网络结构和生成器网络结构类似,也是一个全卷积的网络结构,含有 5 个卷积层,卷积核的大小为 4×4 ,不含池化层和全连接层。在输入端,训练数据和类别标签分别通过卷积层进行维度转换,转换成 $(64, 16, 16)$ 的三维张量,再将二者连接成 $(128, 16, 16)$ 的三维张量。接着依次经过两个卷积层输出 $(512, 4, 4)$ 的三维张量到达输出端。输出端分为两部分:一部分判断数据的真实来源,另一部分输出分类结果。

与生成器相同,本文在除了输入端和输出端的卷积层后添加 CBN,提升训练效果。本文在判别器中使用 LeakyReLU 激活函数,它是 ReLU 激活函数的改进版,在判别器中表现得更好。因此,本文在中间两个卷积层后添加 LeakyReLU 激活函数层。在输出端,本文在最后两个卷积层后分别使用 Sigmoid 分类器和 Softmax 分类器完成不同的任务,一是 Sigmoid 分类器输出数据的真假判别结果,二是 Softmax 分类器输出真实数据和生成器生成的数据的分类结果。

在分类时,假设数据集有 n 个类别。首先,每个由生成器生成的假数据通过网络向前传递,并通过获取概率预测向量的最大值来分配一个标签。因此,这些假数据可用于在网络中使用这些标签进行训练。此外,假数据不属于任何类型的真实数据。由于真实数据和假数据的不同,所以创建了一个新的类别标签($n+1$ 类)划分它们,每个假数据都被赋予这个新的类别标签。本文采用这种方法最后输出 $n+1$ 个类别的识别结果。CBN-CGAN 的判别器网络结构如图 3 所示。

3.4 目标损失函数

本文设计的网络模型的目标损失函数包括两部分:一部分是判断数据来源真假的 L_S ,另一部分是对真实数据和生成器生成的数据进行分类的 L_C 。该目标损失函数如下所示:

$$\min_G \max_D V(G, D) = L_S \pm L_C, \quad (6)$$

$$L_S = E_{x \sim p_{\text{data}}(x)} [\log D(x | c)] + E_{z \sim p_z(z)} [\log(1 - D(G(z | c)))] \quad (7)$$

$$L_C = E_{x \sim p_{\text{data}}(x)} [\log D(x | c, n)] + E_{z \sim p_z(z)} [\log D(G(z | c), n + 1)] \quad (8)$$

式中, L_S 将输入的真实数据判断为真,将生成器

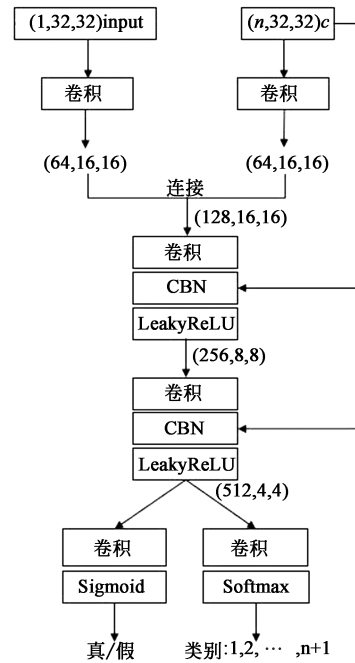


图 3 CBN-CGAN 判别器网络结构图

Fig.3 Structure diagram of CBN-CGAN discriminator

生成的数据判断为假。 L_C 分为两部分,一部分是真实数据对应的分类结果,它应该被检测为前 n 种类别中其唯一对应的类别,即检测为真实类别的概率为 1;另一部分是使生成器生成的假数据可以被分类为第 $n+1$ 类,即假数据被检测成真实类别的概率为 0,也即假数据对应的分类结果为第 $n+1$ 类的概率为 1。对于判别器,其最终目的是最大化 $L_S + L_C$;生成器的目的是最小化 $L_S - L_C$ 。

4 实验结果

本文实验在 Windows 操作系统下进行,基于开源深度学习框架 Pytorch,使用的编程语言为 Python,实验设备包括 Intel(R) Core(TM) i5-6500 CPU @ 3.2 GHz 处理器,16 GB 运行内存(RAM),NVIDIA GeForce GTX 1060 GPU。实验所用的数据集为 MNIST 数据集,它是机器学习领域最常见的手写字体数据集,包含 0~9 的 10 类手写数字图像,每类数字包含 60 000 个训练样本和 10 000 个测试样本,每幅图像是 28×28 像素的灰度图像。

4.1 实验设计

实验阶段,为了让生成器尽可能学习到所有

样本的数据分布,需要对样本进行归一化处理,对类别标签数据进行 one-hot 编码处理。网络输入的图像大小为 32×32 ,批处理大小设置为 32,由于数据集为灰度的手写数字图像,特征相对较少,因此训练迭代了 30 个 Epoch。在训练阶段,网络采用 Adam 优化器,学习率设置为 0.000 2,权重衰减设置为 0.000 5。

4.2 实验结果分析

为了评估本文所设计的 CBN-CGAN 网络模

型的性能,将本文网络模型与传统的深度学习网络做了对比。实验评价标准一方面是生成图像的质量,另一方面是识别的准确率。图 4 为本文方法与 CGAN 和深度卷积生成对抗网络^[13](DCGAN)在 Epoch 次数为 30 时对生成图像所做的对比。从对比图中可以直接看出,本文提出的 CBN-CGAN 网络比其他网络生成的图像质量更好,更清晰,例如数字“2”和数字“8”的轮廓更明显,更容易辨别。

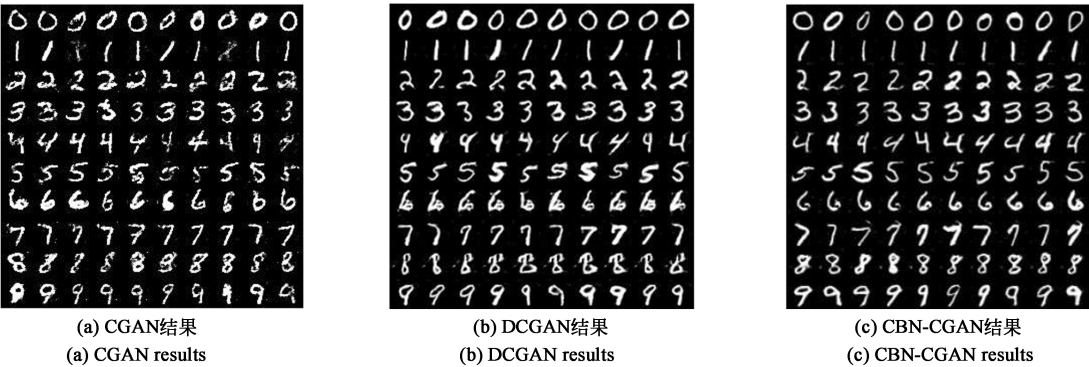


图 4 不同方法生成 MNIST 图像对比结果

Fig.4 Comparison results of MNIST images generated by different methods

表 1 MNIST 数据集识别准确率和识别时间

Tab.1 Recognition accuracy and recognition time of MNIST image set

分类模型	准确率/%	识别时间/s
Decision Tree	87.93	1 272.92
SVM	96.60	1 384.86
Random Forest	97.15	1 736.59
CNN	98.36	2 164.28
CGAN + Softmax	98.57	2 485.96
DCGAN + Softmax	98.86	2 794.27
CBN-CGAN + Softmax	99.43	2 983.74

表 1 为本文识别方法与其他方法的识别精度对比。其中决策树(Decision Tree)的最大深度设置为 100;支持向量机(SVM)中的核函数设置为径向基核函数(rbf),rbf 的系数默认为“auto”,惩罚参数设置为 100;随机森林(Random Forest)中树的数目设置为 200;卷积神经网络

(CNN)的网络结构与本文判别器网络结构相似,卷积核大小也为 4×4 ,网络中加入 ReLU 激活函数,BN 层,Softmax 分类器,数据输入之前做归一化处理,学习率设置为 0.001;在 CGAN 和 DCGAN 的判别器输出端分别添加 Softmax 分类器,使这两个网络可以完成分类任务,学习率都设为 0.000 2。

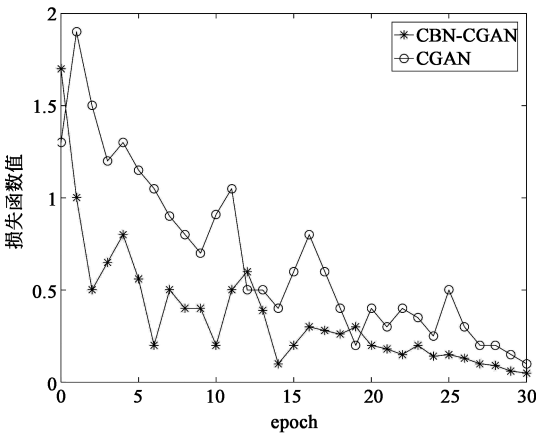


图 5 判别器损失函数曲线对比

Fig.5 Comparison of discriminator loss function curve

从表 1 可以看出,本文所提出的 CBN-CGAN 网络模型达到的识别准确率最高,为 99.43%,分别比 DCGAN、CGAN、CNN、随机森林、SVM 和决策树高出 0.57%,0.86%,1.07%,2.28%,2.83% 和 11.5% 的准确率。虽然本文模型在识别时所消耗的时间有一定程度的增加,但是却提高了准确率,因此证明了本文所提方法在手写体数字识别时可以更好地提取特征,有效提高了识别准确率。

图 5 为本文所提出的 CBN-CGAN 网络的判别器损失函数曲线与原始 CGAN 网络的判别器损失函数曲线的对比。从图中可以看出本文所提出的网络判别器的收敛速度更快,且随着迭代次

数的增加越来越稳定。

5 结 论

本文结合条件生成对抗网络与条件批处理归一化提出 CBN-CGAN、改进生成器和判别器的结构,提出新的损失函数。利用 CBN 融合类别标签的优势,使网络提取更多的图像特征,提高了模型的鲁棒性。在 MNIST 数据集上的实验结果表明,相比于其他方法,本文所提出的方法虽然消耗的时间较长,但是生成的图像质量更好,同时手写体数字识别的准确率达到 99.43%,验证了本文所提出的 CBN-CGAN 模型在手写体数字识别领域中的有效性。

参 考 文 献:

- [1] 郑欣,田博,李晶晶.基于 YOLO 模型的宫颈细胞簇团智能识别方法[J].液晶与显示,2018,33(11):965-971.
ZHENG X, TIAN B, LI J J. Intelligent recognition method of cervical cell cluster based on YOLO model [J]. *Chinese Journal of Liquid Crystals and Displays*, 2018, 33(11): 965-971. (in Chinese)
- [2] 张万征,胡志坤,李小龙.基于 LeNet-5 的卷积神经网络图像识别算法[J].液晶与显示,2020,35(5):486-490.
ZHANG W Z, HU Z K, LI X L. Convolutional neural image recognition algorithm based on LeNet-5 [J]. *Chinese Journal of Liquid Crystals and Displays*, 2020, 35(5): 486-490. (in Chinese)
- [3] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [C]//*Proceedings of the 27th International Conference on Neural Information Processing Systems*. Cambridge: MIT Press, 2014, 3: 2672-2680.
- [4] RATLIFF L J, BURDEN S A, SASTRY S S. Characterization and computation of local Nash equilibria in continuous games [C]//*Proceedings of the 51st Annual Allerton Conference on Communication, Control, and Computing*. Monticello, IL, USA: IEEE, 2013: 917-924.
- [5] ZHU L, CHEN Y S, GHAMISI P, et al. Generative adversarial networks for hyperspectral image classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(9): 5046-5063.
- [6] 刘坤,王典,荣梦学.基于半监督生成对抗网络 X 光图像分类算法[J].光学学报,2019,39(8):0810003.
LIU K, WANG D, RONG M X. X-ray image classification algorithm based on semi-supervised generative adversarial networks [J]. *Acta Optica Sinica*, 2019, 39(8): 0810003. (in Chinese)
- [7] 杨旺功,淮永建.基于深度卷积生成对抗网络的花朵图像增强与分类[J/OL].计算机科学,2020,47(6):176-179.
YANG W G, HUAI Y J. Flower image enhancement and classification based on deep convolution generative adversarial network [J/OL]. *Computer Science*, 2020, 47(6): 176-179. (in Chinese)
- [8] KUANG Y, LAN T, PENG X Q, et al. Unsupervised multi-discriminator generative adversarial network for Lung nodule malignancy classification [J]. *IEEE Access*, 2020, 8: 77725-77734.
- [9] MIRZA M, OSINDERO S. Conditional generative adversarial nets [J]. *arXiv*: 1411.1784, 2014.
- [10] ZHANG H, GOODFELLOW I, METAXAS D, et al. Self-attention generative adversarial networks [EB/OL]. [2019-06-14]. <https://arxiv.org/pdf/1805.08318.pdf>.
- [11] MIYATO T, KOYAMA M. CGANs with projection discriminator [EB/OL]. [2018-08-15]. <https://arxiv.org/pdf/1802.05637.pdf>.
- [12] LOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift [C]//*Proceedings of the 32nd International Conference on Machine Learning*. Lille, France: PMLR, 2015: 448-456.

- [13] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks [J]. *arXiv*: 1511.06434, 2015.

作者简介:



王爱丽(1979—),女,天津人,博士,副教授,2008 年于哈尔滨工业大学获得博士学位,主要从事深度学习、图像分类、图像超分辨重建方面的研究。
E-mail: aili925@hrbust.edu.cn



吴海滨(1977—),男,上海人,博士,教授,2008 年于哈尔滨理工大学获得博士学位,主要从事机器视觉、医学虚拟现实、光电检测方面的研究。
E-mail: woo@hrbust.edu.cn