双判别器生成对抗网络及其在接触网鸟巢检测与半监督学习中的应用

金炜东*, 杨沛, 唐鹏

西南交通大学电气工程学院, 成都 611731

* 通信作者. E-mail: wdjin@home.swjtu.edu.cn

收稿日期: 2018-03-20; 接受日期: 2018-05-03; 网络出版日期: 2018-07-20

国家自然科学基金重点项目 (批准号: 61134002) 和国家重点研发计划子任务 (批准号: 2016YFB200401-102F) 资助项目

摘要 基于图像的接触网异常检测中, 鸟巢检测是其中一种典型的异常状况. 但在接触网图像数据中含有鸟巢的图像只占总数据很小的一部分, 属一类典型的不平衡数据分类问题. 使用机器学习模型解决该类不平衡数据问题, 模型对于数据特征的学习能力非常关键. 生成对抗网络 (generative adversarial networks, GANs) 可以无监督地学习到丰富的数据特征, 其优异的数据特征捕获能力被越来越多地证实和应用. 但由于 GANs 模型自身结构及原理所限, 直接将其用作图像分类模型效果并不理想. 本文针对图像分类任务对 GANs 模型进行了研究和改进, 提出了双判别器生成对抗网络 (double discriminator generative adversarial networks, DDGANs), 在接触网鸟巢检测中取得了令人满意的分类效果. 同时 DDGANs 也是一种有效的半监督学习模型, 对 MNIST 标准数据集的实验表明, 与其他模型相比, 准确率和收敛速度获得了明显提高.

关键词 生成对抗网络, 半监督学习, 不平衡数据分类, 异常检测, 接触网

1 引言

高速铁路电气化建设与运营管理中, 接触网异常检测是目前业界特别关注的关键技术, 基于智能视频分析的检测方法是其中的主要技术途径, 我们采用了多种数据分析方法对该智能视频分析问题进行了研究. 本文结合接触网鸟巢检测的应用背景, 探讨生成对抗网络的改进模型. 鸟巢是接触网异常检测中的一种典型状况, 搭建于电气化铁路接触网支架上的鸟巢极易引起接触网跳闸、绝缘子击穿等严重故障. 目前主要依靠人工巡检方式进行鸟巢检测, 成本高、效率低. 利用视频图像处理进行自动巡检是业界十分重视的技术发展方向.

引用格式: 金炜东, 杨沛, 唐鹏. 双判别器生成对抗网络及其在接触网鸟巢检测与半监督学习中的应用. 中国科学: 信息科学, 2018, 48: 888–902, doi: 10.1360/N112017-00290
Jin W D, Yang P, Tang P. Double discriminator generative adversarial networks and their application in detecting nests built in catenary and semisupervised learning (in Chinese). Sci Sin Inform, 2018, 48: 888–902, doi: 10.1360/N112017-00290

在接触网异常视频检测中,正常与异常状态的样本数据量相差悬殊(本文使用的京广高铁接触网视频监控图像中,正常数据与鸟巢异常数据之比约为50:1).因其数据不平衡性,使用常用图像分类方法如卷积神经网络(convolutional neural networks, CNN)和支持向量机(support vector machine, SVM)等对其进行分类效果并不理想.机器学习模型解决不平衡数据问题,模型对于数据特征的学习能力非常关键.

生成对抗网络(generative adversarial networks, GANs)是由Goodfellow等^[1]在2014年提出的一种基于深度神经网络的生成式模型. GANs提供了一种通过无监督学习获得数据深度表征的方法,通过反向传播算法和不断更新两个网络并执行竞争性学习达到其训练目的. GANs被证明可以生成视觉上足够逼真的图像^[2],其在图像合成、视频预测、图像语义解释、图像超分辨率重建等应用中都取得了非常不错的效果^[3~7].目前GANs仍然有一定的局限性,比如训练过程可能不稳定、可解释性不高、生成数据往往不能表达出真实数据的全部分布等.尽管存在这些问题,相比传统概率生成模型,GANs仍是更好的生成模型.其一,有别于传统的概率生成模型,GANs避免了Markov链式的学习机制.传统概率生成模型通常难以避免使用Markov链式的采样和推断以逼近真实数据分布,因此其计算复杂度非常高.而GANs是直接对数据进行采样和推断,提高了应用效率.其二,GANs的设计框架非常灵活,其实际应用场景更为广泛.其三,当概率密度不可计算时,传统依赖于数据自然性解释的一些生成模型就很难对其进行训练学习和应用,而由于GANs内部引入了一个对抗训练机制,因此在这种情况下依然可用.

GANs由生成器和判别器两部分组成,生成器旨在生成逼近真实样本的数据,判别器则用于判断输入数据是属于真实数据还是生成数据. GANs的目标是在具有高维参数的连续非凸问题中找到纳什均衡(Nash equilibrium)^[8],使生成器能够生成足够接近真实样本的数据,而判别器无法判断出输入数据来自生成样本还是真实样本.可见GANs的根本目标是让生成器学习到真实样本的数据特征,在GANs达到纳什均衡后,其对于生成样本和真实样本分类准确度会稳定在50%左右.因此,直接使用GANs模型无法有效解决分类问题.之后的一些改进模型虽然具有分类能力,如CGAN^[9],ACGAN^[10]等,但其分类效果不够理想.为解决前述不平衡数据分类问题,本文利用GANs对数据特征较强的学习表达能力,提出了新的生成对抗模型——双判别器生成对抗网络(double discriminator generative adversarial networks, DDGANs),并且在接触网鸟巢图像识别中获得了良好的检测效果.

同时,使用本文的DDGANs模型在半监督学习中也获得了令人满意的分类效果.深度神经网络通常是对大量标记数据进行训练,如何从大量未标记的数据中学习到有意义的特征表示,一直是一个备受关注的问题^[11].半监督学习旨在标记数据有限的情况下,利用大量未标记数据来提高模型性能.半监督分类问题是半监督学习中常见的问题,通过使用大量的无标签样本 $P_{ul} = \{x_{ul}^{(1)}, \dots, x_{ul}^{(n)}\}$ 改善有标签样本 $P_l = \{(x_l^{(1)}, y^{(1)}), \dots, (x_l^{(n)}, y^{(n)})\}$ 不足造成的监督方法过拟合、欠拟合以及其他性能缺陷,得到性能更优良的分类器^[12].应用中DDGANs模型对半监督学习表现出了优良的分类性能,在MNIST数据集中仅使用100张有标签样本进行训练的情况下,测试准确度达到98%以上.

2 算法模型

2.1 GANs

2.1.1 GANs 模型基本结构

如图1所示, GANs模型由两部分组成:生成器(generator, G)和判别器(discriminator, D).

图 1 GANs 模型结构图

Figure 1 Structure diagram of GANs

判别器 D 为多层感知机, D 是一个判别输入数据 $x \sim p_{\text{data}}(x)$ 来自真实数据分布 (而不是来自生成器生成的数据) 的概率函数: $D: D(x)$, 其输出为单个标量.

生成器 G 是输入为先验的噪声变量 $z \sim p_z(z)$ 的多层感知机, G 将 $p_z(z)$ 映射到数据空间中, 输出与 $x \sim p_{\text{data}}(x)$ 维度一致的生成数据, 表示为: $G(z)$. 训练 D 最大化对真实样本 $x \sim p_{\text{data}}(x)$ 和生成样本 $x \sim G(z)$ 的判别准确度, 同时通过最小化 $\log[1 - D(G(z))]$ 来训练 G .

D 和 G 使用下述价值函数式 (1) 进行二人极大极小化博弈 (two-player minimax game):

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log D(x)] + \mathbb{E}_{x \sim p_z(z)} [\log (1 - D(G(x)))]. \quad (1)$$

2.1.2 GANs 性能分析

引理1 (见参文 [1]) 在式 (1) 所示价值函数给定的 G , 最优判别器 D 为

$$D_G^*(x) = \frac{p_{\text{data}}(x)}{p_{\text{data}}(x) + p_g(x)}. \quad (2)$$

对于任意的 G , 将式 (2) 中最优判别器 $D_G^*(x)$ 带入式 (1) 得到

$$\begin{aligned} C(G) &= V(G, D_G^*(x)) \\ &= \mathbb{E}_{x \sim p_{\text{data}}} \log \frac{p_{\text{data}}(x)}{p_{\text{data}}(x) + p_g(x)} + \mathbb{E}_{x \sim p_g} \log \frac{p_g(x)}{p_{\text{data}}(x) + p_g(x)}. \end{aligned} \quad (3)$$

引理2 (见参文 [1]) 当且仅当 $p_g = p_{\text{data}}$ 时, 式 (3) 中 $C(G)$ 达到全局最小.

由引理 1 和 2 得, 当 GANs 趋近最优状态时, $D_G^*(x) = \frac{p_{\text{data}}(x)}{p_{\text{data}}(x) + p_g(x)} \rightarrow \frac{1}{2}$, 也就意味着, 此时不论生成还是真实数据, D 都已无法判断出其属性, 而 D 的分类性能亦不会随着训练的继续而进一步提高. 因此直接使用 GANs 原模型解决分类问题难以达到令人满意的效果.

2.1.3 直接使用 GANs 模型进行鸟巢检测实验分析

考虑将 GANs 中的判别器 D 直接作为二分类器对正常工况和含有鸟巢的图像进行分类. 只使用正常工况下的图像数据对 GANs 网络进行训练, 希望在 GANs 收敛后可以判别出正常工况及含有鸟巢的图像. 实验结果如图 2 和 3 所示, 图 3 中横坐标为训练的循环次数, 纵坐标为使用测试集对 D 网络进行测试的准确率结果. 从这两张图中可看出:

(1) 直接将 GANs 中判别器 D 用于分类任务的效果不稳定;

(2) 将 GANs 应用到分类问题时, 判别器准确率最高可达到 90% 左右, 且生成器可以生成具有接触网局部特征的图像. 这说明将 GANs 应用于鸟巢异常分类问题是可行的.

图 2 生成图像与真实图像对比

Figure 2 Generated images compared with real images

图 3 使用 GANs 对真实数据的分类性能

Figure 3 Classification accuracy of GANs

图 4 DDGANs 模型结构图

Figure 4 Structure of DDGANs

2.2 双判别器生成对抗网络 DDGANs

2.2.1 DDGANs 模型基本原理及结构

基于 2.1.3 小节中的实验分析, 为了利用 GANs 对数据特征较强的学习表达能力, 同时得到一个具有良好分类性能的模型. 通过对 GANs 原理的深入研究, 本文提出了一种可用于分类问题的改进生成对抗网络模型——双判别器生成对抗网络 DDGANs. 该模型结构如图 4 所示, DDGANs 由判别器 1 (discriminator 1, D_1) 和判别器 2 (discriminator 2, D_2) 及生成器 (generator, G) 组成. 针对图像分类问题, 本文的 DDGANs 模型中判别器以及生成器均为卷积神经网络.

生成器 G 的设计参考了 ACGAN^[10], 输入为先验的噪声变量 $z \sim p_z(z)$ 和数据类别标签 $c \sim p_c(c)$, G 将两者作为输入映射到数据空间中, 生成与 c 对应类别的生成数据 $x \sim G(z, c)$, 该生成数据与真实数据 $x \sim p_{\text{data}}(x)$ 具有相同的维度. G 将 $z \sim p_z(z)$ 和 $c \sim p_c(c)$ 映射到对应的数据空间中, 从大量无标签数据中学习特征表示, 并根据输入类别标签 $c \sim p_c(c)$ 生成对应的数据 $(x, c) \sim G(z, c)$, 解决了半监督学习中有标签数据不足的问题.

判别器 D_1 的设计同 GANs, 其输出为区间在 $(1, 0)$ 上的单个标量, $D_1(x)$ 表示 $x \sim p_{\text{data}}(x)$ 来源于真实数据 (而不是来自 G 生成数据) 的概率.

判别器 D_2 的输入数据分为两类: 有标签数据 (labeled data, $(x, y) \sim p_{\text{data}}(x_l, y_l) \cup (x, c) \sim G(z, c)$)

图 5 分类器性能与 Shannon 熵

Figure 5 Performance of classifier and Shannon entropy

和无标签数据 (unlabeled data, $x \sim p_{\text{data}}(x_{ul})$), 其中有标签数据由真实数据 (real data) 中的有标签数据 $(x, y) \sim p_{\text{data}}(x_l, y_l)$ 和 G 生成的数据 $(x, c) \sim G(z, c)$ 共同组成, 无标签数据 $x \sim p_{\text{data}}(x_{ul})$ 全部来自真实数据中无标签的数据样本。

在判别器 D_2 中, 对于有标签的数据使用交叉熵 (cross entropy, CE) 作为损失函数, 对于无标签数据, 如式 (4) 所示, 使用 Shannon 熵 (Shannon entropy, SE) 作为损失函数, 其中 c 为分类器类别个数。

$$\begin{aligned}
 l_{ul} &= E_{x \sim p_{\text{data}}(x_{ul})} S[D_2(i|x)] \\
 &= \frac{1}{N} \sum_{i=1}^N S[D_2(i|x)] \\
 &= \frac{1}{N} \sum_{i=1}^N \left[- \sum_{i=0}^c D_2(i|x) \log D_2(i|x) \right].
 \end{aligned} \quad (4)$$

如图 5 所示, 分类器 1 (classifier 1) 对于输入数字 9 的分类结果: $p_i = p(y = i|x)$, $i = 0, 1, 2, \dots, 9$ 均为 0.1, 其分类输出 Shannon 熵为 2.30. 而分类器 2 (classifier 2) 对于输入数字 9 的分类结果: $p_i = p(y = i|x)$, $i = 0, 1, 2, \dots, 9$, 其中 $\max p_i = p_9 = 0.5$, 其分类输出 Shannon 熵降为 1.22. 在理想状况下分类器输出应为某一类为 1.0 其余均为 0, 即该分类器对于输入数据的判断确定度很高, 此时该分类器各类别概率输出的 Shannon 熵为 0. 故使用 Shannon 熵作为无标签数据的损失函数, 以衡量分类器对于无标签数据分类的不确定度。

D_1 和 G 使用式 (5) 价值函数进行二人极大极小化博弈:

$$\begin{aligned}
 l_{d_1} &= \min_G \max_{D_1} V(D_1, G) \\
 &= E_{x \sim p_{\text{data}}(x)} [\log D_1(x)] + E_{z \sim p_z(z), c \sim p_c(c)} [\log(1 - D_1(G(z, c)))].
 \end{aligned} \quad (5)$$

D_2 通过无标签数据和有标签数据进行训练, 同时监督 G 使其生成与输入类别 c 对应的数据. 其价值函数如式 (6) 所示:

$$\begin{aligned}
 l_{d_2} &= \max_{G, D_2} V(D_2, G) \\
 &= E_{(x, y) \sim p_{\text{data}}(x_l, y_l)} [\log D_2(y|x)] + E_{z \sim p_z(z), c \sim p_c(c)} [\log(D_2(c|G(z, c)))] - l_{ul}.
 \end{aligned} \quad (6)$$

最终, DDGANs 目标函数为

$$\min_{G, D_2} \max_{D_1} V(D_1, D_2, G) = l_{d_1} - l_{d_2}. \quad (7)$$

2.2.2 DDGANs 理论证明

定理1 对于给定的和 D_2 和 G , 最优 D_1 为

$$D_1^*(x) = \frac{p_{\text{data}}(x)}{p_{\text{data}}(x) + p_g(x)}. \quad (8)$$

证明 D_1 的训练标准为: 对于给定的 D_2 和 G , D_1 需要最大化 $V(D_1, D_2, G)$,

$$V(D_1, D_2, G) = \int_x p_{\text{data}}(x) \log(D_1(x)) + p_g(x) \log(1 - D_1(x)) dx - H(D_2, G), \quad (9)$$

其中,

$$H(D_2, G) = \iint p_{\text{data}}(x, y) \log(D_2(y|x)) + p_g(x, y) \log(D_2(y|x)) dx dy - l_{ul}. \quad (10)$$

对于 $\forall(a, b) \in (0, 1)$, 函数 $y \rightarrow a \log(y) + b \log(1 - y)$ 在 $[0, 1]$ 区间内, 当 $y = \frac{a}{a+b}$ 取得最大值, 故当 $D_1(x) = \frac{p_{\text{data}}(x)}{p_{\text{data}}(x) + p_g(x)}$ 时, 式 (9) 中 $V(D_1, D_2, G)$ 取得最大值.

定理2 对于给定的 D_1 和 G , 最优 D_2 为

$$D_2^*(y|x) = 1. \quad (11)$$

证明 D_2 的训练标准为: 对于给定的 D_1 和 G , D_2 需要最小化 $V(D_1, D_2, G)$,

$$V(D_1, D_2, G) = I(D_1, G) - \iint p_{\text{data}}(x, y) \log(D_2(y|x)) + p_g(x, y) \log(D_2(y|x)) dx dy + l_{ul}, \quad (12)$$

其中,

$$I(D_1, G) = \int_x p_{\text{data}}(x) \log(D_1(x)) + p_g(x) \log(1 - D_1(x)) dx. \quad (13)$$

对于 $\forall(a, b) \in (0, 1)$, 函数 $f(x) = a \log(x) + b \log(1 - x)$, $x \in [0, 1]$, 当 $x = 1$ 时函数取得最大值 0, l_{ul} 作为利用 Shannon 熵进一步优化 D_2 的损失函数, 其优化目标为 D_2 对于样本分类的确定性, 故当 $D_2(y|x) = 1$ 时, l_{ul} 取得最小值 0, 式 (12) 中 $V(D_1, D_2, G)$ 取得最小值.

对于任意的 G , 将定理 1 和 2 中结果带入式 (7) 得

$$\begin{aligned} C(G) &= V(D_1^*, D_2^*, G) \\ &= \mathbb{E}_{x \sim p_{\text{data}}(x)} \log \frac{p_{\text{data}}(x)}{p_{\text{data}}(x) + p_g(x)} + \mathbb{E}_{x \sim p_g(x)} \log \frac{p_g(x)}{p_{\text{data}}(x) + p_g(x)} - 0. \end{aligned} \quad (14)$$

由 2.1.2 小节中引理 2 可得如下定理.

定理3 当且仅当 $p_g = p_{\text{data}}$ 时, 从 $C(G)$ 达到全局最小.

通过以上定理可得: DDGANs 中, D_1 的优化目标为 $D_1^*(x) = \frac{p_{\text{data}}(x)}{p_{\text{data}}(x) + p_g(x)}$, G 的优化目标为 $p_g = p_{\text{data}}$, D_2 的优化目标为 $D_2^*(y|x) = 1$. 由此可见 DDGANs 中在 D_1 和 G 进行二人极大极小化博弈的对抗环节, D_1 和 G 优化目标与 GANs 一致, 而 D_2 的加入对于 D_1 和 G 对抗并达到纳什均衡并没有产生影响, 同时在后续的实验中 DDGANs 收敛速度比 GANs 更快, 也证明了 DDGANs 相比 GANs 更容易达到纳什均衡.

表 1 判别器网络结构

Table 1 The structure of the discriminator network

Discriminator in ACGAN	D_2 in DDGANs
Input image	Input image
5×5 convolutional layer 32 lReLU	5×5 convolutional layer 32 lReLU
3×3 max-pool, stride 2	3×3 max-pool, stride 2
3×3 convolutional layer 64 lReLU	3×3 convolutional layer 64 lReLU
3×3 convolutional layer 64 lReLU	3×3 convolutional layer 64 lReLU
3×3 max-pool, stride 2	3×3 max-pool, stride 2
3×3 convolutional layer 128 lReLU	3×3 convolutional layer 128 lReLU
1×1 convolutional layer 10 lReLU	1×1 convolutional layer 10 lReLU
flatten	flatten
128 fc ELU with l2(0.01) re	128 fc ELU with l2(0.01) re
10 fc softmax and 1 fc sigmoid	10 fc softmax

2.2.3 DDGANs 训练过程

DDGANs 训练过程伪代码如下:

算法 1 DDGANs 的 Minibatch 随机梯度下降训练过程

- 1: **for** number of training iterations **do**
- 2: Sample minibatch of m noise samples $\{z^{(1)}, \dots, z^{(m)}\}$ from noise prior $p_z(z)$.
- 3: Sample minibatch of m category samples $\{c^{(1)}, \dots, c^{(m)}\}$ from noise prior $p_c(c)$.
- 4: Sample minibatch of m unlabeled examples $\{x_{ul}^{(1)}, \dots, x_{ul}^{(m)}\}$ from data generation distribution $p_{\text{data}}(x_{ul})$.
- 5: Sample minibatch of m labeled examples $\{(x_l^{(1)}, y^{(1)}), \dots, (x_l^{(m)}, y^{(m)})\}$ from data generation distribution $p_{\text{data}}(x_l)$.

- 6: Update D_1 by ascending its stochastic gradient:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m [\log D_1(x_{ul}^{(i)}) + \log(1 - D_1(G(c^{(i)}, z^{(i)})))] \quad (15)$$

- 8: Update D_2 by ascending its stochastic gradient:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m \left[\log(D_2(y^{(i)} | x_l^{(i)})) + \log(D_2(c^{(i)} | G(c^{(i)}, z^{(i)}))) + \left(- \sum_{j=0}^9 D_2(j | x_{ul}^{(i)}) \times \log D_2(j | x_{ul}^{(i)}) \right) \right] \quad (16)$$

- 10: Update G by descending its stochastic gradient:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m [\log(1 - D_1(G(c^{(i)}, z^{(i)}))) - \log(D_2(c^{(i)} | G(c^{(i)}, z^{(i)})))] \quad (17)$$

- 12: **end for**

2.2.2 小节中证明了 D_2 的优化目标, 同时得出 DDGANs 可以在增加一个判别器的情况下达到纳什均衡. 以 ACGAN 为参照, 通过实验对比 ACGAN 和 DDGANs 在分类以及生成性能的差异, 并进一步说明引入 D_2 的意义和优点.

使用表 1 所示结构构造 ACGAN 中判别器 D 以及 DDGANs 中 D_2 , 使它们具有相同的网络结构.

图 6 DDGANs 与 ACGAN 有监督分类准确度对比

Figure 6 Accuracy of supervised classification of DDGANs and ACGAN

图 7 DDGANs 与 ACGAN 生成图像对比

Figure 7 Generated images by DDGANs and ACGAN

由于 ACGAN 不具备半监督学习能力, 本文实验使用 MNIST 数据集, 对两个模型进行有监督分类实验. 在每个周期 (epoch) 训练完成后使用 MNIST 标准测试集 (10000 张图像) 进行测试, 其准确度如图 6 所示.

图 6 中, 在第 5 个周期后 ACGAN 判别器出现轻微的过拟合现象, 而 DDGANs 中 D_2 准确度保持稳定没有出现过拟合. 每个周期训练完成后生成器生成的图像对比如图 7 所示, 相比 ACGAN, DDGANs 生成的手写数字图像更加清晰真实, 如在第 7 个循环中, ACGAN 生成的图像质量普遍不如 DDGANs, 其中数字 6 的生成效果尤其差, 同样的情形在第 9 个循环数字 8 的生成中亦非常明显.

在实验中, 我们尝试使用深度更深的卷积神经网络作为判别器构造 ACGAN 时, 发生了生成器崩溃的问题, 这是由于在 GANs 训练中, 生成器以及判别器的结构和参数对于模型收敛情况具有非常显著的影响. 在 DDGANs 中, 由于 D_2 网络不参与对抗, 故其参数及结构不影响对抗网络的对抗训练过程. 这样, 就可以根据任务需求, 更加自由地选择适合的网络模型以及参数, 同时做到生成器的性能不受影响. 在下面的实验分析中, 构造了一个层数更深的卷积神经网络作为 D_2 , 在有限的有标签数据情况下达到了非常高的分类准确度.

图 8 鸟巢位置示意图

Figure 8 The location of the nest on catenary

图 9 使用 Sobel 算子对图像进行边缘化处理

Figure 9 Edged image by Sobel operator

3 实验分析

3.1 接触网鸟巢检测实验分析

3.1.1 接触网鸟巢检测图像预处理

2012 年, 原铁道部运输局供电部、中国铁道科学研究院和西南交通大学共同制定了《高速铁路供电安全检测监测系统 (6C 系统) 总体技术规范》, 旨在对高速铁路的牵引供电系统进行全方位、全覆盖的综合检测监测. 在 6C 系统中, 大量采用了视频图像方式进行检测监测. 随之, 对视频图像的自动分析受到业界的高度关注.

本文采用的接触网鸟巢视频图像数据来自于车载设备拍摄的京广高铁接触网支柱图像. 摄像机固定在列车驾驶室的前向位置. 接触网支柱在接触网视频中出现的位置相对固定, 同时通过分析接触网设备的筑巢位置, 可以发现几乎绝大多数鸟巢只会搭建在接触网支柱的几个固定位置 (如图 8). 对于接触网鸟巢检测使用下文所述分割方法并进行分类检测取得了较满意的效果, 对于其他位置不确定性较高的故障, 可以将 R-CNN 算法提取候选区域的方法与 DDGANs 图像分类结合进行不平衡数据的目标检测及分类.

如图 8 所示, 基于接触网支柱特征在巡检视频中提取出接触网支柱部分, 将其等份分割为 25 张 40×40 的小图, 将这些小图中含有鸟巢的图像 (方框所圈) 标记为 1, 其余不含鸟巢的图像标记为 0, 得到训练数据集.

由于拍摄设备白平衡、曝光等参数及拍摄时不同天气状况的影响, 造成数据图像曝光程度以及图像质量不一致. 本文对原始图像使用 Sobel 算子对图像进行了边缘化处理, 以去除背景明暗影响, 突出图像内容, 如图 9 所示.

图 10 两种 GANs 模型初步验证结果比较

Figure 10 Comparison of preliminary results of two GANs models

图 11 接触网图像预测方法示意图

Figure 11 Prediction method of the catenary images

3.1.2 接触网鸟巢检测的训练及测试数据

本文使用的接触网鸟巢检测数据包含鸟巢的数据 (正样本) 为 500 张; 正常工况数据 (负样本) 数量为 3000 张. 为了更加有效地利用数量不多的正样本, 通过拉伸、旋转、截取图像的方法, 将 500 张正样本图像增加为 1000 张.

训练集使用数据为: 正样本 500 张, 负样本 2000 张. 测试集使用数据为: 正样本 500 张, 负样本 1000 张.

3.1.3 接触网鸟巢检测实验结果分析

相比使用普通图像分类方法对接触网鸟巢进行检测, DDGANs 通过生成器生成对应类别数据, 以补偿不平衡数据中特定类别数据不足的问题, 使得模型能够更为有效地学习到数据特征.

使用 3.1.1 小节中方法, 对测试集进行分割处理, 均分为 25 张, 每张为 40×40 像素的小图, 得到 500×25 张正样本小图数据, 1000×25 张负样本小图数据. 使用这些小图数据来初步验证 DDGANs 的性能, 同时与此前直接使用 GANs 网络进行分类进行比较 (图 10). 可见, 使用 DDGANs 针对样本数据进行分类的性能大大提高.

在初步验证模型性能后, 本文使用如下方法训练及预测模型.

如图 11, 将训练数据中接触网支柱图像等份分割为 25 张有标记的小图, 每张小图作为单独数据训练 DDGANs, 再将整张图片 5×5 的标签矩阵作为输入, 训练一个用于判断整体图像是否为正的全连接神经网络. 同样, 将待预测的图像等份分割为 25 张小图, 输入到训练好的模型中, 得到 5×5 的预测矩阵. 将该预测矩阵输入全连接神经网络得到结果.

		Real category	
		1	0
Prediction category	1	True positive TP	False positive FP
	0	False negative FN	True negative TN

图 12 混淆矩阵
Figure 12 Confusion matrix

表 2 接触网鸟巢检测实验结果
Table 2 Test results of detecting nests in catenary

Algorithm	TP	FP	FN	TN	Recall rate (%)
SIFT	341	322	159	678	68.2
SURF	347	358	153	642	69.4
CNN	414	157	86	843	82.8
KEFEH	469	73	31	927	93.9
DDGANs	490	85	10	915	98.0

测试实验使用 3.1.2 小节中所述测试数据集, 500 张负样本 (正常工况) 的图像数据以及 1000 张正样本 (含有鸟巢) 的图像数据对 DDGANs 进行测试. 基于图 12 中的混淆矩阵, 同时考虑到接触网鸟巢检测在工程应用中的情形, 少量的 FP (false positive) 对实际应用效果影响不大, 而当 FN (false negative) 较高时整个系统的应用意义将不再显著, 故我们使用召回率 R 来评判模型对于接触网鸟巢检测的性能.

$$R = \frac{TP}{TP + FN}. \tag{18}$$

我们使用以下 4 种方法比较评估本文模型在接触网鸟巢检测中的性能, 将其与 SIFT 特征匹配方法^[13]、SURF 特征匹配方法^[14]、CNN^[15]以及关键区域 HOG 特征提取 (key area feature extraction with HOG, KEFEH)^[16]进行对比, 结果见表 2.

3.2 采用 MNIST 数据集的半监督学习实验分析

本文 DDGANs 模型不仅可用于有监督学习, 对半监督学习的图像分类也获得了理想的性能. 半监督图像分类是计算机视觉中一个重要的研究方向, 近年来在半监督学习图像分类中涌现出了许多优秀的算法模型. MNIST 数据集作为机器学习、计算机视觉、人工智能、深度学习等领域研究的算法衡量基准数据集, 在很多相关研究中发现其身影, 如面向半监督学习的深度生成模型 DGN (deep generative models)^[17]以及同样基于生成对抗网络的半监督学习方法 CatGAN (categorical generative adversarial networks)^[18]等.

MNIST 数据集 (如图 13 所示) 是由 Yann LeCun 等在由美国国家标准与技术研究院 (National Institute of Standards and Technology) 提供的手写数据集的基础上筛选并进行了尺寸标准化及数字中

图 13 MNIST 数据示意图
Figure 13 Example of MNIST data

图 14 半监督学习性能比较
Figure 14 Comparison of performance of semi-supervised learning

表 3 基于 MINST 数据集进行半监督学习分类结果
Table 3 Classification results of semi-supervised learning on MNIST dataset

N_L	TSVM	CNN	AtlasRBF	VA	DGN	CatGAN	DDGANs
100	83.19	87.13	91.90	97.88	96.67	98.09	98.12
1000	93.84	94.67	96.32	98.68	97.60	98.27	99.26

心化等处理的标准数据集. 总共包含 70000 张有标签的手写数字图像 (每张图像尺寸为 28×28 像素), 其中 60000 张为训练样本, 10000 张为测试样本.

本文实验, 在 60000 张训练数据中随机提取了 100 张有标记的数据集, 再从其余 59900 张数据中随机提取 50000 张作为无标签数据. 在每个训练周期 (epoch) 结束后, 使用 MNIST 数据集标准测试数据进行测试. 与在图像分类中应用较多的深度学习模型中的 CNN 进行比较分析, 随着训练进行的模型分类准确率如图 14 所示.

针对 MNIST 数据集, 本文还将 DDGANs 模型与下列 6 种半监督学习模型进行性能对比, 结果如表 3 所示.

- (1) TSVM (transductive SVMs): 基于 SVM, 以最大边缘分类为目标, 最少化接近决策边界的无标签样本 [19];
- (2) CNN (卷积神经网络) [15]: 该 CNN 使用了 Shannon 熵 ((4) 式) 作为损失函数对无标签数据进行训练;
- (3) AtlasRBF: 基于流形的半监督学习方法 [20];
- (4) VA (virtual adversarial): 通过将在每个训练数据点周围的预测分布的平滑度纳入目标函数改进深度神经网络的方法 [21];
- (5) DGN (deep generative models) [17]: 深度生成模型;
- (6) CatGAN (categorical generative adversarial networks) [18]: 基于生成对抗网络的半监督学习模型.

表 3 中除 CNN 及本方法外, 其余方法均使用原文献中实验结果.

CNN 及本方法使用如下方法得出准确率结果: 在表 3 中, N_L 为有标签数据数量, 无标签数据均为 50000. 总共进行 10 次独立实验, 每次训练数据 (含有标签和无标签) 均从 MINST 训练数据中随机提取, 算法性能采用准确率 A 来衡量 (式 (19)), 其中 N_C 预测正确的数量, N_T 为测试集总样本数

图 15 使用 MNIST 数据集生成的图像

Figure 15 Generated images trained by MNIST dataset. (a) Generated image by DCGAN at epoch 2; (b) generated image by DDGANs at epoch 2

(10000 张). 实验结果为 10 次实验结果的均值.

$$A = \frac{N_C}{N_T}. \quad (19)$$

在使用 MNIST 进行训练并且同样网络结构及参数设置的情况下, 相比普通 GANs, DDGANs 的生成网络 G 收敛更快, 如图 15 为第 2 个循环周期中 DDGANs 以及 DCGANs^[11] 生成的图像. 图中 DCGANs 还未收敛, 本实验中在第 21 个循环周期后, DCGANs 才开始收敛, 生成有意义的图像.

4 结论

本文利用 GANs 较强的数据特征分布表达能力, 针对分类任务构造了一种新的 GANs 网络 DDGANs, 在接触网鸟巢异常检测中取得了良好的检测效果, 在 MINST 标准数据集上的实验结果进一步证明了本模型在不平衡数据和半监督学习上相较于其他 6 种方法拥有更好的性能.

GANs 网络仍存在一定的缺陷^[22], 如何改进 GANs 模型使其数据特征表达能力更强, 同时结合本文中针对分类任务的改进思路, 达到更好的分类和生成效果, 还值得深入研究.

参考文献

- Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. In: Proceedings of International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2014. 2672–2680
- Dai Z H, Yang Z L, Yang F, et al. Good semi-supervised learning that requires a bad gan. 2017. ArXiv: 1705.09783
- Shrivastava A, Pfister T, Tuzel O, et al. Learning from simulated and unsupervised images through adversarial training. 2017. ArXiv: 1612.07828
- Li J, Monroe W, Shi T, et al. Adversarial learning for neural dialogue generation. 2017. ArXiv: 1701.06547
- Reed S, Akata Z, Yan X, et al. Generative adversarial text to image synthesis. 2016. ArXiv: 1605.05396
- Mathieu M, Couprie C, LeCun Y. Deep multi-scale video prediction beyond mean square error. 2015. ArXiv: 1511.05440
- Huang R, Zhang S, Li T, et al. Beyond face rotation: global and local perception gan for photorealistic and identity preserving frontal view synthesis. 2017. ArXiv: 1704.04086
- Salimans T, Goodfellow I, Zaremba W, et al. Improved techniques for training gans. 2016. ArXiv: 1606.03498
- Mirza M, Osindero S. Conditional generative adversarial nets. 2014. ArXiv: 1411.1784
- Odena A, Olah C, Shlens J. Conditional image synthesis with auxiliary classifier GANs. 2016. ArXiv: 1610.09585
- Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. 2015. ArXiv: 1511.06434

- 12 Liu J W, Liu Y, Luo X L. Semi-supervised learning methods. *Chin J Comput*, 2015, 38: 1592–1617 [刘建伟, 刘媛, 罗雄麟. 半监督学习方法. *计算机学报*, 2015, 38: 1592–1617]
- 13 蔡学敬. 基于图像处理技术的高速接触网动态检测系统研究. 博士学位论文. 成都: 西南交通大学, 2008
- 14 Lin P, Wang L, Luo L. High-speed locomotive porcelain positioning method based on SIFT. *Electric Drive Locomot*, 2012, 5: 87–89 [林萍, 王黎, 罗林. 基于 SIFT 的高速列车车顶瓷瓶识别方法. *机车电传动*, 2012, 5: 87–89]
- 15 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, 2012. 1097–1105
- 16 段旺旺, 唐鹏, 金炜东, 等. 基于关键区域 HOG 特征的铁路接触网鸟巢检测. *中国铁路*, 2015, 8: 73–77
- 17 Kingma D P, Mohamed S, Rezende D J, et al. Semi-supervised learning with deep generative models. In: *Proceedings of Annual Conference on Neural Information Processing Systems*, Montreal, 2014. 3581–3589
- 18 Springenberg J T. Unsupervised and semi-supervised learning with categorical generative adversarial networks. 2015. ArXiv: 1511.06390
- 19 Joachims T. Transductive inference for text classification using support vector machines. In: *Proceedings of the 16th International Conference on Machine Learning*. San Francisco: Morgan Kaufmann Publishers Inc., 1999. 99: 200–209
- 20 Pitelis N, Russell C, Agapito L. Semi-supervised learning using an unsupervised atlas. In: *Proceedings of Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Berlin: Springer, 2014. 565–580
- 21 Miyato T, Maeda S, Koyama M, et al. Distributional smoothing by virtual adversarial examples. 2015. ArXiv: 1507.00677
- 22 Wang K F, Gou C, Duan Y J, et al. Generative adversarial networks: the state of the art and beyond. *Acta Autom Sin*, 2017, 43: 321–332 [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. *自动化学报*, 2017, 43: 321–332]

Double discriminator generative adversarial networks and their application in detecting nests built in catenary and semisupervised learning

Weidong JIN*, Pei YANG & Peng TANG

School of Electrical Engineering, Southwest Jiaotong University, Chengdu 611731, China

* Corresponding author. E-mail: wdjin@home.swjtu.edu.cn

Abstract In image-based detection of catenary anomalies, detection of bird's nest anomalies is a typical situation. However, the image data containing the nests is only a small portion of the total data, which makes nest detection a typical problem of imbalanced data classification. For using a machine learning algorithm to solve imbalanced data classification, the learning ability of data features is of much importance. Generative adversarial networks (GANs) can learn prosperous data features from unlabeled data, which has been widely confirmed and applied. Nonetheless, because of the limitation of GANs' structure and theory, it is not an ideal model for image classification. In this research, the GANs model is improved with respect to image classification tasks. The improved model is named "double discriminator generative adversarial networks" (DDGANs). With DDGANs, the classification results of nest detection are satisfactory, and it is also an effective semisupervised learning model. Experiments on the MNIST standard dataset show that the accuracy and convergence rate have been significantly improved compared with other models.

Keywords GANs, semisupervised learning, unbalanced data classification, anomaly detection, catenary

Weidong JIN was born in 1959. He received his Ph.D. degree in electrification and automation for railway traction at Southwest Jiaotong University, Chengdu, Sichuan Province, China, in 1998. He is presently a professor at the School of Electrical Engineering of Southwest Jiaotong University. His current research interests include intelligent information processing, pattern recognition, and system simulation and optimization.

Pei YANG was born in 1989. He is currently studying for his M.S. degree at the School of Electrical Engineering at Southwest Jiaotong University. His main research interests include deep learning and pattern recognition.

Peng TANG was born in 1979. He received a bachelor's degree in automation from Chongqing University, Chongqing, China, in 2002, and a Ph.D. degree in computer application and technology with an emphasis on imaging and vision at Sichuan University, Chengdu, Sichuan Province, China, in 2010. Currently, he is a lecturer at Southwest Jiaotong University, China. His current research interests include the areas of computer vision and machine learning, with contributions in robotic vision, medical image processing, and video understanding.