

# 基于条件生成式对抗网络的数据增强方法

陈文兵, 管正雄\*, 陈允杰

(南京信息工程大学 数学与统计学院 南京 210044)

(\* 通信作者电子邮箱 zhengxguan@163.com)

**摘要:** 深度卷积神经网络(CNN)在大规模带有标签的数据集训练下,训练后模型能够取得高的识别率或好的分类效果,而利用较小规模数据集训练 CNN 模型则通常出现过拟合现象。针对这一问题,提出了一种集成高斯混合模型(GMM)及条件生成式对抗网络(CGAN)的数据增强方法并记作 GMM-CGAN。首先,通过围绕核心区域随机滑动采样的方法增加数据集样本数量;其次,假定噪声随机向量服从 GMM 描述的分布,将它作为 CGAN 生成器的初始输入,图像标签作为 CGAN 条件,训练 CGAN 以及 GMM 模型的参数;最后,利用已训练 CGAN 生成符合样本真实分布的新数据集。对包含 12 种雾型 386 个样本的天气形势图基准集利用 GMM-CGAN 方法进行数据增强,增强后的数据集样本数多达 38 600 个,将该数据集训练的 CNN 模型与仅使用仿射变换增强的数据集及 CGAN 方法增强的数据集训练的 CNN 模型相比,实验结果表明,前者的平均分类正确率相较于后两个模型分别提高了 18.2% 及 14.1%,达到 89.1%。

**关键词:** 图像分类;深度卷积神经网络;高斯混合模型;有条件对抗神经网络;数据增强算法

中图分类号: TP391.41 文献标志码: A

## Data augmentation method based on conditional generative adversarial net model

CHEN Wenbing, GUAN Zhengxiong\*, CHEN Yunjie

(School of Mathematics and Statistics, Nanjing University of Information Science & Technology, Nanjing Jiangsu 210044, China)

**Abstract:** Deep Convolutional Neural Network (CNN) is trained by large-scale labelled datasets. After training, the model can achieve high recognition rate or good classification effect. However, the training of CNN models with smaller-scale datasets usually occurs overfitting. In order to solve this problem, a novel data augmentation method called GMM-CGAN was proposed, which was integrated Gaussian Mixture Model (GMM) and CGAN (Conditional Generative Adversarial Net). Firstly, sample number was increased by randomly sliding sampling around the core region. Secondly, the random noise vector was supposed to submit to the distribution of GMM model, then it was used as the initial input to the CGAN generator and the image label was used as the CGAN condition to train the parameters of the CGAN and GMM models. Finally, the trained CGAN was used to generate a new dataset that matched the real distribution of the samples. The dataset was divided into 12 classes of 386 items. After implementing GMM-CGAN on the dataset, the total number of the new dataset was 38 600. The experimental results show that compared with CNN's training datasets augmented by Affine transformation or CGAN, the average classification accuracy of the proposed method is 89.1%, which is improved by 18.2% and 14.1%, respectively.

**Key words:** image classification; deep Convolution Neural Network (CNN); Gaussian Mixture Model (GMM); Conditional Generative Adversarial Net (CGAN); data augmentation algorithm

## 0 引言

卷积神经网络(Convolution Neural Network, CNN)是一种有监督学习模型,在视觉处理和图像分类中性能优越<sup>[1-7]</sup>。LeCun 等<sup>[1]</sup>提出的 LeNet-5 网络是 CNN 的最初模型,该模型采用基于梯度的反向传播(Back Propagation, BP)算法对网络进行有监督的训练;经过训练的网络通过交替连接的卷积层和下采样层将原始图像转换成一系列的特征图,再通过全连接层实现对图像特征图分类或识别,卷积层中的卷积核发挥人类视觉的感受野功能,卷积核将图像的低级局部区域信息转换成人类视觉的更高级形式。Krizhevsky 等<sup>[2]</sup>提出一种

AlexNet 网络架构,该架构在大小为 1400 万张样本、涵盖 2 万个类别的图像数据集 ImageNet 上参加图像分类竞赛,它以准确超越第二名 11% 的巨大优势夺得了 2012 年冠军,这一惊人的成绩引起了研究人员的普遍关注,并使得 CNN 成为近年的研究热点。Simonyan 等<sup>[3]</sup>基于 AlexNet 针对 CNN 深度进行了专门研究,并提出了 VGGNet 网络架构,该网络架构的各卷积层均采用  $3 \times 3$  的卷积核,通过对比基于不同深度网络架构的图像分类性能,证明了增加网络架构的深度有助于提升图像分类的准确度。近年来,对 CNN 模型架构的研究及应用仍然在迅速发展之中,在模型架构研究方面,GoogLeNet<sup>[4]</sup>、ResNet<sup>[5]</sup>等受到广泛关注;另一方面,由前述

收稿日期:2018-05-14;修回日期:2018-06-26;录用日期:2018-07-03。

基金项目:国家自然科学基金资助项目(61672291);北极阁基金资助项目(BJG201504)。

作者简介:陈文兵(1964—),男,安徽东至人,副教授,硕士,主要研究方向:计算数学、模式识别、图像处理;管正雄(1993—),男,安徽芜湖人,硕士研究生,主要研究方向:模式识别、图像处理;陈允杰(1980—),男,江苏南京人,教授,博士,主要研究方向:计算数学、模式识别、图像处理。

模型的训练、测试及分类应用可以看出,良好性能的取得依赖于大规模图像数据集的支撑,如 LeNet-5 采用的训练集是样本数为 60 000、分类标签个数为 10 的 MNIST (Modified National Institute of Standards and Technology) 数据集, AlexNet、VGGNet 等均采用训练集大小为 1 400 万张、涵盖 2 万个类别的 ImageNet 数据集进行训练、测试。由此可见,训练集的规模对 CNN 性能发挥着至关重要的影响。

然而,在现实世界中由于受自然因素的影响和数据记录条件的限制,得到大尺度有标签的数据集通常是不现实的,往往仅有少量的、带标签的数据样本。如某地区为了建立基于浓雾天气形势场的智能预报模型,由于天气形势场实际上就是一些等高线组成的纹理图,雾型与纹理之间具有高度的关联性,因此,利用 CNN 建模是解决这一问题的最佳选择。然而,该地区仅记录了 2010 年以来的天气形势图及其对应的出雾记录,样本集收集了 386 个样本,对应的雾型 12 类(即分类标签数 12 个)。若直接采用该样本集训练 CNN 模型,则训练出的模型必然缺乏泛化性<sup>[8]</sup>,因此缺乏可信性及可靠性。因此,在建立可信性及可靠性 CNN 模型之前,需要寻找一种可靠的扩展数据样本及多样性的方法,即所谓的数据增强(Data Augmentation)方法。

在数据增强研究方面,Bjerrum 等<sup>[9]</sup>通过使用仿射变换生成新样本,将样本和新样本混合作为训练集输入到神经网络中,训练完成后模型分类结果误差控制在 0.35% 以下。Goodfellow 等<sup>[10]</sup>提出的生成式对抗网络(Generative Adversarial Net, GAN)是一种生成式模型,其主要思想如下:在结构上受博弈论中的二人零和博弈(即二人的利益之和为零,一方的所得正是另一方的所失)的启发,由一个生成器  $G$  和一个判别器  $D$  构成。 $G$  捕捉真实数据样本的数学分布模型,并由学习到的分布模型生成新的数据样本; $D$  是一个二值分类器,用处是判别输入是真实数据还是生成的样本。二者不断学习,提高各自的生成能力和判别能力。Mirza 等<sup>[11]</sup>提出条件生成式对抗网络(Conditional Generative Adversarial Network, CGAN)模型,该模型是有条件控制的 GAN,通过对生成器和判别器添加相同的条件  $Y$ (例如数据的标签),从而实现 GAN 模型控制条件。目前有很多研究自动编码器(AutoEncoder, AE)、变分自动编码器(Variational AutoEncoder, VAE)结合 GAN 的工作<sup>[12-14]</sup>,目的在于提升 GAN 生成图像的真实性和多样性。

将现有的数据增强算法如仿射变换、GAN 等应用于天气形势图,实验显示生成的新数据集出现重复率高、多样性低等问题,利用生成的数据集训练 CNN 模型,所训练模型分类的正确率仍不理想。综上,为了更好地解决天气形势图问题,提出一种集成高斯混合模型(Gaussian Mixture Model, GMM)及 CGAN 模型的数据增强方法,该方法不仅生成类似样本的新图像,在提升生成样本的多样性方面与传统方法相比有显著改进。

## 1 相关数据增强算法

### 1.1 仿射变换

仿射变换是一种二维坐标  $(x, y)$  到二维坐标  $(u, v)$  的线性变换,其数学表达式如式(1):

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} \cos \theta & -\sin \theta & a - a \cos \theta + b \sin \theta \\ \sin \theta & \cos \theta & b - a \sin \theta - b \cos \theta \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (1)$$

式(1)中的系数矩阵被称为仿射变换矩阵。其中: $\theta$ 为图像旋转的角度, $a$ 为图像平移的横坐标移动距离, $b$ 为图像平移的纵坐标移动距离。

Bjerrum 等<sup>[9]</sup>提出了基于仿射变换(Affine Transformation)的数据增强方法,通过对样本图像进行放大、缩小、平移、旋转以实现生成类似样本。实验中,随机生成  $x$  轴的位移为  $a$ ,  $y$  轴的位移为  $b$  和旋转角度为  $\theta$  的仿射变换矩阵  $A$ ,作用在输入图像  $x$  上,变换后的新图像为  $Ax$ 。由于仿射变换是一种全局图像变换,因而在聚焦于局部区域的多样性方面该变换无法实现。

### 1.2 GAN 及衍生模型

事实上,这个学习优化过程是一个极小极大博弈(Minimax game)问题,即寻找二者之间的一个平衡点,如果达到该平衡点, $D$  无法判断数据来自  $G$  还是真实样本,此时  $G$  达到最优状态。大量的实践已经证明可利用 GAN 解决训练集中样本数量过少的问题,如 Gurumurthy 等<sup>[15]</sup>利用改进的 GAN 增强小数据集以提升训练器的分类精度;王坤峰等<sup>[16]</sup>提出多个 GAN 衍生模型以增强数据集。

GAN 的结构如图 1 所示, $D$  和  $G$  分别表示判别器和生成器,它们的结构都为 CNN。 $D$  的输入为真实数据  $x$ ,输出为 1 或 0; $G$  的输入是一维随机噪声向量  $z$ ,输出是  $G(z)$ 。训练的目标是使得  $G(z)$  的分布尽可能接近真实数据的分布  $p_{\text{data}}$ 。 $D$  的目标是实现对输入数据的二值分类,若输入来源于真实样本,则  $D$  的输出为 1;若输入为  $G(z)$ ,则  $D$  的输出为 0。 $G$  的目标是使自己生成的数据  $G(z)$  在  $D$  上的表现  $D(G(z))$  和真实数据  $x$  在  $D$  上的表现  $D(x)$  尽可能一致。 $G$  的损失函数按式(2)计算:

$$\min_G V_G(D, G) = \min_G (E_{z \sim p_z} (\ln (1 - D(G(z)))) \quad (2)$$

式(2)描述的是, $G$  在不断对抗学习的过程中,生成的数据  $G(z)$  越来越接近真实样本, $D$  对  $G(z)$  的判别也越来越模糊。 $D$  的损失函数按式(3)计算:

$$\max_D V_D(D, G) = \max_D (E_{x \sim p_{\text{data}}} (\ln D(x)) + E_{z \sim p_z} (\ln (1 - D(z)))) \quad (3)$$

综上, $G$  和  $D$  的总体损失函数可以描述如式(4)所示:

$$\min_G \max_D V(D, G) = \min_G \max_D (E_{x \sim p_{\text{data}}} (\ln D(x)) + E_{z \sim p_z} (\ln (1 - D(z)))) \quad (4)$$

传统的 GAN 模型一次只能学习一类数据,对于包含多个类的数据样本集,需逐类学习及生成相应类的被增强样本集,因此,效率低是模型的主要缺陷。为了解决以上问题,Mirza 等<sup>[11]</sup>提出了 CGAN 模型,CGAN 的结构如图 2 所示。该模型通过对生成器和判别器添加相同的条件  $Y$ (例如:数据的标签),从而使 GAN 模型具有多类数据的生成能力。

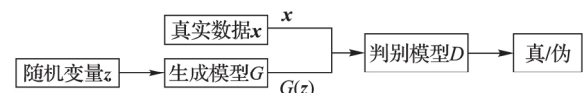


图1 GAN结构示意图

Fig. 1 Schematic diagram of GAN structure

与传统 GAN 对比,CGAN 模型仅对前者的总体损失函数进行了修改,新的总体损失函数如式(5):

$$\min_G \max_D V(D, G) = \min_G \max_D (E_{x \sim p_{\text{data}}} (\ln D(x|Y)) + E_{z \sim p_z} (\ln (1 - D(z|Y))) ) \quad (5)$$

然而, GAN 及 CGAN 在训练样本过少的情况下, 均存在  $G$  和  $D$  过早达到平衡点现象, 致使  $G$  生成的数据重复度高, 数据多样性不足。

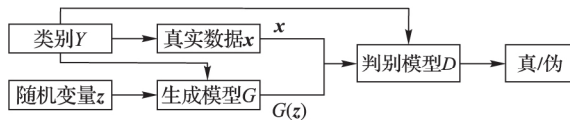


图2 CGAN结构示意图

Fig. 2 Schematic diagram of CGAN structure

## 2 GMM-CGAN

如前所述, 生成器  $G$  通过单一分布描述训练数据样本的分布, 不难理解单一分布对样本数据特征多样性难以反映, 其直接后果是训练的生成器  $G$  生成的数据样本特征单一, 难以达成样本数据集增强的目的。而高斯混合模型 (GMM) 的实质是利用  $m$  ( $m \geq 3$ ) 个正态分布来刻画样本整体的多样性特征, 通过训练学习后, 建立由  $m$  个组件 (即  $m$  个正态分布) 构成的混合分布模型。一方面多组件构成的混合模型能够更好地刻画样本的多样性特征, 另一方面这种数据特征的多样性又受到每个组件的约束, 使得混合模型生成的新样本既具有多样性又保持与原样本之间特征的相似性。基于此, 为了解决上述存在的问题, 将 GMM 集成到 CGAN 模型进而提出一种全新的 GMM-CGAN 数据增强框架, 这个框架在理论上是可行的。

GAN 中的生成器  $G$  的目标是使得  $p_{\text{data}}(G(z))$  尽可能接近样本分布, 其中  $p_{\text{data}}(G(z))$  是描述  $G(z)$  的分布。根据概率的乘法公式,  $p_{\text{data}}(G(z)|z)$  可写成一个已知的先验分布密度函数  $p_z(z)$ , 乘以  $p_{\text{data}}(G(z)|z)$ , 如式(6)所描述。结合前面的分析, 通过提升先验分布的多样性, 从而提升  $G(z)$  的多样性, 达到生成样本多样性的目的。首先, 假设先验分布的密度函数  $p_z(z)$  是有  $m$  个组件 GMM, 如式(7), 同时假设每个高斯组件的协方差矩阵为对角阵。

$$p_{\text{data}}(G(z)) = \int_z p(G(z)|z) dz = \int_z p_{\text{data}}(G(z)|z) p_z(z) dz \quad (6)$$

$$p_z(z) = \sum_{i=1}^m \pi_i N(z; \mu_i, \sigma_i) \quad (7)$$

其中  $N(x; \mu_i, \sigma_i)$  表示高斯混合模型的概率密度函数, 具体形式如式(8), 在 GAN 训练的过程中, 由于参数  $\pi_i$  不能被优化, 设  $\pi_i = 1/m$  以简化计算:

$$N(x; \mu_i, \sigma_i) = \frac{1}{(2\pi^{(n/2)}) |\sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T \sigma^{-1}(x-\mu)} \quad (8)$$

接着, 利用 Kingma 等<sup>[12]</sup> 提出的重复调参技术 (Reparameterization trick) 生成服从先验分布的一维随机噪声向量  $z$ ,  $z$  如式(9)计算:

$$z = \mu_i + \sigma_i \delta; \delta \sim N(0, 1) \quad (9)$$

其中:  $\mu_i, \sigma_i$  为第  $i$  个高斯组件的均值和标准差。重复调参技术优点在于: 可将高斯组件的参数看作为网络参数的一部分进而与网络参数一起训练及优化。

综合式(6)、(7)、(9), 可导出式(10):

$$p_{\text{data}}(G(z)) = \int_{i=1}^m \frac{p_{\text{data}}(G(\mu_i + \sigma_i \delta) | \delta) p(\delta) d\delta}{m} \quad (10)$$

式(10)中,  $u = [u_1, u_2, \dots, u_N]^T$ ,  $\sigma = [\sigma_1, \sigma_2, \dots, \sigma_N]^T$ ,  $m$  为高斯组件个数,  $N$  为  $z$  的维度。高斯组件个数与生成样本多样性密切相关, 实验分析表明, 当  $m$  在 [20, 30] 内变化时, 生成的样本效果较好。为了防止在实验中  $\sigma$  的值变为 0, 在生成器  $G$  的损失函数中添加关于  $\sigma$  的  $L_2$  正则化项, 修改后的生成器损失函数如式(11):

$$\min_G V_G(D, G) = \min_G E_{z \sim p_z} [\ln(1 - D(G(z)))] + \lambda \sum_{i=1}^N \frac{(1 - \sigma_i)^2}{N} \quad (11)$$

GMM-CGAN 模型结构如图 3 所示。GMM-CGAN 的参数需初始化, 由于对应于不同  $Y$  条件 (样本的标签) 的数据分布不相同的, 因此, 对于每一  $Y$  条件需要对  $\mu, \sigma$  向量初始化, 令  $\mu_i \sim U(-1, 1)$ ,  $\sigma_i \in (0, 1)$ , 其中  $U(-1, 1)$  表示区间  $(-1, 1)$  上的均匀分布, 标准差  $(0, 1)$  区间上随机选取。

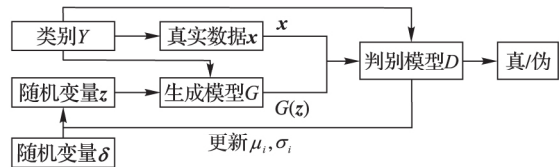


图3 GMM-CGAN结构示意图

Fig. 3 Schematic diagram of GMM-CGAN structure

参数  $\mu, \sigma$  按上述方法初始化后, 令  $z = \mu_k + \sigma_k \delta$ ,  $\delta \sim N(0, 1)$ ,  $k$  按照顺序从 1 到  $m$  取值, 将  $z$  输入  $G$  进入 CGAN 的训练程序, 从而达到逐个训练、优化高斯组件参数  $\mu_k, \sigma_k$ ,  $k \in (1, m)$  的目的。

在 CGAN 被训练后, 利用  $G$  生成新的样本, 步骤如下:

- 1) 选定需生成样本的标签;
  - 2) 在该标签下从  $\mu, \sigma$  向量中任选一对分量  $\mu_h, \sigma_h$ ,  $h \in (1, m)$ , 并计算  $z = \mu_h + \sigma_h \delta$ ,  $\delta \sim N(0, 1)$ ;
  - 3) 将  $z$  输入生成器  $G$  后, 即为生成的新样本  $G(z)$ ;
- 重复 1) ~ 3), 即可生成需要更具多样性的被增强的数据样本集。

## 3 实验分析与评价

### 3.1 原始数据集

#### 3.1.1 浓雾天气形势图

江苏省气象科学研究所整理收集了自 2010 年以来所有雾型天气形势图, 雾型个例 77 个, 每个雾型个例由记录一个完整成雾过程的若干幅天气形势图组成, 一般由 4 ~ 5 张纹理类似、尺寸为  $1600 \times 1500$  图像组成。气象工作人员根据雾型将这 77 个例分为 12 类别。然而, 深入分析这 12 个雾型类别对应的天气形势图发现, 即使两个个例同属于一个类别, 不同个例的形势图纹理间的差异性却很大, 故样本的标签不能以类别进行标记, 而以个例标记更为适当。采用 77 个分类的 one-hot 编码编制样本标签。在这样的编码机制下, 每个类中有至少 4 张形势图, 由于在首个历时及最后的历时天气形势图未入型, 故剔除首尾历时未入型图后构成对应个例的样本集。通过这样的预处理后, 样本数据集中样本数为 386, 标签类别数为 77。对样本集按标准的 70% 对 30% 随机划分, 分割

后训练集样本个数为 231, 测试集样本个数为 155。

### 3.1.2 MNIST

MNIST<sup>[1]</sup> 是机器学习的常用数据集, 它由数字 0~9 共计 10 类别 6000 张手写数字图像组成。从每个类别中随机抽取 50 张, 可以得到样本数为 500 的子集。对这样的数据集按标准的 70% 对 30% 随机分割, 将样本个数为 350 的数据集作为训练集, 样本个数为 150 的数据集作为测试集。

### 3.1.3 CIFAR 10

CIFAR 10 是另外一个机器学习的常用数据集, 它由 10 个类别, 每个类别 6000 张图, 共计 60000 张彩色图像组成。实验中将所有图像进行灰度化预处理, 从每个类别中随机抽取 50 张图像, 可以得到样本数为 500 张灰度图像的子集。对样本数为 500 张图像的子集, 按 70% 对 30% 随机分割, 将样本个数为 350 的数据集作为训练集, 样本个数为 150 的数据集作为测试集。

### 3.2 数据预增强

将样本中影响 CNN 分类的关键区域称为核心区域。在样本个数较少时, 通过滑动围绕核心区域的窗口反复重采样, 以实现数据的初步增强。

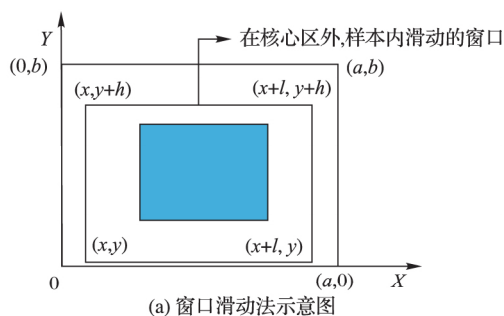
如图 5 所示, 设在长为  $a$ 、宽为  $b$  的样本图像上取长为  $l$ 、宽为  $h$  的区域为核心区域, 该核心区域左下角坐标为  $(x, y)$ 。再设滑动窗口长为  $\alpha$ 、宽为  $\beta$  ( $l < \alpha < a, h < \beta < b$ )。假设随机变量  $\delta, \xi$  是任意实施一次滑动所得到的滑动窗口左下角坐标 ( $\delta, \xi$ ), 则滑动窗口方算法步骤如下所述:

第 1 步 随机生成参数  $\delta \in [0, a - \alpha], \xi \in [0, b - \beta]$ 。

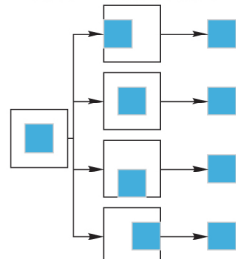
第 2 步 判断窗口的端点是否在核心区外, 即  $(\delta, \xi)$  是否满足不等式  $\begin{cases} x + l - \alpha \leq \delta \leq x \\ y + h - \beta \leq \xi \leq y \end{cases}$ 。

第 3 步 若满足第 2 步, 则输出截取的窗口图像; 否则返回第 1 步。

将滑动窗口法应用于样本集中的每个样本, 可得到一个以核心区域为主导的、被初步增强的样本集, 图 4 展示了该方法的演化过程。



(a) 窗口滑动法示意图



(b) 对同一张样本多次使用

图 4 窗口滑动法

Fig. 4 Sliding window method

具体到浓雾天气形势图数据集, 由于图像的经纬度及大

小均一样, 这里设定图像中心的  $800 \times 800$  正方形区域为核心区域。在滑动窗口法中设窗口长  $a = 1000$ , 宽为  $b = 1000$  并应用该方法, 对每一张样本作用 100 次, 可使得样本个数扩展 100 倍。图 5 所示是部分预处理结果, 训练集及测试集的样本数分别达到 23100、15500, 总 38600 张。随机抽取的 MNIST 和 CIFAR 10 子集无需预增强操作, 直接进入 GMM-CGAN 处理阶段。

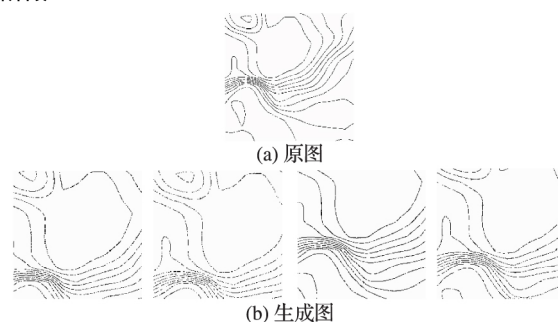
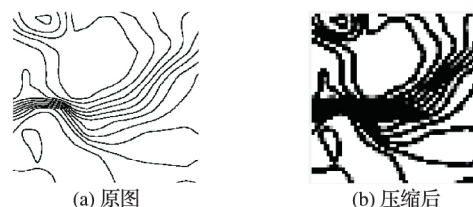


图 5 滑动窗口法生成图

Fig. 5 Results of sliding window method

LeNet-5、AlexNet 等 CNN 模型, 为了保证网络学习的效率以及限制参数的数量级在可控范围内, 在保持原有图像特征不丢失的情况下尽可能压缩输入图像的尺寸, 使 CNN 的参数在可训练的范围内。例如, LeNet-5 的输入图片尺寸为  $28 \times 28$ ; AlexNet 的输入图片尺寸为  $224 \times 224$ 。浓雾天气形势图经过前面的预增强, 其尺寸由  $1600 \times 1500$  转化为  $1000 \times 1000$ , 仍需进一步压缩处理。对比多种压缩算法的处理结果后, 最优的压缩方法为下采样法, 在保留纹理的情况下图像的尺寸由  $1000 \times 1000$  压缩到  $56 \times 56$ , 如图 6 所示。



(a) 原图

(b) 压缩后

图 6 压缩后的样本图像

Fig. 6 Compressed sample image

### 3.3 基于 GMM-CGAN 模型的数据增强

#### 3.3.1 实验采用的 CGAN 结构

实验中使用的 CGAN, 生成器  $G$  和判别器  $D$  为 CNN, 具体的结构如表 1~2 所示。

表 1 CGAN 的生成器结构

Tab. 1 Structure of generator in CGAN

层数	生成器结构
1	卷积层: $5 \times 5$ 卷积核, 64 个通道
2	非线性激活函数: ReLu
3	卷积层: $3 \times 3$ 卷积核, 128 个通道
4	非线性激活函数: ReLu
5	卷积层: $3 \times 3$ 卷积核, 256 个通道
6	非线性激活函数: tanh
7	全连接层

CGAN 的其他训练参数如下: 输入、输出图像尺寸为  $56 \times 56$ , 条件信息  $Y$  为数据集的标签, 训练批次为 50 个样本一组, 最大迭代次数设置为 1000, 梯度优化算法选择的是 Adam 优化器。CGAN 的结构图如图 7 所示。由于 CGAN 的生成器和



判别器的结构都被设置为浅层卷积神经网络,而且卷积核尺寸较小,通道数量较少,CGAN 需要学习的参数规模不大。GMM 为非神经网络结构,计算量小,故 GMM-CGAN 训练的计算复杂度与 CGAN 相比变化不大。

表 2 CGAN 的判别器结构  
Tab. 2 Structure of discriminator in CGAN

层数	判别器结构
1	卷积层:3×3 卷积核 64 个通道
2	非线性激活函数:ReLU
3	卷积层:3×3 卷积核 128 个通道
4	非线性激活函数:ReLU
5	卷积层:5×5 卷积核 256 个通道
6	非线性激活函数:Sigmoid
7	全连接层

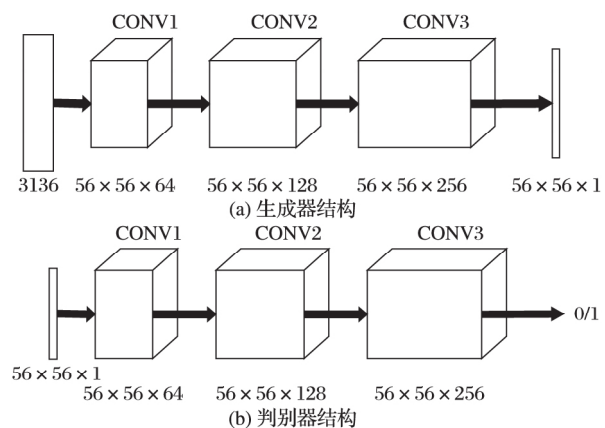


图 7 CGAN 的实例结构

Fig. 7 Structure of instance CGAN

3.3.2 基于 GMM-CGAN 模型生成的样本

浓雾天气形势图数据集已通过滑动窗口法进行了预增强,得到新样本集。以预增强的浓雾天气形势图新样本集,随机抽取的 MNIST 和 CIFAR 10 子集作为训练集,按照上述步骤训练 GMM-CGAN 模型。训练后模型中的生成器生成的 3 个数据集上的新样本与原样本对比如图 8 所示。

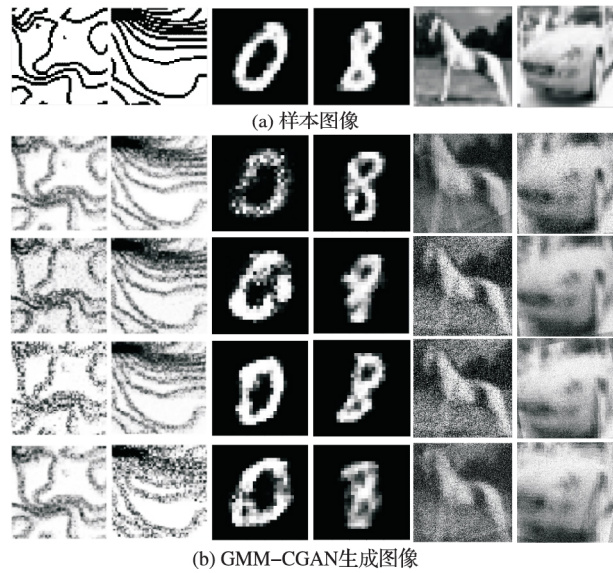


图 8 GMM-CGAN 的生成器生成结果

Fig. 8 Generated images by GMM-CGAN

3.4 增强后数据集的 CNN 分类

3.4.1 CNN 结构参数设置及卷积操作可视化

GMM-CGAN 模型的性能体现在其增强后的样本集上,如增强后的样本集训练出的 CNN 具有很高的分类准确率,那么认为所提模型是高效的。CNN 的结构选择是一个重要的问题,因为近年提出的一些高性能 CNN 模型,例如 GOOGLE-NET、VGG and ResNet 等。与传统的 CNN 相比,这些网络的特殊模块可以一定程度上减弱模型过度拟合数据,例如:残差模块<sup>[5]</sup>、Inception 模块<sup>[4]</sup>、Dropout 层<sup>[3]</sup>等。如果使用以上提到的高性能 CNN 作为测试网络,在分析 CNN 分类效率时会产生困难,因为无法区分数据增强算法增强对分类结果的影响还是这些网络的结构使然。

综上,自定义了被称为 TestNet 的 CNN 模型,结构如图 9 所示,可以更加精确比较和分析不同数据增强算法应用到浓雾天气形势图集的数据增强的效果。

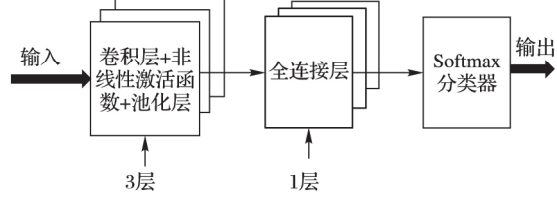


图 9 TestNet 的数据流图

Fig. 9 Data flow diagram of TestNet

该模型共有 3 个卷积层,其中第一卷积层(C1 层)中的卷积核尺寸为 5×5,第二卷积层(C2 层)和第三卷积层(C3 层)的卷积核尺寸为 3×3,卷积核选择尺寸较小利于减小网络的复杂度;三层卷积层的卷积核尺寸由大变小的目的是为了先总体后局部的学习样本特征;设置多个不同的卷积核有利于网络学习样本图像中不同的特征,卷积操作后生成的图像被称为特征图,C1 层共有 96 个特征图,C2 和 C3 分别有 256 和 384 个不同的特征图,网络具体的结构参数设置如表 3 所示。选择 ReLU 函数为网络的非线性激活函数,如式(12)所示:

$$f_{ReLU} = \max(0, x) \tag{12}$$

ReLU 函数与其他非线性激活函数<sup>[13]</sup>相比,具有计算简便、不易发生梯度爆炸等特点;在卷积层后设置下采样层,用于减少 CNN 整体的参数量。TestNet 的其他训练参数如下,输入图像尺寸为 56×56,训练批次(Batch)为 50,最大迭代次数设置为 1000,损失函数为交叉熵,梯度优化算法是 Adam 算法。为了观察逐层卷积后的特征图,每一层卷积后图像的输出结果,如图 10 所示。

表 3 TestNet 的结构  
Tab. 3 Structure of TestNet

层数	结构
1	卷积层:5×5 卷积核 96 个通道
2	非线性激活函数:ReLU
3	下采样层:采样步长 2×2,下采样区域大小 2×2
4	卷积层:3×3 卷积核 256 个通道
5	非线性激活函数:ReLU
6	下采样层:采样步长 2×2,下采样区域大小 2×2
7	卷积层:3×3 卷积核 384 个通道
8	非线性激活函数:ReLU
9	下采样层:采样步长 2×2,下采样区域大小 2×2
10	全连接层
11	Softmax 分类器

### 3.4.2 对比实验设计及评价指标

为了对比所提 GMM-CGAN 模型与传统数据增强算法的增强效果,设计了其他 3 个对比实验,其中一个为空白对照组,即不使用数据增强算法 (None),其余对比实验的数据增强算法为仅使用仿射变换、仅使用 CGAN。将在不同数据集上实现这 4 个实验。在相同数据集上,除各个实验使用的数据增强算法不同,训练集、测试集中的样本数量相同,其余控制变量均相同。实验环境的配置如下,硬件方面:CPU 是 Intel Core i7 9280,内存为 16 GB DDR4, GPU 采用的是 NVIDIA GTX1080。软件方面:操作系统是 Windows 10 64 b 版本,实现的平台是基于 Python 的 Tensorflow 框架,其中有 CUDA9.1 以及 CUDNN7 加速包的支持。

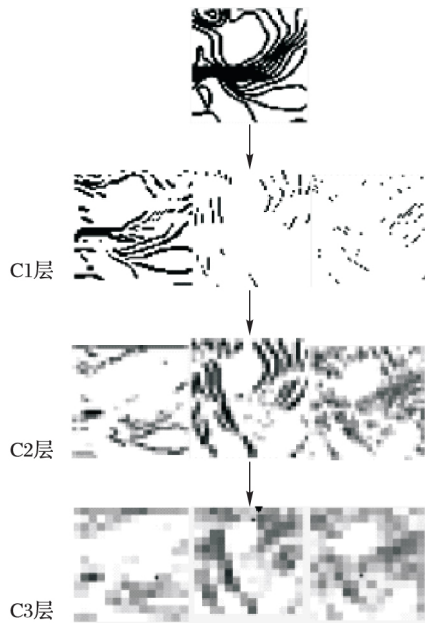


图 10 各个卷积层的输出结果

Fig. 10 Outputs of each CONV

利用数据增强后的数据集训练 TestNet,针对网络的分类结果,以平均分类正确率和过拟合比评价数据增强算法的性能。对于数据集中某一类图像的正确分类情况,定义了分类正确率如式(13):

$$Acc = \text{images\_correct} / \text{images} \quad (13)$$

其中:images\\_correct 表示在该类中网络分类正确的图像数量,images 表示该类中图像的总数。

对包含  $n$  类的样本集,定义了平均分类准确率如式(14):

$$AvgAcc = \frac{1}{n} \sum_{i=1}^n Acc_i \quad (14)$$

其中  $Acc_i$  表示第  $i$  类的分类正确率。

反映 CNN 是否过拟合训练数据的指标为过拟合比 (OverfitRatio),其定义如式(15):

$$OverfitRatio = \text{Train\_AvgAcc} / \text{AvgAcc} \quad (15)$$

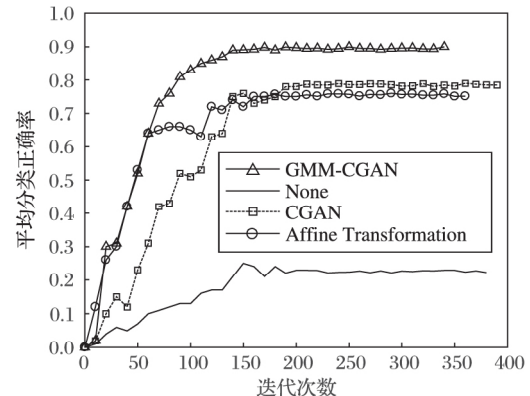
其中 OverfitRatio 中的 Train\_AvgAcc 表示用训练后的网络测试原训练集的平均分类正确率。

### 3.5 实验结果分析与评价

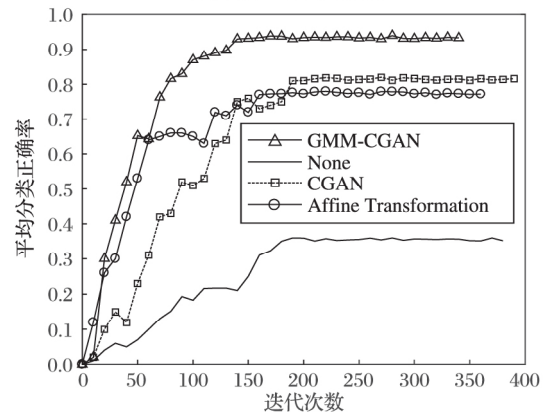
表 4 记录了在 3 个数据集上分别实现 4 个实验,共 12 个

实验的平均分类正确率及过拟合比。

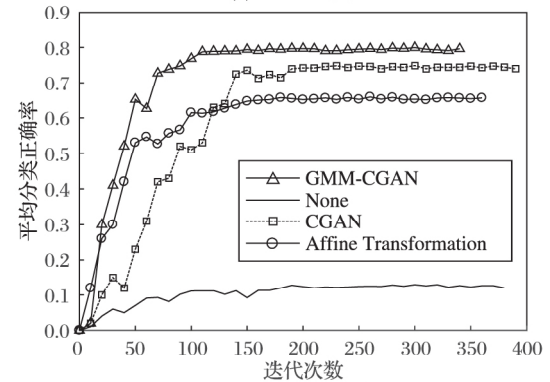
图 11 按数据集展示了在该数据集上实验的平均分类正确率随迭代次数增加的变化趋势,未使用数据增强算法的实验组,在各个数据集上平均分类正确率最低;使用所提 GMM-CGAN 的实验组,在各个数据集上平均分类正确率最高。过拟合比是反映模型过拟合数据程度的指标,过拟合比越低模型的泛化性越好,反之泛化性越差。表 4 展示了 12 个实验的过拟合比,因未使用数据增强算法的实验组中训练样本的相似度较高,所以过拟合比最高;其中使用所提的 GMM-CGAN 的实验组过拟合比最低,说明所提模型提升数据的多样性最高。



(a) 浓雾天气形势图结果



(b) MNIST结果



(c) CIFAR 10结果

图 11 不同增强算法在各数据集上的正确率曲线

Fig. 11 Curves of accuracy on different augmented datasets

综上所述,所提的 GMM-CGAN 模型具有收敛快,在相同迭代稳定后平均正确率高的特点。使用真实数据的实验证明所提模型是可靠的、高效的。

表 4 各数据集增强后的分类结果

Tab. 4 Classification results on different augmented datasets

编号	数据集	数据增强算法	平均分类 正确率/%	过拟 合比
1	浓雾天气 形势图	None	22.3	4.48
2		Affine Transformation	75.4	1.33
3		CGAN	78.1	1.28
4		GMM-CGAN	89.1	1.08
5	MNSIT	None	35.6	2.17
6		Affine Transformation	77.2	1.21
7		CGAN	81.4	1.08
8		GMM-CGAN	93.8	0.94
9	CIFAR 10	None	12.1	3.47
10		Affine Transformation	65.2	1.31
11		CGAN	74.3	1.42
12		GMM-CGAN	79.5	1.28

## 4 结语

本文所提的 GMM-CGAN 模型,在原有浓雾天气形势图基准集的基础上有效扩展了浓雾天气形势图数量,解决了浓雾天气形势图基准集因数据量偏小无法有效训练 CNN 的问题。GMM-CGAN 方法生成的新数据集所训练的 CNN,其平均分类准确率达到 89.1%,证明 GMM-CGAN 方法及所训练的 CNN 架构性能均高度可靠。未来工作将进一步研究其他类型小数据集场景(如数值型小数据集的增强)的增强模型。

### 参考文献 (References)

- [1] LECUN Y, BOSER B, DENKER J S, et al. Back propagation applied to handwritten zip code recognition[J]. Neural Computation, 1989, 1(4): 541–551.
- [2] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]// Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, Nevada: Curran Associates Inc., 2012: 1097–1105.
- [3] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J/OL]. arXiv Preprint, 2014, 2014: arXiv:1409.1556 (2014-09-04) [2015-04-10]. <http://arxiv.org/abs/1409.1556>.
- [4] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2015: 1–9.
- [5] HE K, ZHANG X, REN S. Deep residual learning for image recognition[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2016: 770–778.
- [6] 刘雨桐,李志清,杨晓玲.改进卷积神经网络在遥感图像分类中的应用[J]. 计算机应用, 2018, 38(4): 949–954. (LIU Y T, LI Z Q, YANG X L. Application of improved convolution neural network in remote sensing image classification[J]. Journal of Computer Applications, 2018, 38(4): 949–954.)
- [7] 安旭骁,邓洪敏,史兴宇.基于迷你卷积神经网络的停车场空车位检测方法[J]. 计算机应用, 2018, 38(4): 935–938. (AN X X, DENG H M, SHI X Y. Parking lot space detection method based on mini convolutional neural network[J]. Journal of Computer Applications, 2018, 38(4): 935–938.)
- [8] PEREZ L, WANG J. The Effectiveness of data augmentation in image classification using deep learning[J/OL]. arXiv Preprint, 2017, 2017: arXiv:1712.04621 [2017-12-13]. <http://arxiv.org/abs/1712.04621>.
- [9] BJERRUM E J. SMILES enumeration as data augmentation for neural network modeling of molecules[J/OL]. arXiv Preprint, 2017, 2017: arXiv:1703.07076 (2017-03-21) [2017-05-17]. <http://arxiv.org/abs/1703.07076>.
- [10] 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: 2672–2680.
- [11] MIRZA M, OSINDERO S. Conditional generative adversarial nets[J/OL]. arXiv Preprint, 2014, 2014: arXiv:1411.1784 [2014-11-06]. <http://arxiv.org/abs/1411.1784>.
- [12] KINGMA D P, WELING M. Auto-encoding variational Bayes[J/OL]. arXiv Preprint, 2013, 2013: arXiv:1312.6114 (2013-12-20) [2014-05-01]. <http://arxiv.org/abs/1312.6114>.
- [13] ROSCA M, LAKSHMINARAYANAN B, WARDEFARLEY D, et al. Variational approaches for auto-encoding generative adversarial networks[J/OL]. arXiv Preprint, 2017, 2017: arXiv:1706.04987 (2017-05-15) [2017-10-21]. <http://arxiv.org/abs/1706.04987>.
- [14] LARSEN A B L, LAROCHELLE H, WINTHER O. Autoencoding beyond pixels using a learned similarity metric[C]// Proceedings of the 33rd International Conference on International Conference on Machine Learning. New York: JMLR.org, 2016: 1558–1566.
- [15] GURUMURTHY S, SARVADEVABHATLA R K, BABU R V. DeLiGAN: Generative adversarial networks for diverse and limited data[C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2017: 4941–4949.
- [16] 王坤峰,苟超,段艳杰等.生成式对抗网络 GAN 的研究进展与展望[J].自动化学报, 2017, 43(3): 321–332. (WANG K F, GOU C, DUAN Y J, et al. Generative adversarial networks: the state of the art and beyond[J]. Acta Automatica Sinica, 2017, 43(3): 321–332.)

This work is partially supported by the National Natural Science Foundation of China (61672291), the Beijige Foundation (BJG201504).

**CHEN Wenbing**, born in 1964, M. S., associate professor. His research interests include computational mathematics, pattern recognition, image processing.

**GUAN Zhengxiong**, born in 1993, M. S. candidate. His research interests include pattern recognition, image processing.

**CHEN Yunjie**, born in 1980, Ph. D., professor. His research interests include computational mathematics, pattern recognition, image processing.