基于条件深度卷积生成对抗网络的图像识别方法

唐贤伦 [1] 杜一铭 [1] 刘雨微 [2] 李佳歆 [2] 马艺玮 [2]

摘 要 生成对抗网络 (Generative adversarial networks, GAN) 是目前热门的生成式模型. 深度卷积生成对抗网络 (Deep convolutional GAN, DCGAN) 在传统生成对抗网络的基础上, 引入卷积神经网络 (Convolutional neural networks, CNN) 进行无监督训练; 条件生成对抗网络 (Conditional GAN, CGAN) 在 GAN 的基础上加上条件扩展为条件模型. 结合深度卷积生成对抗网络和条件生成对抗网络的优点, 建立条件深度卷积生成对抗网络模型 (Conditional-DCGAN, C-DCGAN), 利用卷积神经网络强大的特征提取能力, 在此基础上加以条件辅助生成样本, 将此结构再进行优化改进并用于图像识别中, 实验结果表明, 该方法能有效提高图像的识别准确率.

关键词 生成对抗网络, 卷积神经网络, 条件模型, 特征提取, 图像识别

引用格式 唐贤伦, 杜一铭, 刘雨微, 李佳歆, 马艺玮. 基于条件深度卷积生成对抗网络的图像识别方法. 自动化学报, 2018, **44**(5): 855−864

DOI 10.16383/j.aas.2018.c170470

Image Recognition With Conditional Deep Convolutional Generative Adversarial Networks

TANG Xian-Lun[1] DU Yi-Ming[1] LIU Yu-Wei[2] LI Jia-Xin[2] MA Yi-Wei[2]

Abstract Generative adversarial network (GAN) is a prevalent generative model. Deep convolutional generative adversarial network (DCGAN), based on traditional generative adversarial networks, introduces convolutional neural networks (CNN) into the training for unsupervised learning to improve the effect of generative networks. Conditional generative adversarial network (CGAN) is a conditional model which adds condition extension into GAN. The generative model of conditional-DCGAN (C-DCGAN) is a combination of DCGAN and CGAN, which integrates the feature extraction of convolutional networks and condition auxiliary generative sample for image recognition. The result of simulation experiments shows that this model can improve the accuracy of image recognition.

Key words Generative adversarial network (GAN), convolutional neural networks (CNN), conditional models, feature extraction, image recognition

Citation Tang Xian-Lun, Du Yi-Ming, Liu Yu-Wei, Li Jia-Xin, Ma Yi-Wei. Image recognition with conditional deep convolutional generative adversarial networks. *Acta Automatica Sinica*, 2018, **44**(5): 855−864

生成对抗网络 (Generative adversarial network, GAN) 是 Goodfellow 等在 2014 年提出的一种生成模型[1]. 不同于传统生成模型, 其在网络结构上除了生成网络外, 还包含一个判别网络. 生成网络与判别网络之间是一种对抗的关系. 对抗的思想源自博弈论 (Game theory), 博弈双方在平等的对局中各自利用对方的策略变换自己的对抗策略, 以此达到获胜目的[2]. 引申到生成对抗网络中, 即生成器和判别器为博弈双方, 生成器拟合数据的产生过程生成模型样本, 优化目标是达到纳什均衡[3], 使生成器估测到数据样本的分布. GAN 目前在图像和视觉领域得到了广泛的研究和应用, 已经可以生成数字和人脸等物体对象, 构成各种逼真的室内外场景, 从分割图像恢复原图像, 给黑白图像上色, 从物体轮廓恢复物体图像, 从低分辨率图像生成高分辨率图像等[4]. 此外, GAN 已经开始被应用到语音和语言处理[5−6]、电脑病毒监测[7]、棋类比赛程序[8]等问题的研究中. 然而生成对抗网络在图像识别领域的应用却不多, 虽然图像识别现有方法如卷积神经网络 (Convolutional neural networks, CNN) 等已经有很高的识别率, 但这些方法依赖大量数据并且收敛速度较慢. 本文结合条件生成对抗网络 (Conditional GAN, CGAN) 与深度卷积生成对抗

收稿日期 2017-08-29 录用日期 2017-12-14
Manuscript received August 29, 2017; accepted December 14, 2017
国家自然科学基金 (61673079, 61703068), 重庆市基础科学与前沿技术研究项目 (cstc2016jcyjA1919) 资助
Supported by National Natural Science Foundation of China (61673079, 61703068) and Chongqing Research Program of Basic Research and Frontier Technology (cstc2016jcyjA1919)
本文责任编委 李力
Recommended by Associate Editor LI Li
1. 重庆邮电大学计算机科学与技术学院 重庆 400065 2. 重庆邮电大学自动化学院 重庆 400065
1. College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065 2. College of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065

网络 (Deep convolutional GAN, DCGAN) 建立条件深度卷积生成对抗网络模型 (Conditional-DCGAN, C-DCGAN), 利用该网络模型的判别器提取特征用于图像分类. 实验结果表明, 该方法能有效提高图像识别的正确率. 本文第 1 节介绍 GAN 的原理; 第 2 节和第 3 节分别介绍 CGAN 和 DCGAN; 第 4 节结合 CGAN 和 DCGAN 建立 C-DCGAN 模型用于图像分类; 第 5 节对实验结果进行分析; 第 6 节对本文进行总结和展望.

1 生成对抗网络原理

生成对抗网络 (GAN) 由两个模型构成, 生成模型 G 和判别模型 D, 随机噪声 z 通过 G 生成尽量服从真实数据分布 $P_{\rm data}$ 的样本 $G(z)$, 判别模型 D 可以判断出输入样本是真实数据 x 还是生成数据 $G(z)$. G 和 D 都可以是非线性的映射函数, 比如多层感知器. GAN 的流程如图 1 所示.

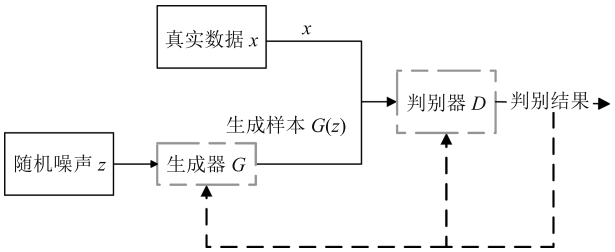

图 1 GAN 流程图

Fig. 1 GAN flow chart

GAN 核心原理的算法描述如下:

首先, 在生成器给定的情况下, 优化判别器. 判别器为一个二分类模型, 训练判别器是实现最小化交叉熵的过程. $\mathrm{E}(\cdot)$ 为期望值的计算, x 采样于真实数据分布 $P_{\rm data}(x)$, z 采样于先验分布 $P_z(z)$.

生成器为了学习数据 x 的分布, 由先验噪声分布 $P_z(z)$ 构建了一个映射空间 $G(z;\theta_g)$, 对应的判别器映射函数为 $D(x;\theta_d)$, 输出一个标量表示 x 为真实数据的概率.

$$\min_G \max_D V(D,G) = \mathrm{E}_{x\sim P_{\rm data(x)}}[\log D(x)] + \mathrm{E}_{z\sim P_{z(z)}}[\log(1-D(G(z)))] \tag{1}$$

其中, $\mathrm{E}_{x\sim P_{\rm data}(x)}[\log D(x)]$ 中, x 表示真实样本, $D(x)$ 表示 x 通过判别网络判断其为真实样本的概率; $\mathrm{E}_{z\sim P_z(z)}[\log(1-D(G(z)))]$ 中, z 表示输入生成样本的噪声, $G(z)$ 表示生成网络由噪声 z 生成的样本, $D(G(z))$ 表示生成样本通过判别网络后, 判断其为真实样本的概率. 生成网络的目的是让生成样本越接近真实样本越好, 即 $D(G(z))$ 越接近 1 越好, 这时 $V(D,G)$ 会变小; 而判别网络的目的是让 $D(x)$ 接近 1, 而 $D(G(z))$ 接近 0, 此时 $V(D,G)$ 会增大.

相比其他生成模型, 从实际结果看, GAN 能产生更好的生成样本.

但原始的 GAN 存在很多问题. 训练 GAN 需要达到纳什均衡, 训练 GAN 模型是不稳定的. 另外, 它也很难去学习生成离散的数据, 为了取得 “胜利” 生成器会选择容易生成的样本.

2 条件生成对抗网络

条件生成对抗网络 (CGAN) 是在 GAN 的基础上加上了条件扩展为条件模型, 如果生成器和判别器都适用于某些额外的条件 c, 例如类标签, 那么可以通过将 c 附加到输入层中输入到生成器和判别器中进行调节, 可以指导数据生成过程[9].

在生成器中, 输入噪声的同时输入相应条件 c, 而真实数据 x 和条件 c 作为判别器的输入. 其目标函数 $V(D,G)$ 如式 (2) 所示:

$$\min_G \max_D V(D,G) = \mathrm{E}_{x\sim P_{\rm data(x)}}[\log D(x|c)] + \mathrm{E}_{z\sim P_{z(z)}}[\log(1-D(G(z|c)))] \tag{2}$$

由式 (2) 可知, CGAN 对于目标函数 $V(D,G)$ 的优化过程与 GAN 相似: $\mathrm{E}_{x\sim p_{\rm data}(x)}[\log D(x|y)]$ 表示将数据 x 与条件 c 输入判别器 D 得到是否为真实数据的概率; $\mathrm{E}_{z\sim P_z(z)}[\log(1-D(G(z|c)))]$ 表示随机噪声结合条件 c 输入生成器产生的生成样本, 然后通过判别器判断其为真实数据的概率. 图 2 是一个简单的 CGAN 结构.

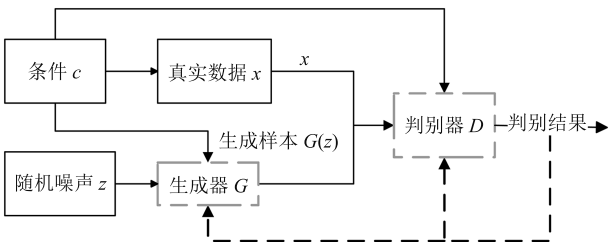

图 2 CGAN 流程图

Fig. 2 CGAN flow chart

若条件 c 为类别标签 y, 则可以认为 CGAN 是将无监督的 GAN 模型变为有监督模型的改进.

3 深度卷积生成对抗网络

深度卷积生成对抗网络 (DCGAN) 首次将卷积网络引入 GAN 的结构, 利用卷积层强大的特征提取能力[10] 来提高 GAN 的效果.

DCGAN 相比于传统 GAN 有以下特点[11−13]:

1) 在判别器模型中使用带步幅 (Strided convolutions) 的卷积代替池化层 (Pooling); 在生成器模

型中使用 Four fractionally-strided convolution 完成从随机噪声到图片的生成过程.

2) 在网络结构中, 除了生成器模型的输出层及其对应的判别器模型的输入层, 其他层上都使用了批量归一化 (Batch normalization), 加入 Batch normalization 层这一操作解决了初始化差的问题, 同时保住梯度传播到每一层, 也能够防止生成器把所有的样本都收敛到同一个点.

3) 去除全连接层, 直接使用卷积层连接生成器和判别器的输入层和输出层; 需要注意, 取消全连接层增加了模型的稳定性, 但却使得收敛速度变慢.

4) 生成器的输出层使用 Tanh (双切正切函数) 激活函数, 其余层使用 ReLU (Rectified linear unit); 判别器的所有层使用 Leaky ReLU (Leaky rectified linear unit).

4 基于条件深度卷积生成对抗网络的图像识别

本文综合 CGAN 和 DCGAN 的特点, 使用其相结合的模型 — 有条件的深度卷积生成对抗网络 (C-DCGAN) 模型. 将条件加入深度卷积生成对抗网络中的生成器, 利用卷积网络提取特征的能力加上条件辅以训练. 之后将训练好的 C-DCGAN 中的判别器部分抽取出来, 添加 Softmax 后形成用于图像识别的新网络结构.

4.1 C-DCGAN 模型结构

C-DCGAN 的生成器模型如图 3 所示 (此处以 MNIST 数据集为例, 数据为 28 像素 × 28 像素的手写数字图像): 100 维的噪声与 10 维的标签数据连接 (Concat) 成 110 维的数据作为输入, 通过两个全连接层后进行维度转换 (Reshape) 成 (7, 7, 128) 的三维张量, 之后通过一个卷积核为 3 像素 × 3 像素、步幅为 2 的转置卷积层, 输出一个 (14, 14, 128) 的三维张量, 经过三个卷积核为 3 像素 × 3 像素步幅为 1 的转置卷积层, 输出再次经过一个步幅为 2 以及三个步幅为 1 的转置卷积层后输出 (28, 28, 1) 张量, 即为一个生成图像样本. 需注意的是, 与原始 CGAN 生成器模型只在输入时连接标签数据不同, 为了增强标签数据在训练中的引导作用, 模型中每一层的输入数据都要连接标签数据, 当输入的是标量数据时, 需要连接 (Concat) 标签数据 y 的 10 维标量, 例如: 输入层 100 维噪声连接 10 维标签数据后得到 110 维的数据作为输入层的输入; 当输入的是三维张量时, 需要将标签数据 y 转换为三维张量之后通过第三个通道进行连接, 如图 3 Deconv1 层的输入, 先将标签数据转换成形如 (1, 1, 10) 的三维张量, 再将此张量乘上 (7, 7, 10) 的全 1 三维张量, 标签数据即转换为 (7, 7, 10) 的三维张量, 此时就可将上一层的输出 (7, 7, 128) 与转换好的标签数据 (7, 7, 10) 通过第三维连接起来, 即得到 (7, 7, 138) 的张量.

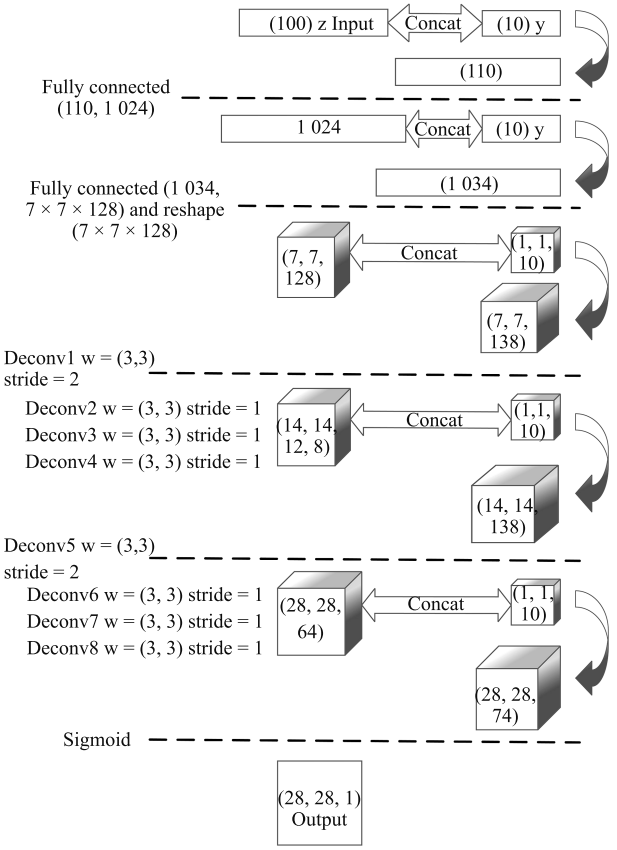

图 3　C-DCGAN 生成器的结构

Fig. 3　The structure of C-DCGAN generator

C-DCGAN 的判别器结构 (如图 4 所示) 与生成器结构正好相反. 输入一张 (28, 28, 1) 的样本数据经过一个卷积核大小为 3 像素 × 3 像素步幅为 1 的卷积层, 输出 (28, 28, 64) 的三维标量后经过两层步幅为 1 和一个步幅为 2 的卷积层后变为 (14, 14, 128) 的三维张量, 类似再经过 3 个步幅为 1, 一个步幅为 2 的卷积层后转换成 (7, 7, 128), 将三维张量展开, 通过两个全连接层后, 输出 1 维的结果. 与原始 CGAN 的判别器有所不同, 判别器不需要在输入时连接标签数据, 同样也不需要如上述 C-DCGAN 生成器在每一层连接标签数据. 这样修改原始模型的原因是由于本文的目的是要将训练好的 C-DCGAN 的判别器提取出来用于分类, 所以在预训练时, 判别器的结构中要避免出现标签数据的影响, 另外, 在后面训练分类器时也方便整体提取; 在生成器每一层输入条件信息, 相比只在输入层加入条件信息更能引导生成器的输出结果.

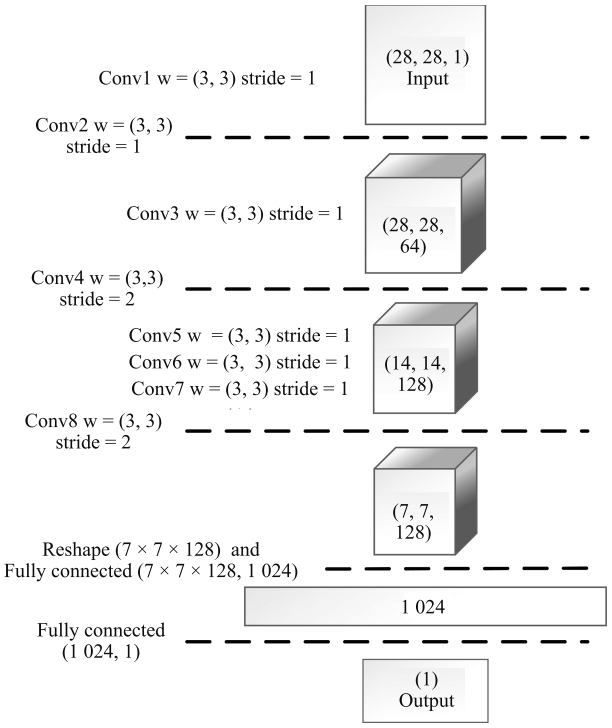

图 4 C-DCGAN 判别器的结构

Fig. 4 The structure of C-DCGAN discriminator

4.2 C-DCGAN 模型训练

生成器的目标是要生成器的输出通过判别器后结果接近 1 (生成样本接近真实样本); 而判别器的目的, 一方面要让真实样本通过判别器后结果接近 1, 另一方面让生成器生成的样本通过判别器的结果接近 0.

$$\min_G \max_D V(D,G) = \mathrm{E}_{x\sim P_{\mathrm{data(x)}}}[\log D(x)] + \mathrm{E}_{z\sim P_{z(z)}}[\log(1-D(G(z|c)))] \tag{3}$$

生成器的损失函数定义为生成器的输出与 “1” 做交叉熵 (Cross entropy). 判别器的损失函数由两部分组成: 1) 真实样本通过判别器的输出与 “1” 做交叉熵; 2) 生成器生成的样本通过判别器的输出与 “0” 做交叉熵. 判别器的损失函数是两个部分之和.

获得判别器和生成器的损失函数后, 选择 Adam 优化器优化损失函数.

实际训练过程中, 判别器会很容易在与生成器的对抗训练中取得胜利, 导致生成器出现梯度消失 (Vanishing gradient)[14]. 因此在训练中, 更新一次判别器需要更新 k ($k > 1$) 次生成器来使判别器在训练过程中不能快速达到 (近似) 最优, 以此保持生成器与判别器的对抗平衡. k 值的选择也要根据不同规模的数据集选定, 如果 k 值偏小, 会使判别器达到 (近似) 最优, 生成器就会出现梯度消失, 损失函数降不下去的情况; 如果 k 值偏大, 会导致生成器的梯度不准, 来回震荡.

4.3 基于 C-DCGAN 的图像识别

本文将训练好的 C-DCGAN 应用到图像识别中. 如图 5 所示, 将训练好的 C-DCGAN 判别器部分 (去除最后一层) 提取出来, 并在新的结构中进行参数微调, 与判别器结构不同, 最后一个全连接层输出维数为 n (n 等于该数据集类别数, 如 MNIST 数据集 $n = 10$). 输出结果通过 Softmax 分类器. 与标签 y 做交叉熵得到的损失函数, 同样使用 Adam 来优化. 另外, 在训练此结构时, 将 Dropout 的值设为 0.5, 防止全连接层出现过拟合.

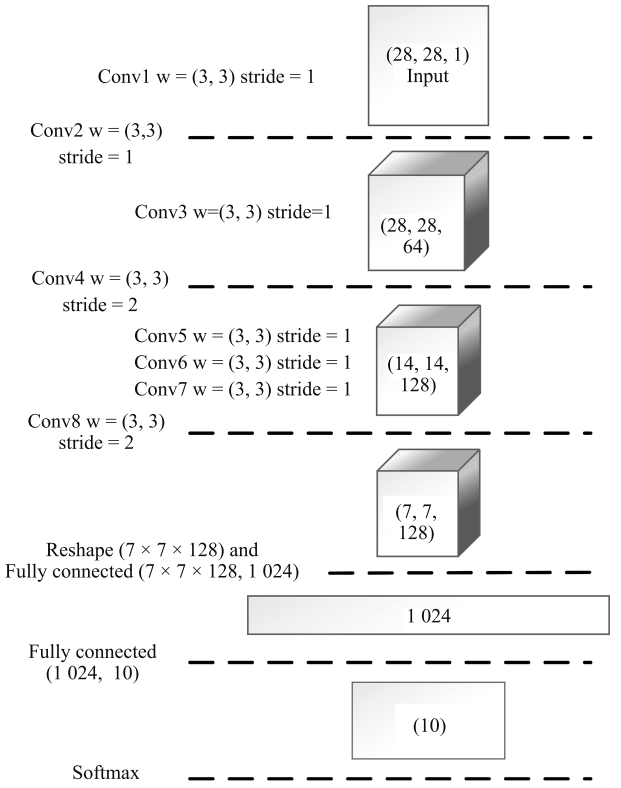

图 5 C-DCGAN 在 MNIST 上分类的结构

Fig. 5 The structure of C-DCGAN′s classification on MNIST

与一般的有监督学习用于图像分类的方式相比, 本文方法所用分类模型中的所有特征提取层会作为 C-DCGAN 的判别器在数据集中进行预训练, 其优势在于除数据集中的真实样本外, C-DCGAN 生成器生成的样本同样会作为数据输入到判别器中, 起到数据增强的作用, 使得判别器能够通过训练获得更多的特征.

5 实验结果与分析

本文在 MNIST 和 CIFAR-10 数据集上进行实验分析. 实验环境为 Intel (R) Core (TM) i5-6500

CPU @ 3.20 GHz 处理器, 8 GB 运行内存 (RAM), NVIDIA GeForce GTX 1050 Ti GPU, TensorFlow 平台.

5.1 MNIST 实验

MNIST 数据集包含 0～9 的 10 类手写数字灰度图像, 图像大小为 28 像素 × 28 像素, 整个数据集有 60 000 个训练样本, 10 000 个测试样本. 在使用数据时, 需要对图像样本归一化处理, 标签数据进行独热编码 (One-hot encoding).

使用图 3 和图 4 的结构作为对抗网络的生成器和判别器对 MNIST 数据集预训练, 为了保持对抗平衡, 判别器与生成器的更新次数为 1 : 3. C-DCGAN 训练完成后, 提取判别器各层的权值使用图 5 的结构进行图像分类. 根据生成器和判别器的特点且经过实验测试, 各层激活函数使用如下策略效果较好.

1) 生成器除了最后一层使用 Sigmoid 函数外, 其余每一层都使用 ReLU 作为激活函数;

2) 判别器使用 Leaky ReLU 作为激活函数.

在生成器输出层使用 Sigmoid 的原因是因为 Sigmoid 在特征相差明显时的效果很好, 在循环过程中会不断地增强特征效果. 而 ReLU 函数虽然收敛速度快, 但其最大的缺点是当输入小于零时的梯度为 0, 这样就导致负值的梯度被设置为 0, 此神经元会处于失活状态, 不会对任何数据有所反应, 尤其是当学习率很大时, 会出现神经元大面积坏死. 生成器的学习率通常较小, 所以在其余层可以使用 ReLU 加快训练速度. 判别器的所有层选用收敛速度快又不易使神经元坏死的 Leaky ReLU.

为了防止训练时出现过拟合的情况, 在模型训练时使用如下策略: 在生成器的两个全连接层加入 Dropout[15], Dropout 率设置为 0.5; 判别器的第一个全连接层后加入 Dropout, 设为 0.9; 另外, 对判别器的所有步幅为 1 的卷积层以及所有全连接层的参数进行 L2 正则化.

为了防止训练时梯度消失的问题, 生成器和判别器每一层都使用 Batch normalization 对隐含层的输入进行批量归一化[16−17] 处理.

5.1.1 MNIST 生成样本

实验中设置优化器 Adam 的学习率为 0.0002, 动量为 0.5. 每个批次 64 个样本. 训练迭代次数与损失函数的变化如图 6～9 所示.

图 6 和图 7 分别表示判别器判别真实样本 (d_loss_real) 和生成样本 (d_loss_fake) 的损失函数, 随训练次数增加而变化的情况. 根据第 4.2 节的分析 d_loss 的值为 d_loss_real 和 d_loss_fake 两值求和得到. 图 8 表示 C-DCGAN 在 MNIST 数据集上判别器的损失函数 (d_loss) 随着训练次数增加而变化的情况. 图 9 表示 C-DCGAN 在 MNIST 数据集上生成器的损失函数 (g_loss) 变化情况.

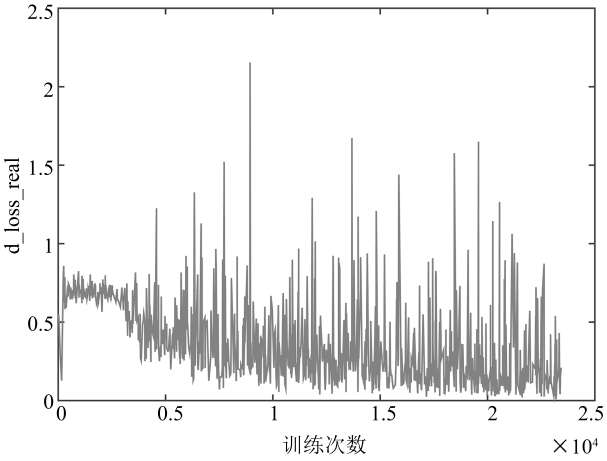

图 6　MNIST 上 d_loss_real 变化趋势

Fig. 6　Trends of d_loss_real on MNIST

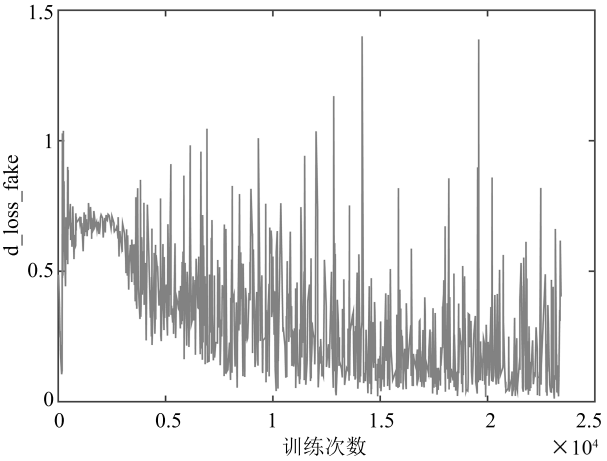

图 7　MNIST 上 d_loss_fake 变化趋势

Fig. 7　Trends of d_loss_fake on MNIST

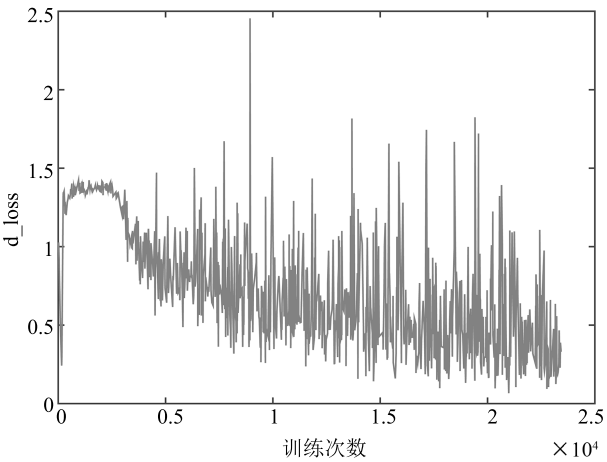

图 8　MNIST 上 d_loss 变化趋势

Fig. 8　Trends of d_loss on MNIST

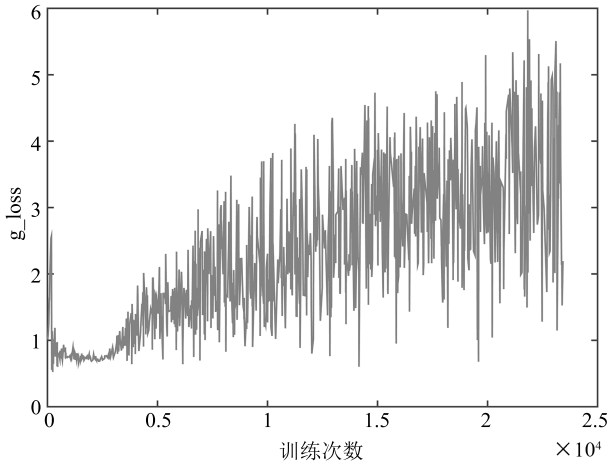

图 9 MNIST 上 g_loss 变化趋势

Fig. 9 Trends of g_loss on MNIST

从图 8 和图 9 可以看, 出生成器和判别器在训练初期较为平滑, 随着训练次数的增加, 模型逐渐稳固, 两个网络结构相互对抗, 呈现出图中大幅震荡状态. 尽管在训练过程中, 判别器与生成器的更新次数为 1 : 3, 但在总体趋势上, 判别器的损失函数处于逐渐下降, 生成器的损失函数处于逐步上升. 实验仿真表明, 在对抗过程中, 判别器能够以微弱优势胜过生成器.

图 10 为数据集前 64 个生成样本随着 Epoch 次数的增加, 生成样本的变化, 可以看出类似 0 和 6, 3 和 8, 这些形似的数字会在训练中出现变化.

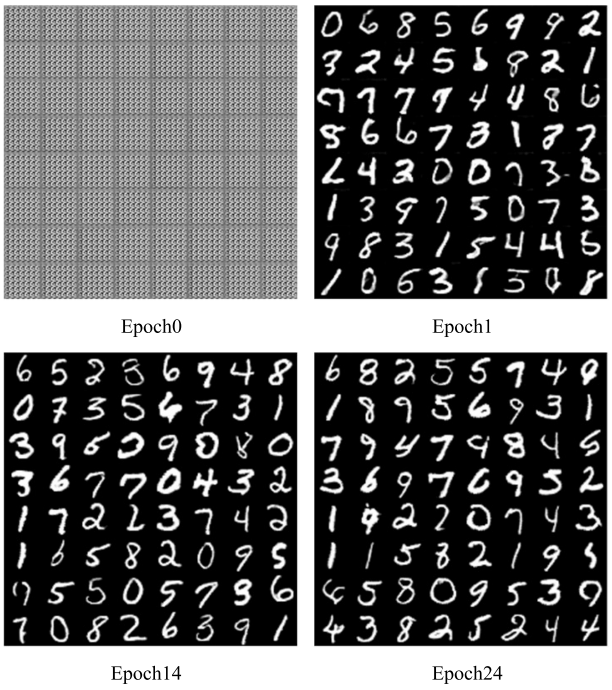

图 10 MNIST 生成样本

Fig. 10 The samples generated by MNIST

5.1.2 MNIST 分类结果

分类器同样选取学习率为 0.0002 的 Adam 优化器, 为了验证 C-DCGAN 算法在图像分类上的优势, 本文单独训练了一个 CNN 模型作为对比, 在结构上该 CNN 模型与本文 C-DCGAN 判别器完全相同, 且在训练前对数据进行了归一化处理. 图 11 为两个模型损失函数随迭代次数变换的对比.

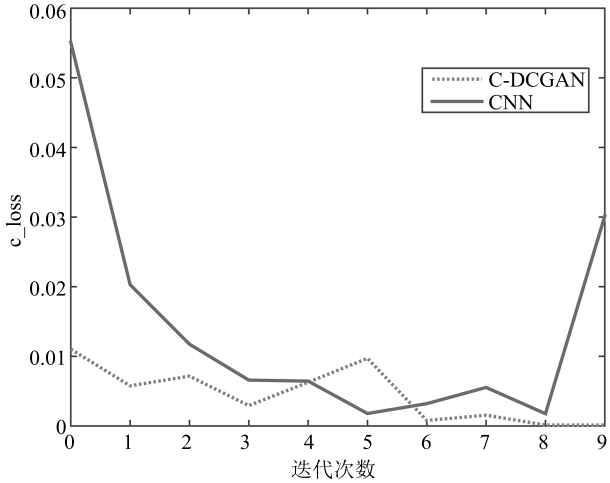

图 11 MNIST 上 c_loss 变化趋势

Fig. 11 Trends of c_loss on MNIST

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 10 个 Epoch. 由于本文分类模型中的特征提取层在 C-DCGAN 模型中作为判别器完成了对抗训练, 有效提取到了数据集的特征, 因此在分类训练开始时就取得了很大优势. 在相同的迭代次数下, 本文方法较传统 CNN 方法能更快地收敛.

表 1 为本文方法与其他方法[18−20] 识别结果的对比. 本文对比数据仅列举与本文结构相似的非大型卷积神经网络网络, 表 1 中 Convolutional net LeNet-5, [huge distortions] 和 Convolutional net LeNet-5, [distortions] 两种方法是在 LeNet-5 模型中通过对数据进行扭曲处理实现数据增强来提升识别效果, 由表 1 可知, 两种方法与无扭曲的方法 Convolutional net LeNet-5, [no distortions] 相比, 识别准确率有所提升, 但本文方法识别率优于上述两种方法的结果, 证明在 C-DCGAN 模型中以对抗训练方式提取特征达到数据增强的效果优于对原始数据预处理的数据增强效果.

表 1 中 CNN 为上文实验中用作对比的模型, 与 C-DCGAN 判别器模型的结构相同, 且在输入时对数据进行了归一化处理, 而本文方法没有对图像进行去斜处理和归一化处理, 但识别率相对上述传统方法有所提高, 证明该方法的可行性.

表 1　MNIST 上各方法准确率对比

Table 1　The recognition accuracy comparison on MNIST

识别方法	预训练	准确率 (%)
linear classifier (1-layer NN)	去斜	91.60
K-nearest-neighbors, Euclidean (L2)	–	95.00
40 PCA + quadratic classifier	–	96.70
SVM, Gaussian Kernel	–	98.60
Trainable feature extractor + SVMs [no distortions]	–	99.17
Convolutional net LeNet-5, [no distortions]	–	99.05
Convolutional net LeNet-5, [huge,distortions]	huge distortions	99.15
Convolutional net LeNet-5, [distortions]	distortions	99.20
CNN	归一化	98.40
C-DCGAN + Softmax	–	**99.45**

5.2　CIFAR-10 实验

CIFAR-10 数据集包含 10 类彩色图像, 图像大小为 32 像素 × 32 像素, 数据集有 50 000 个训练样本, 10 000 测试样本. 与 MNIST 数据集一样, 在使用数据时需要对图像样本归一化处理, 标签数据进行独热编码. 与 MNIST 数据集的实验相似, 使用类似图 3～5 的结构分别作为生成器、判别器和分类器, 生成器的输出以及判别器和分类器的输入为 (32, 32, 3) 的样本, 即长 32 宽 32 通道为 3 的彩色图. C-DCGAN 对抗训练时, 为了保持对抗平衡, 判别器与生成器的更新次数为 1 : 2.

5.2.1　CIFAR-10 生成样本

与 MNIST 实验相同, 优化器 Adam 的学习率设为 0.0002, 动量设为 0.5. 每个批次 64 个样本.

图 12 和图 13 真实数据 (d_loss_real) 和生成数据 (d_loss_fake) 通过判别器的损失函数随着训练次数增加的变化. 两图总体都呈现出下降趋势, 训练后期都出现大幅度震荡, 这种现象是对抗网络模型稳定后其与判别器对抗效果的表现.

图 14 是判别器损失函数 (d_loss) 的变化; 图 15 是生成器损失函数 (g_loss) 的变化.

从图 14 和图 15 可以看出, 生成器和判别器训练前期都相对平滑, 后期震荡明显, 这一现象说明随着训练次数的增加两个网络不断变得成熟, 由于两者之间对抗的关系, 因此会出现图中此消彼长的震荡.

图 16 为 CIFAR-10 数据集前 64 个样本在 4 个阶段 (第 1、6、16、25 Epoch) 生成样本的情况.

5.2.2　CIFAR-10 分类结果

分类器同样采用学习率 0.0002 的 Adam 优化器, 同样与归一化处理数据后采用传统 CNN 模型的

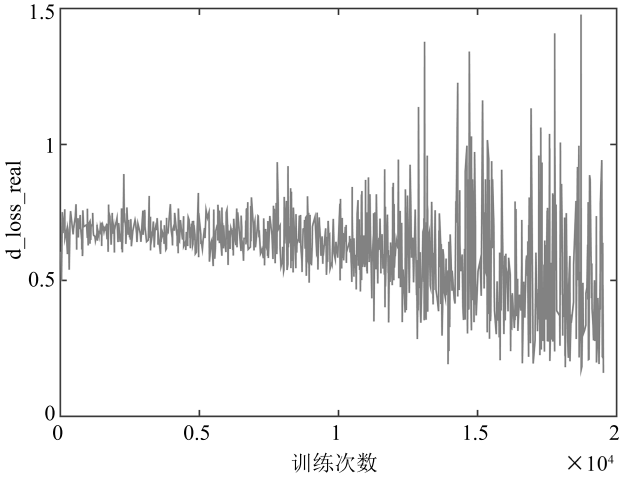

图 12　CIFAR-10 上 d_loss_real 变化趋势

Fig. 12　Trends of d_loss_real on CIFAR-10

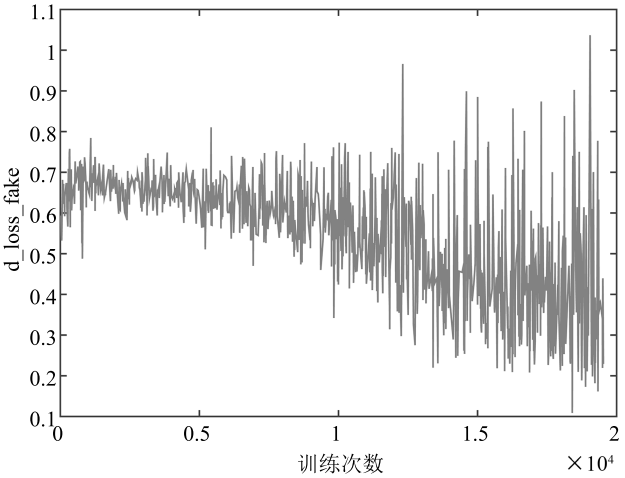

图 13　CIFAR-10 上 d_loss_fake 变化趋势

Fig. 13　Trends of d_loss_fake on CIFAR-10

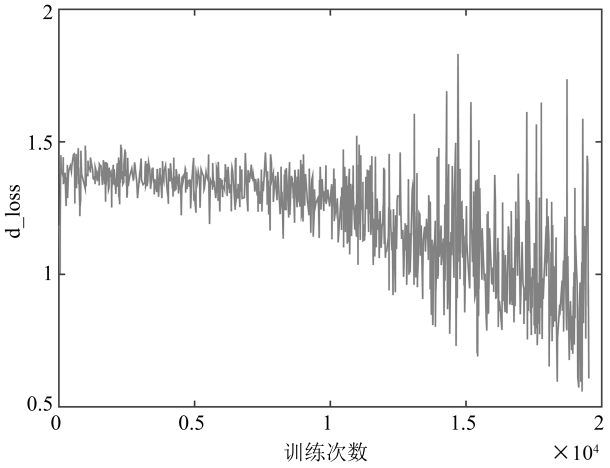

图 14 CIFAR-10 上 d_loss 变化趋势

Fig. 14 Trends of d_loss on CIFAR-10

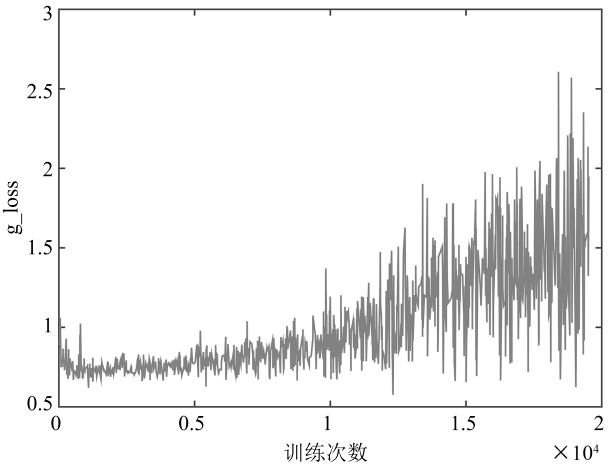

图 15 CIFAR-10 上 g_loss 变化趋势

Fig. 15 Trends of g_loss on CIFAR-10

结果进行对比, 图 17 为两个模型损失函数随迭代次数变换的对比, C-DCGAN 与传统 CNN 相比优势明显, 证明分类器的特征提取层在对抗训练中提取特征的有效性.

两个模型准确率随迭代次数变化的对比如图 18 所示, C-DCGAN 的准确率从开始就能达到 79 %, 迭代 10 个 Epoch 时结果明显高于传统 CNN, 再次证明用生成对抗模型预训练达到数据增强方法的可行性.

表 2 为本文方法与其他方法[11, 21] 识别结果的对比.

表 2 中 DCGAN + L2-SVM 方法同为生成对抗网络模型用于分类, 该方法先在 Imagenet-1k 上以无监督学习的方式进行了预训练, 之后将训练好的特征连接一个 L2-SVM 分类器, 在 CIFAR-10 数据集上进行分类, 而本文方法仅使用 CIFAR-10 进行训练和测试, 结果优于传统 DCGAN 方法, 证明 C-DCGAN 模型相比于传统 DCGAN 模型, 更具有在相对较小的数据集中提取到更多特征的能力.

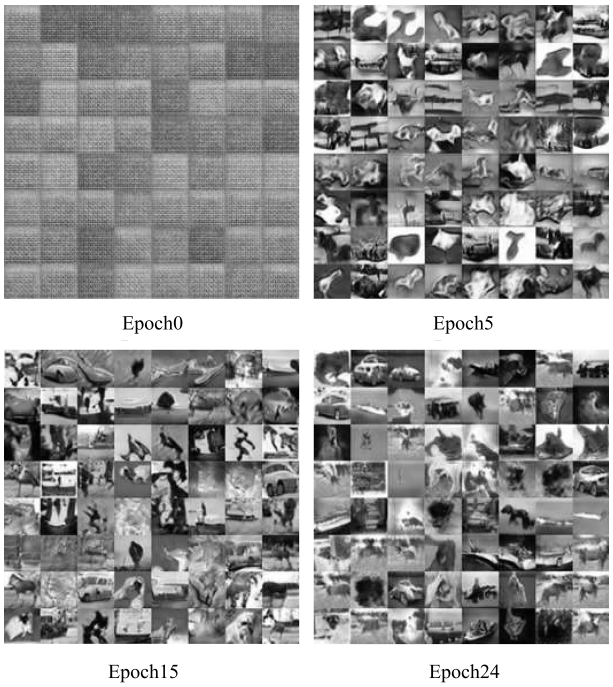

图 16 CIFAR-10 生成样本

Fig. 16 The samples generated by CIFAR-10

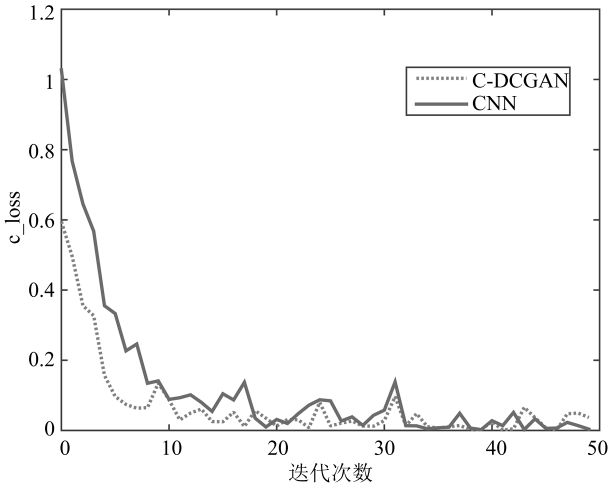

图 17 CIFAR-10 上 c_loss 变化趋势

Fig. 17 Trends of c_loss on CIFAR-10

除此之外, 本文方法的识别率均优于表 2 中其他传统无监督和有监督算法, 证明了该方法的可行性.

6 结论

本文结合条件生成对抗网络与深度卷积生成对抗网络提出条件深度卷积对抗网络 (C-DCGAN), 利用该模型的判别器提取特征用于图像识别, 条件深

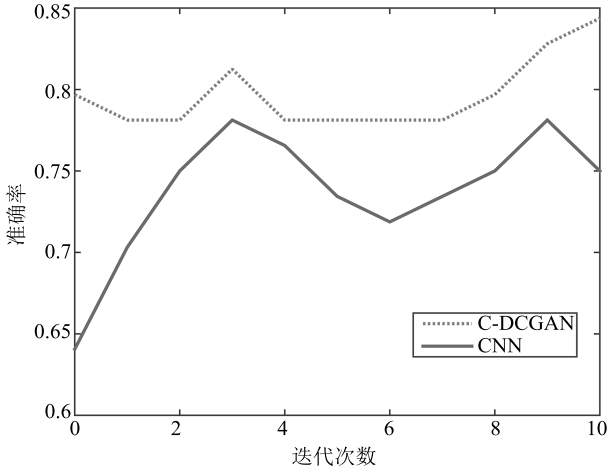

图 18　CIFAR-10 上准确率变化趋势

Fig. 18　Trends of accuracy on CIFAR-10

表 2　CIFAR-10 上各方法准确率对比

Table 2　The recognition accuracy comparison on CIFAR-10

识别方法	准确率 (%)
1 Layer K-means	80.6
3 Layer K-means Learned RF	82.0
View Invariant K-means	81.9
Cuda-convnet (CNN)	82.0
DCGAN + L2-SVM	82.8
C-DCGAN + Softmax	**84**

度卷积对抗网络不仅可以在条件的限制下生成预期的样本, 而且利用卷积层提取特征的能力较为高效地生成高质量样本. 将条件深度卷积生成对抗网络用于图像分类, 在 MNIST 和 CIFAR-10 数据集上进行仿真实验, 结果表明, 相比于其他方法, 本文所提方法不但加快了收敛速度, 减少了迭代次数, 同时有效提高了图像分类识别率, 证明本文所提的条件深度卷积生成对抗网络在图像识别领域中的可行性. 下一步将针对生成模型与判别模型在对抗训练过程中速度较慢, 以及判别器和生成器需要一个衡量标准来告知模型何时能够达到最优等问题进行深入研究.

References

1 Wang Kun-Feng, Gou Chao, Duan Yan-Jie, Lin Yi-Lun, Zheng Xin-Hu, Wang Fei-Yue. Generative adversarial networks: the state of the art and beyond. *Acta Automatica Sinica*, 2017, **43**(3): 321−332
(王坤峰, 苟超, 段艳杰, 林懿伦, 郑心湖, 王飞跃. 生成式对抗网络 GAN 的研究进展与展望. 自动化学报, 2017, **43**(3): 321−332)

2 Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. In: Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, Canada: ACM, 2014. 2672−2680

3 Ratliff L J, Burden S A, Sastry S S. Characterization and computation of local Nash equilibria in continuous games. In: Proceedings of the 51st Communication, Control, and Computing (Allerton). Monticello, IL, USA: IEEE, 2013. 917−924

4 Goodfellow I. NIPS 2016 tutorial: generative adversarial networks. arXiv preprint arXiv: 1701.00160, 2016.

5 Li J W, Monroe W, Shi T L, Jean S, Ritter A, Jurafsky D. Adversarial learning for neural dialogue generation. arXiv preprint arXiv: 1701.06547, 2017.

6 Yu L T, Zhang W N, Wang J, Yu Y. SeqGAN: sequence generative adversarial nets with policy gradient. In: Proceedings of the 31st AAAI Conference on Artificial Intelligence. San Francisco, CA, USA: AAAI, 2017. 2852−2858

7 Hu W W, Tan Y. Generating adversarial malware examples for black-box attacks based on GAN. arXiv preprint arXiv: 1702.05983, 2017.

8 Chidambaram M, Qi Y J. Style transfer generative adversarial networks: learning to play chess differently. arXiv preprint arXiv: 1702.06762, 2017.

9 Mirza M, Osindero S. Conditional generative adversarial nets. arXiv preprint arXiv: 1411.1784, 2014.

10 Chang Liang, Deng Xiao-Ming, Zhou Ming-Quan, Wu Zhong-Ke, Yuan Ye, Yang Shuo, Wang Hong-An. Convolutional neural networks in image understanding. *Acta Automatica Sinica*, 2016, **42**(9): 1300−1312
(常亮, 邓小明, 周明全, 武仲科, 袁野, 杨硕, 王宏安. 图像理解中的卷积神经网络. 自动化学报, 2016, **42**(9): 1300−1312)

11 Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv: 1511.06434, 2015.

12 Jin Lian-Wen, Zhong Zhuo-Yao, Yang Zhao, Yang Wei-Xin, Xie Ze-Cheng, Sun Jun. Applications of deep learning for handwritten Chinese character recognition: a review. *Acta Automatica Sinica*, 2016, **42**(8): 1125−1141
(金连文, 钟卓耀, 杨钊, 杨维信, 谢泽澄, 孙俊. 深度学习在手写汉字识别中的应用综述. 自动化学报, 2016, **42**(8): 1125−1141)

13 Chen Rong, Cao Yong-Feng, Sun Hong. Multi-class image classification with active learning and semi-supervised learning. *Acta Automatica Sinica*, 2011, **37**(8): 954−962
(陈荣, 曹永锋, 孙洪. 基于主动学习和半监督学习的多类图像分类. 自动化学报, 2011, **37**(8): 954−962)

14 Arjovsky M, Bottou L. Towards principled methods for training generative adversarial networks. arXiv preprint arXiv: 1701.04862, 2017.

15 Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 2014, **15**(1): 1929−1958

16 Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Proceedings of the 32nd International Conference on Machine Learning. Lille, France: PMLR, 2015. 448−456

17 Simon M, Rodner E, Denzler J. ImageNet pre-trained models with batch normalization. arXiv preprint arXiv: 1612.01452, 2016.

18 Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, Nevada: ACM, 2012. 1097−1105

19 LeCun Y, Cortes C, Burges C J C. The MNIST database of handwritten digits [Online], available: http://yann.lecun.com/exdb/mnist/, July 12, 2016

20 Xu Ke. Study of Convolutional Neural Network Applied on Image Recognition [Master thesis], Zhejiang University, China, 2012.
(许可. 卷积神经网络在图像识别上的应用的研究 [硕士学位论文]. 浙江大学, 中国, 2012.)

21 Krizhevsky A, Nair V, Hinton G. The CIFAR-10 dataset [Online], available: http://www.cs.toronto.edu/kriz/cifar.html, July 24, 2017

唐贤伦 重庆邮电大学计算机科学与技术学院教授. 主要研究方向为模式识别与智能系统, 深度学习.
E-mail: tangxl@cqupt.edu.cn
(**TANG Xian-Lun** Professor at the College of Computer Science and Technology, Chongqing University of Posts and Telecommunications. His research interest covers pattern recognition and intelligent system, deep learning.)

杜一铭 重庆邮电大学计算机科学与技术学院硕士研究生. 主要研究方向为图像识别, 生成对抗网络. 本文通信作者.
E-mail: jimmy4code@gmail.com
(**DU Yi-Ming** Master student at the College of Computer Science and Technology, Chongqing University of Posts and Telecommunications. His research interest covers image recognition, generative adversarial networks. Corresponding author of this paper.)

刘雨微 重庆邮电大学自动化学院硕士研究生. 主要研究方向为深度学习, 模式识别.
E-mail: yuweiliu1993@hotmail.com
(**LIU Yu-Wei** Master student at the College of Automation, Chongqing University of Posts and Telecommunication. Her research interest covers deep learning, pattern recognition.)

李佳歆 重庆邮电大学自动化学院硕士研究生. 主要研究方向为深度学习, 文本识别.
E-mail: suggercandy@outlook.com
(**LI Jia-Xin** Master student at the College of Automation, Chongqing University of Posts and Telecommunication. Her research interest covers deep learning, text recognition.)

马艺玮 重庆邮电大学自动化学院副教授. 主要研究方向为智能控制, 系统优化.
E-mail: mayw@cqupt.edu.cn
(**MA Yi-Wei** Associate professor at the College of Automation, Chongqing University of Posts and Telecommunications. Her research interest covers intelligent control, system optimization.)