

基于WGAN-GP的建筑垃圾数据集的优化与扩充

邬欣诺

同济大学，电子与信息工程学院，上海

收稿日期：2022年12月28日；录用日期：2023年1月24日；发布日期：2023年1月31日

摘要

随着人工智能技术的火爆与不断成熟，人们越来越倾向于用神经网络的方式去解决现有的问题。同时随着城市化的推进，建筑的拆毁与重建过程中产生了大量的建筑垃圾。现有的建筑垃圾回收装置回收工序复杂，效率低下，导致回收成本远远高于资源本身价值，因此探索高效率，低成本的建筑垃圾回收设备尤为重要。然而现在几乎不存在针对建筑垃圾的专有数据集，且大多数数据质量都不高，导致神经网络无法应用于建筑垃圾领域。为了解决上述问题，提高神经网络在建筑垃圾方面的应用，本文提出了一个新的建筑垃圾数据集，通过手工采集与清理的方式形成初始数据集，在通过WGAN-GP模型对其完成高质量的样本扩充，从而填补建筑垃圾数据集的空白。

关键词

数据增强，生成对抗网络，建筑垃圾数据集

Optimization and Expansion of Construction Waste Dataset Based on WGAN-GP

Xinnuo Wu

School of Electronics and Information Engineering, Tongji University, Shanghai

Received: Dec. 28th, 2022; accepted: Jan. 24th, 2023; published: Jan. 31st, 2023

Abstract

With the explosion and continuous maturity of artificial intelligence technology, people are increasingly inclined to use neural networks to solve existing problems. Meanwhile, with the advancement of urbanization, the process of building demolition and reconstruction has produced a

large amount of construction waste. The existing construction waste recycling devices have complicated recycling processes and low efficiency, resulting in recycling costs much higher than the value of the resources themselves, so it is important to explore high-efficiency and low-cost construction waste recycling devices. However, there are almost no proprietary data sets for construction waste, and most of the data are of low quality, which makes it impossible to apply neural networks to the construction waste field. In order to solve the above problems and improve the application of neural networks in construction waste, this paper proposes a new construction waste dataset, the initial dataset is formed by manual collection and cleaning, and then it is expanded with high quality samples by WGAN-GP model to fill the gap of construction waste dataset.

Keywords

Data Augmentation, Generative Adversarial Network, Construction Waste Dataset

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着互联网技术的不断发展[1], 虚拟现实技术也逐渐走入人们的生活, 图像生成任务也受到越来越多的人关注, 生成对抗网络(GANs)凭借其出色的表现革新了图像生成领域, 也逐渐的进入工业界的视野。虽然生成对抗网络在图像生成任务中取得了巨大的成功, 但在训练期间的不稳定性也一直被人诟病。这种不稳定性一个被普遍接受的原因是当真实分布与虚假分布没有足够的重叠时, 从判别器传递到生成器的梯度会变得没有信息。

当前, 加强环境保护已经成为国家的基本政策[2]。而保护好环境, 一个很重要的因素就是要把人类产生的垃圾处理好。很多地方已经开始了垃圾分类来处理人类产生的生活垃圾, 取得了很大的进步[3]。但在垃圾当中, 生活垃圾仅仅占据一小部分。事实上, 工业垃圾是对环境造成污染的更重要的因素。因此如何处理好工业垃圾成为一个紧迫的课题[4]。在工业垃圾当中, 建筑垃圾则有着独特的地位。这是因为一方面我国是世界基础建设大国, 每年产生了大量的建筑垃圾, 另一方面, 建筑垃圾里面还包含着各种有用的资源, 能够回收和再利用[5]。因此, 有必要建立起良好的建筑垃圾处理能力。智能化垃圾分类系统能够加速绿色环保的垃圾处理过程, 对绿色都市和智能化城市管理都有着重大意义。然而, 当前的建筑垃圾样本集是十分稀缺的, 且现有的数据集未必能够适用于真实的工厂环境以及垃圾的分拣情况。

为了解决上述问题, 越来越多的研究者提出了更多的网络结构企图让模型训练更稳定[6] [7]。确实有一些工作通过修改网络结构让生成对抗网络变的更稳定, 修改起来也简单, 但这样的方式终究也缺乏一些可解释性, 没有让损失函数变得有意义, 还是不能够指示训练过程, 并未明确指出生成对抗网络训练不稳定的原因以及解决方法, 且目前的工作还是没能解决 GAN 对超参数敏感的问题, 无法做到能够持续稳定的保持判别器与生成器之间的平衡。目前 GAN 仍然面临着两个问题: 模式崩塌与训练不稳定。当生成器网络只能捕捉到数据分布中一个子集的方差时, 就会出现模式崩塌的问题。在我们这项工作中, 主要是来解决训练不稳定的问题。

综上所述, 为了扩充建筑垃圾的样本数量以及多样性, 以及尽量保持训练过程的稳定以降低训练成本, 本文采用了 Wasserstein GAN (WGAN)这种网络结构来生成建筑垃圾。本文的主要贡献如下:

- 1) 我们收集了一个由建筑垃圾组成的新数据集, 该数据集主要包括木材、红砖、塑料、织物、泡沫、

混凝土等多种建筑垃圾,采用 2D 高速工业相机对现场物料进行实时检测获取位置、颜色以及纹理等信息;

2) 采用了 WGAN 去生成建筑垃圾样本, WGAN 抛弃了之前模型采用的 Kullback-Leibler 散度(KL 散度)以及 Jensen-Shannon 散度(JS 散度),采用 Wasserstein 距离衡量两个分布之间的距离,彻底解决了 GAN 以往的训练不稳定的问题;

3) 由于直接求解 Wasserstein 距离是很难做到的,采用 Lipschitz 约束函数的最大上界,从而转化为更节约计算代价的问题,然后通过神经网络的方式去优化该问题。

2. 相关工作

2.1. 生成对抗网络

GAN [8]是“Generative Adversarial Networks”的缩写,由 Ian Goodfellow 在 2014 年首次提出。GAN 是非监督式学习的一种方法,是深度学习领域的一个重要模型。生成对抗网络由一个生成网络与一个判别网络组成。生成网络从潜在空间中随机采样作为输入,其输出结果需要尽量模仿训练集中的真实样本。判别网络的输入则为真实样本或生成网络的输出,其目的是将生成网络的输出从真实样本中尽可能分辨出来。而生成网络则要尽可能地去欺骗判别网络。两个网络相互对抗、不断调整参数,最终目的是使判别网络无法判断生成网络的输出结果是否真实。这种对抗方式避免了一些传统生成模型在实际应用中的一些困难,巧妙地通过对抗学习来近似一些不可解的损失函数,在图像、视频、自然语言和音乐等数据的生成方面有着极为广泛的应用。

2.2. 积分概率度量

关于积分概率度量现在已经有很多相关工作了,不同类别的函数可以得到完全不同的度量结果。通常情况下都由 KL 散度、JS 散度去衡量。KL 散度和 JS 散度是突变的,要么最大要么最小,要用梯度下降法优化这个参数时,两者很难提供梯度,高维空间中如果两个分布不重叠或者重叠部分可忽略,则 KL 散度和 JS 散度更反映不了远近。Energy-based GANs (EBGANs)可以被认为是衡量虚假分布与真实分布之间的总体距离的生成方法,采用 Total-Variation 距离(TV 距离)。判别器将最大化这个距离,而它的唯一限制就是处于一个特定常数下。因此,当判别器到达最优时,生成器的成本将接近虚假分布与真实分布之间的总体距离。由于 EBGANs 衡量的变化距离与 JS 散度显示出相同的规律性,不难看出,EBGANs 将遭受与经典 GANs 同样的问题,即无法训练判别器直到最佳状态,从而将其限制在非常不好的梯度上。

2.3. 建筑垃圾数据集

对于机器学习来说,数据的重要性是不言而喻的[9] [10]。而在当前的情况下,几乎不存在专门针对建筑垃圾的数据集,仅有的一些垃圾的数据质量也是参差不齐,完全达不到神经网络的训练标准。这也是导致国内建筑垃圾处理智能化水平低、国内市场中尚未有国产化的建筑垃圾智能分选机器人设备的主要原因。

3. 模型介绍

3.1. 模型整体框架

本文采用的模型框架是 Wasserstein GAN (WGAN) [11], WGAN 提出了一种新的衡量距离的方法,采用 Wasserstein 距离(又叫 Earth-Mover 距离)来衡量真实数据与生成数据分布之间的距离,并通过约束参数的方法来解决 GAN 训练不稳定的问题。在模型中,首先让生成器从隐空间的随机噪声中生成图片,然

后再将真实数据与生成数据依次送入判别器中，然后再让判别器尽可能的区分两者。在损失函数中，引入了拥有更为平滑的特性的 Wasserstein 距离，构建了一种更为松散的拓扑结构，使分布序列更容易收敛，从而解决梯度消失的问题。

- The Earth-Mover (EM) 距离(Wasserstein-1)

$$W(\mathbb{P}_r, \mathbb{P}_g) = \inf_{\gamma \in \Pi(\mathbb{P}_r, \mathbb{P}_g)} \mathbb{E}_{(x,y) \sim \gamma} [\|x - y\|], \tag{1}$$

其中 $\Pi(\mathbb{P}_r, \mathbb{P}_g)$ 表示所有联合分布 $\gamma(x,y)$ 的集合，他们的边界分别为 \mathbb{P}_r 和 \mathbb{P}_g 。不难看出，EM 距离表示的是将一个分布转化成另一个分布的最小成本。图 1 展示了模型的整体框架。

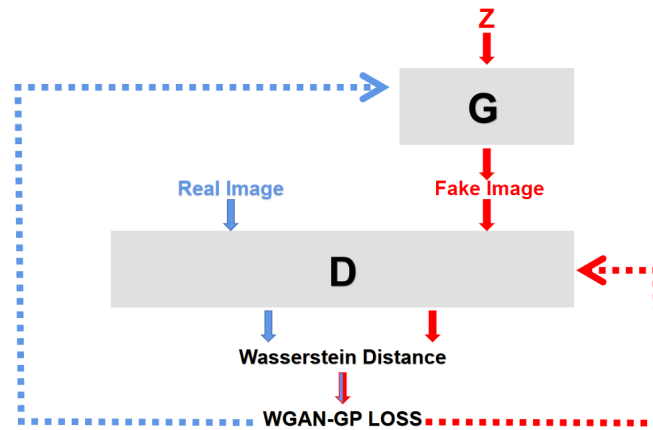


Figure 1. Overall framework of the model
图 1. 模型整体框架

3.2. Wasserstein 距离

我们可以通过对 EM 距离进行梯度下降来学习低维流形上的概率分布。其他距离和散度不能做到这一点，因为它们产生的损失函数甚至不是连续的。这个结论不止在真实分布与虚假分布之间是完全不相交的这种情况之中，当有一个非空的交点包含在一个度量为零的集合中，同样的结论也是成立的。当两个低维流形在一般位置相交时，就会出现这种情况。由于 Wasserstein 距离引发的拓扑结构比 JS 距离要弱很多，对于我们的问题，EM 是一个比至少 Jensen-Shannon 散度更合理的损失函数。EM 距离是连续的和可微的，这意味着我们可以训练判别器直到最优。我们越是训练判别器，我们得到的 Wasserstein 梯度就越可靠。而对于 JS 来说，因为它是局部饱和的，我们得到的梯度是消失的。在其中 KL 最强，其次是 JS 和 TV，EM 最弱。事实上，当学习低维流形支持的分布时，KL、JS 和 TV 距离不是明智的成本函数。然而，EM 距离在这种情况下是明智的。

3.3. Lipschitz 约束

由上一节的定理我们得出，EM 距离对于优化模型具有更好的特性。但观察其公式以及定义，不难看出，EM 距离的下界的求得方法是很难实现的。所以我们通过 Kantorovich-Rubinstein 的理论把问题转化为：

$$W(\mathbb{P}_r, \mathbb{P}_\theta) = \sup_{\|f\|_L \leq 1} \mathbb{E}_{x \sim \mathbb{P}_r} [f(x)] - \mathbb{E}_{x \sim \mathbb{P}_\theta} [f(x)] \tag{2}$$

sup 是指整个 1-Lipschitz 函数 f 的最上界。不难看出，如果我们将 1-Lipschitz 函数换成 k-Lipschitz

函数(这里的 k 指任意常数), 最终结束时收敛到 $K \cdot \mathcal{W}(\mathbb{P}_r, \mathbb{P}_g)$, 因此, 如果我们有一个参数化的函数族 $\{f_w\}_{w \in \mathcal{W}}$, 对于某个 K 都是 K -Lipschitz 函数约束, 我们可以考虑转成解决这个问题:

$$\max_{w \in \mathcal{W}} \mathbb{E}_{x \sim \mathbb{P}_r} [f_w(x)] - \mathbb{E}_{z \sim p(z)} [f_w(g_\theta(z))] \quad (3)$$

现在的问题变成了寻找能解决上述最大值问题的函数。为了大致上接近这个问题, 我们可以做的是训练一个神经网络, 其参数是位于紧凑空间 \mathcal{W} 中的权重, 然后通过 $\mathbb{E}_{z \sim p(z)} [f_w(g_\theta(z))]$ 进行反推, 正如我们对一个典型的 GAN 所做的那样。 \mathcal{W} 空间是紧凑的这一事实意味着所有的函数 f_w 都是符合在特定的常数 k 的取值下的 k -Lipschitz 约束, 且 f_w 只取决于 \mathcal{W} 这个空间而并不取决于某个单独的权重。为了使参数 w 位于一个紧凑的空间中, 我们可以在每次梯度更新后将权重夹在一个固定的区间范围内。权重剪裁显然是执行 Lipschitz 约束的一种最简单的方式。但是如果剪切参数很大, 那么任何权重都需要很长的时间才能达到极限, 从而使训练批判者到最佳状态变得更加困难。如果剪切参数较小, 当层数较多或未使用批量归一化时, 这很容易导致梯度消失, 使 WGAN 的判别器经常不能收敛。

3.4. WGAN-GP

在判别器中, 希望 loss 尽可能的大, 才能来尽可能区分真假样本, 这样会导致在判别器中通过 loss 计算梯度会沿着 loss 越来越大的方向变化, 然而经过截断后, 每个网络参数又被独立的限制了取值范围。这种结果使得所有参数走向极端, 要么取最大值, 要么取最小值, 判别器没能充分利用自身的模型能力, 经过它传给生成器的梯度也会跟着变差。WGAN-GP [12] 现在提出了另一种强制执行 Lipschitz 约束的方法。一个可微分函数是 1-Lipschitz, 当且仅当它的梯度在任何地方都是最多为 1 的, 所以考虑直接约束批评者的输出相对于其输入的梯度规范。为了规避可操作性问题, 我们对随机样本的梯度规范进行惩罚, 强制执行一个软约束版本。WGAN-GP 的新目标是

$$L = \underbrace{\mathbb{E}_{\tilde{x} \sim \mathbb{P}_g} [D(\tilde{x})] - \mathbb{E}_{x \sim \mathbb{P}_r} [D(x)]}_{\text{Original critic loss}} + \lambda \underbrace{\mathbb{E}_{\hat{x} \sim \mathbb{P}_{\hat{x}}} \left[\left(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1 \right)^2 \right]}_{\text{Our gradient penalty}}. \quad (4)$$

3.5. 训练过程

本文采用的模型将传统 GAN 的判别器的将真实图片与生成图片的分类任务转化成衡量真实数据与生成数据分布之间的距离问题。判别器的任务希望二者距离变大, 所以对 L 的前半部分取反。生成器希望生成结果的分布越来越靠近真实数据的分布, 所以希望通过训练让距离 L 最小化, 因为生成器与第一项无关, 所以生成器的 loss 可以简写为:

$$G(\text{loss}) = -D(G(x)) \quad (5)$$

本文采用 Adam 算法进行梯度下降训练, 在训练阶段采用我们自己收集的建筑垃圾数据集去训练。

4. 实验

4.1. 建筑垃圾数据集

针对目前没有质量高的垃圾数据集, 且目前专门适用于建筑垃圾的数据集几乎没有, 难以支撑基于机器学习方法的任务。我们去建筑工地实地手工采集样本, 大致分为织物、砖头、塑料、木材这四类, 将它们批量裁剪为 128×128 的格式。然后, 我们通过手动筛选清洗图片的方式来过滤低质量的、出现错误的和低光照的图片。最后我们获得了约 400 张的可用样本, 形成了我们的建筑垃圾数据集。

4.2. 实验设置

为了验证本模型的有效性，本文采用 Fréchet Inception Distance (FID)来衡量生成的图像质量，来作为主要的评价指标。

在本文方法中，为了保证生成质量，对每个种类垃圾单独使用 WGAN 做图像生成。一些基本的参数设置如下。噪声空间的维度为 100，训练 epoch 设置为 60,000，batch size 设置为 100，训练过程中采用 Adam 算法进行优化，采用激活函数 leaky relu 来增强模型的非线性表达能力。

4.3. 实验结果

本文采用 WGAN-GP 模型生成的结果如下图 2 所示：

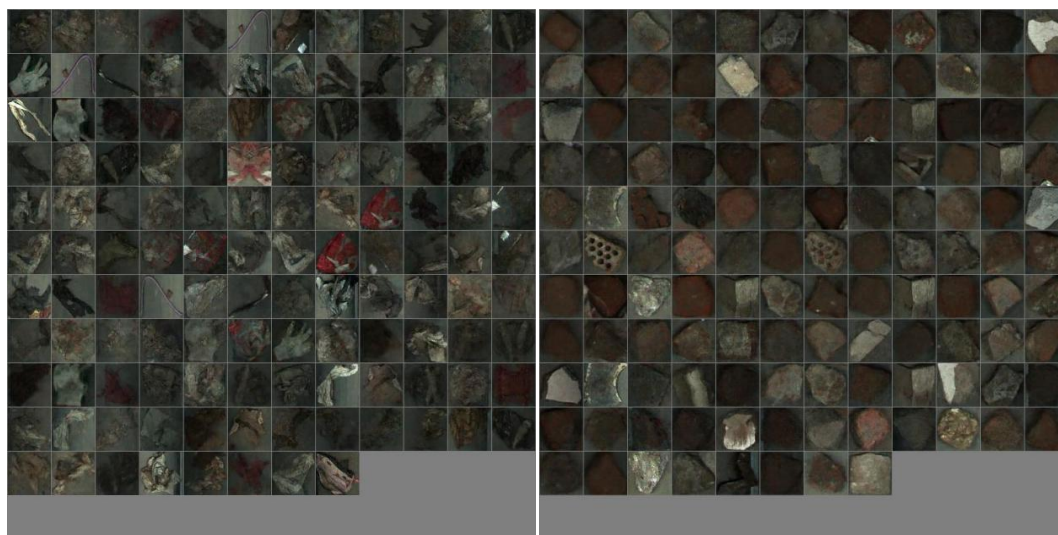


Figure 2. Sample images generated based on each garbage category
图 2. 基于每个垃圾类别生成的样本图像

生成的图像模拟了原始数据特征分布规律与颜色信息，并在此基础上增加了不同的纹理组合，融入了其他图像颜色或纹理上的特点，从而生成了多特征的样本图，保证样本相似性的同时增强了样本的多样性。

Table 1. Main experimental results
表 1. 主要实验结果

	DCGAN	WGAN	WGAN-GP
FID	70.2	35.6	24.1

从表 1 中可以看出，WGAN-GP 的性能在我们的建筑垃圾数据集上超越了其他的模型，是最优的。且收敛所需的 epoch 较少，模型更倾向于在更少的迭代中收敛。且实现简单，效果稳定，不需要根据具体的数据集来调整超参数，也不需要精心设计网络结构来保证 GAN 的稳定性，且生成的图片对比其他方法而言质量要高。综上，WGAN-GP 能够高质量、高效率、低成本的完成对现有的建筑垃圾数据集的扩充。

5. 总结

本文针对现有的垃圾数据集不存在专门针对建筑垃圾、以及图片质量低、噪声高的问题，无论从数

量和质量上都难采用神经网络的模型和方法对建筑垃圾这个领域开展一些如实例分割、降噪等任务[13][14][15]。为此我们提出了自己的建筑垃圾数据集，并采用了用 Lipschitz 约束去近似实现 Wasserstein 距离，以解决原始 GAN 的不稳定问题，并引入了梯度惩罚项代替直接截断，来进一步实现稳定，且在一定程度上规避了模式崩塌的问题的 WGAN-GP 模型，从而可以实现低成本、高效率的生成建筑垃圾样本，以达到数据增强的目的。实验结果表明，通过 WGAN-GP 模型对每一类数据单独进行生成的样本质量较高，且能对同类样本的不同纹理进行特征组合，保证样本相似性的同时增加样本的多样性，最终生成图像与原始数据共同形成最终的样本集，完成样本扩充。

参考文献

- [1] 韩九强, 胡怀中, 等. 机器视觉技术及其应用[M]. 北京: 高等教育出版社, 2009.
- [2] 杨凯. 建筑垃圾的现状分析及资源化广泛利用必要性[J]. 建筑工程技术与设计, 2017(21): 4544.
- [3] 李景茹, 朱姣兰, 蔡红, 等. 设计人员的建筑废弃物减量化行为影响因素分析[J]. 科技管理研究, 2014, 34(14): 183-187.
- [4] 高天霞, 王海兵. 工业化建筑及其废弃物减量化探讨[J]. 广州建筑, 2018, 46(1): 35-40.
- [5] 陈家珑. 我国建筑垃圾资源化利用现状与建议[J]. 建设科技, 2014(1): 8-12.
- [6] 牛斌, 吴鹏, 马利, 等. 一种基于生成对抗网络的行为样本集扩展方法[J]. 计算机技术与发展, 2019(7): 43-48.
- [7] Denton, E.L., Chintala, S., Szlam, A. and Fergus, R. (2015) Deep Generative Image Models Using a Laplacian Pyramid of Adversarial Networks.
- [8] Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., et al. (2014) Generative Adversarial Networks. *Advances in Neural Information Processing Systems*, **3**, 2672-2680.
- [9] 段峰, 王耀南, 雷晓峰, 等. 机器视觉技术及其应用综述[J]. 自动化博览, 2002, 19(3): 59-61.
- [10] 杨壮. 面向 Bin Picking 的虚拟样本集构建及智能识别方法的研究[D]: [硕士学位论文]. 上海: 华东理工大学, 2019.
- [11] Arjovsky, M., Chintala, S. and Bottou, L. (2017) Wasserstein Generative Adversarial Networks. *Proceedings of the 34th International Conference on Machine Learning*, Vol. 70, 214-223.
- [12] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V. and Courville, A. (2017) Improved Training of Wasserstein GANs. *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 5767-5777.
- [13] 章毓晋. 图像理解与计算机视觉[M]. 北京: 清华大学出版社, 2000.
- [14] 王志瑞, 闫彩良. 图像特征提取方法的综述[J]. 吉首大学学报(自然科学版), 2011, 32(5): 43-47.
- [15] Lowe, D.G. (2004) Distinctive Image Features from Scale Invariant Keypoints. *International Journal of Computer Vision*, **60**, 91-110. <https://doi.org/10.1023/B:VISI.0000029664.99615.94>