

基于生成对抗网络的焊缝X射线样本生成方法研究

王雪 黄旭 邹懿

(重庆科技学院电气工程学院, 重庆 401331)

[摘要] 研究压力容器焊缝检测中训练样本的生成问题, 训练样本过少会导致检测模型训练不充分, 影响检测精度。对WGAN (Wasserstein-GAN) 使用AdaBelief优化器进行优化, 代替随机梯度下降, 提升训练GAN的稳定性, 提高生成样本的质量。以解决压力容器焊缝缺陷样本质量差, 数量少的问题。通过GDX-ray焊缝图像生成实验证明, 使用AdaBelief优化器的WGAN能够在小样本的情况下生成高质量的压力容器焊缝缺陷样本。

[关键词] 焊缝; 小样本; 生成对抗网络; AdaBelief

[DOI] 10.12252/j.issn.2096-627X.2021.12.1632

引言

随着深度学习不断发展, 在缺陷检测领域得到广泛应用, 焊缝缺陷检测领域深度学习算法取得了较好的结果。但深度学习算法对样本在数量、质量、分布等方面的要求较高。压力容器的实际焊缝缺陷存在以下问题: 1、可供模型训练的样本极度匮乏, 样本上的缺陷分布极不均匀, 给模型训练带来了很大困难; 2、随着当前加工水平的提高, 使得某些缺陷样本几乎不出现, 对某些缺陷无法进行精确的检测; 3、在实际工程中, 样本被污染被错误标识的情况时有发生, 导致样本的有效性得不到保证。已公开的焊缝数据集中焊缝缺陷样本中不足500个, 因此必须要对实际缺陷样本进行必要的扩充, 本文使用改进后的生成对抗网络 (Generative Adversarial Networks, GAN) 算法进行焊缝样本扩充。

经过近几年 GAN 的不断发展, 由于其强大的图像生成和修复能力, GAN开始用于缺陷检测之中。2018年Mark Sabini[1]使用GAN实现了图片外信息的补充, 在原始图片中推理出额外的信息, 实现了图片修补。2019年, Liu L[2]等人提将GAN用于无缺陷样本表面缺陷检测, 通过人为随机对原图像样本添加缺陷, 用这些人获得的缺陷样本对GAN进行训练。这种方法在对表面缺陷图片的处理上有很好的作用, 在缺陷识别上使用了LBP (Local Binary Pattern) 算法, 能够准确找到缺陷位置。2019年黄旭丰等[3]使用GAN扩充数据集, 并且对轻量卷积神经网络进行Dropblock和全局平均池化优化, 有效提高了焊缝缺陷识别的准确率和泛化能力。这极大减少训练所需要的缺陷样本, 解决了样本不足的问题。2020年谷静[4]等人提出一种多尺度压缩激励网络模型 (SINet), 采用深度卷积对抗生成网络 (DCGAN) 进行数据集的平衡处理, 并在该数据集上验证模型的有效性, 能对焊缝缺陷图像进行有效的分类。但是在模型训练中存在着模型崩塌、收敛等问题。为解决上述问题本文在W-GAN上改变原有的随机梯度下降优化方式, 使用AdaBelief优化器进行优化, 设计合理的网络结构, 解决在小样本的情况下生成高质量的压力容器焊缝缺陷样本的问题。

一、生成对抗网络与AdaBelief优化器

(一) 生成对抗网络

Ian Goodfellow[5]提出生成对抗网络灵感的来源于博弈论中零和博弈, 他设计出由生成器与判别器组成的深度学习网络。由于其不受先验知识限制, 能够解决样本分布不均匀的问题, 得到广泛关注, 并发展出WGAN、DCGAN、CycleGAN等一系列模型。生成对抗网络在解决机器视觉等领域的问题上有着优

异表现。

生成对抗网络中包含生成器G和判别器D两部分。G通过输入一个符合分布的随机噪声z, 输出一个符合采样分布的样本G(z)。设定真实数据分布为, 通过训练让G可以生成贴近的数据G(z), G的目标是输出一个与真实数据极其相似的样本。将G生成的样本与真实样本一起作为D的输入, D判断输入是否为真实数据, 将判断结果D(x)输出, 经过不断训练使D无法判断输入样本的真伪, 即D和G达到纳什均衡点, 这时认为生成对抗网络训练完成, G找到真实数据分布, 可以生成样本。在原始GAN中, 如果让生成的分布函数与十分接近, 需要在中加入参数, 对进行随机采样, 计算最大似然估计并解出:

$$L = \prod_{j=1}^n P_G(x^j, \theta) \quad (1)$$

$$\theta = \arg \min KL(P_{data}(x) || P_G(x, \theta)) \quad (2)$$

当判别器最优即:

$$D(x) = \frac{P_{data}(x)}{P_{data}(x) + P_G(x)} \quad (3)$$

此时似然函数最大, 可以将 P_G 逼近 P_{data} 的问题转化为求解最小化分布的Jensen Shannon散度 (JSD) 的问题。由此得到GAN的目标函数为:

$$\min_G \max_D V(G, D) = E_{x \sim P_{data}(x)} [\log D(x)] + E_{z \sim P_z(z)} [\log(1 - D(G(z)))] \quad (4)$$

从公式中可以看出, 不断训练D会增大公式前半部分的值, 当输入真实数据时 $D(x)=1$ 。当输入由G生成的数据时, D的目标是 $D(G(x))=0$, 这与生成器的目标相悖。如果D的训练效果极好, 会使 $\log(1 - D(G(x)))$ 达到饱和, 造成梯度消失和模式崩塌。现阶段GAN对未知分布无法做到充分建模, 因此需要对目标函数和网络结构进行优化。

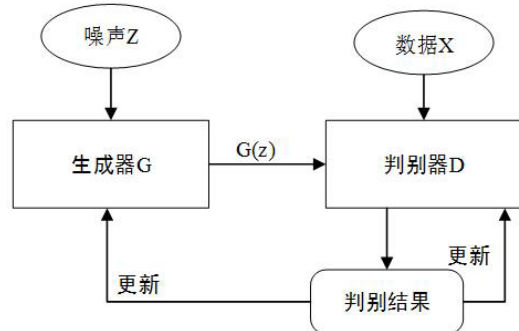


图1 生成对抗网络模型原理

本文使用的WGAN是使用Integral Probability Metric (IPM) 作为两个分布的度量标准, 并通过Wasserstein距离代替JSD对 P_g 和 P_{data} 的差异距离进行计算。

定义如下:

$$W(P_{data}, P_g) = \inf_{\gamma \in \Pi(P_{data}, P_g)} E_{(x,y) \sim \gamma} [\|x - y\|] \quad (5)$$

$$L = E_{x \sim P_{data}} [f(x)] - E_{x \sim P_g} [f(x)] \quad (6)$$

在WGAN中用L近似两个分布之间的Wasserstein距离。WGAN将和结合组成联合分布, 从中采集真实数据和生成数据, 拟合Wasserstein距离及其期望。判别器D的任务变成了计算两个分布集合的距离, 生成器G的目标是两者距离变小, D的目标是两者距离变大。因为此时D是在拟合拟合Wasserstein距离, 所以需要去除网络的最后一层sigmoid激活函数。此时提出WGAN中G、D的损失函数为:

$$G_{loss} = -E_{x \sim P_g} [f_w(x)] \quad (7)$$

$$D_{loss} = E_{x \sim P_g} [f_w(x)] - E_{x \sim P_{data}} [f_w(x)] \quad (8)$$

综上WGAN使用Wasserstein距离作为衡量 P_g 和 P_{data} 的差异方法, 在二者之间交集较小时仍能表示二者差异。很好地解决了原始GAN在D训练过度时, 产生的梯度消失和模式崩溃等问题。因此本文选用WGAN作为生成焊缝缺陷样本的基础模型。图2是WGAN原文中的算法流程图[6]。

Algorithm 1 WGAN, our proposed algorithm. All experiments in the paper use the default values $\alpha = 0.00005$, $c = 0.01$, $m = 64$, $n_{critic} = 5$.

```

Require:  $\alpha$ , the learning rate.  $c$ , the clipping parameter.  $m$ , the batch size.
 $n_{critic}$ , the number of iterations of the critic per generator iteration.
Require:  $w_0$ , initial critic parameters.  $\theta_0$ , initial generator's parameters.
1: while  $\theta$  has not converged do
2:   for  $t = 0, \dots, n_{critic}$  do
3:     Sample  $\{x^{(i)}\}_{i=1}^m \sim \mathbb{P}$ , a batch from the real data.
4:     Sample  $\{z^{(i)}\}_{i=1}^m \sim p(z)$  a batch of prior samples.
5:      $g_w \leftarrow \nabla_w [\frac{1}{m} \sum_{i=1}^m f_w(x^{(i)}) - \frac{1}{m} \sum_{i=1}^m f_w(g_\theta(z^{(i)}))]$ 
6:      $w \leftarrow w + \alpha \cdot \text{RMSProp}(w, g_w)$ 
7:      $w \leftarrow \text{clip}(w, -c, c)$ 
8:   end for
9:   Sample  $\{z^{(i)}\}_{i=1}^m \sim p(z)$  a batch of prior samples.
10:   $g_\theta \leftarrow \nabla_\theta [\frac{1}{m} \sum_{i=1}^m f_w(g_\theta(z^{(i)}))]$ 
11:   $\theta \leftarrow \theta - \alpha \cdot \text{RMSProp}(\theta, g_\theta)$ 
12: end while
    
```

图2 WGAN算法流程图

(二) AdaBelief优化器

2020年由耶鲁大学团队提出AdaBelief优化器[7], 其团队表示AdaBelief有着与Adam的收敛速度快, SDG的泛化性好的优势。WGAN在提出时, 其作者推荐使用RMSProp或者SGD作为WGAN的优化器, 本文考虑使用AdaBelief作为WGAN的优化器, 对WGAN进行训练。

从算法上看, AdaBelief与Adam差别在于, AdaBelief避免新参数的引入, 只是在对 s_t 更新时有所差异。AdaBelief与Adam相同以 θ_t 作为损失函数参数, AdaBelief以 $m_t/\sqrt{s_t}$ 为梯度更新方向, s_t 是 $(g_t - m_t)^2$ 的指数平均, g_t 是步骤t的梯度, 而 m_t 是 g_t 的预测。AdaBelief考虑到损失函数曲率, 当 m_t 与 g_t 的差值较小时, 优化器步长长, 权重继续更新, 当 m_t 与 g_t 的差值较大时, 优化器步长短, 通过 g_t 的不断变化进行更新。图3是耶鲁大学团队论文中的详细算法。

Algorithm 1: AdaBelief

```

Initialize  $\theta_0$ 
 $m_0 \leftarrow 0, s_0 \leftarrow 0, t \leftarrow 0$ 
While  $\theta_t$  not converged
 $t \leftarrow t + 1$ 
 $g_t \leftarrow \nabla_{\theta} f_t(\theta_{t-1})$ 
 $m_t \leftarrow \beta_1 m_{t-1} + (1 - \beta_1) g_t$ 
 $s_t \leftarrow \beta_2 s_{t-1} + (1 - \beta_2) (g_t - m_t)^2 + \epsilon$ 
If AMSGrad
 $s_t \leftarrow \max(s_t, s_{t-1})$ 
Bias Correction
 $\widehat{m}_t \leftarrow m_t / (1 - \beta_1^t), \widehat{s}_t \leftarrow (s_t) / (1 - \beta_2^t)$ 
Update
 $\theta_t \leftarrow \Pi_{\mathcal{F}, \sqrt{\widehat{s}_t}} (\theta_{t-1} - \widehat{m}_t \frac{\alpha}{\sqrt{\widehat{s}_t + \epsilon}})$ 
    
```

图3 Adabelief优化器算法流程图

二、实验

本文使用GDX-ray焊缝缺陷数据集进行模型训练。该数据集有完整焊缝样本10张, 完成切割和标注的样本68张, 通过图片分割将完整焊缝样本与标注样本进行整合获得400张训练样本, 包括条形缺陷样本153张、圆形缺陷样本136张、未焊透样本111张。本文硬件配置为: Intel Core i5-9600KF CPU @ 3.70GHz + Nvidia GeForce RTX 2060 (6 GB), 软件环境为: Python3.7及Tensorflow 1.13。

(一) 模型训练

针对焊缝缺陷小样本图像生成问题, 本文设计WGAN的生成器G包含4层全连接层和2层反卷积层, 除最后一层反卷积层使用Sigmoid函数外其余使用LeakyReLU激活函数。判别器D包含4个卷积层和2个全连接层, 最后一层不再使用Sigmoid函数。每个卷积单元由卷积层、实例归一化层和激活层组成。本文WGAN模型使用AdaBelief优化器进行优化, 学习率设置为0.002。为保证不出现梯度消失, 先训练判别器D, 更新一定次数判别器D后再训练生成器G。训练模型如图4所示。

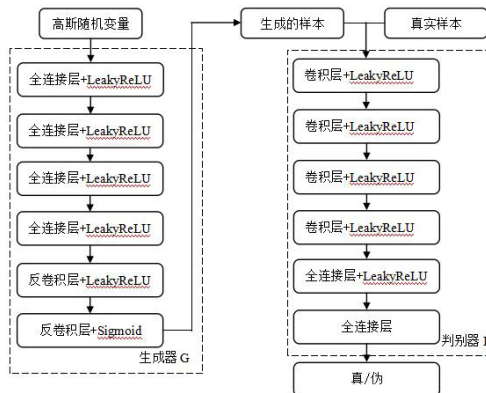


图4 WGAN网络模型

(二) 实验结果评估

WGAN在GDX-ray焊缝缺陷数据集下训练, 图5记录随训练次数的增加模型生成的样本, 每20步输出一次样本。

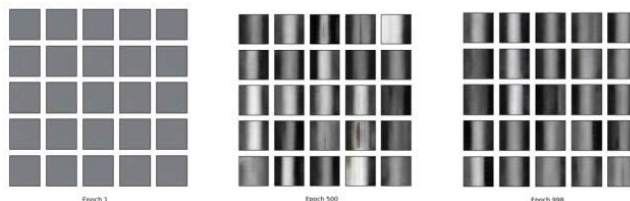


图5 训练时生成的样本

记录迭代每次时的G_loss和D_loss,发现判别器D在第600次时已经完成收敛,生成器G在800次时完成收敛。继续训练损失函数有波动,但是趋于稳定,认为模型训练完成。图6为训练时的G_loss和D_loss。

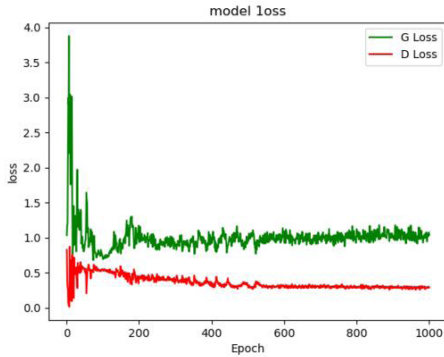
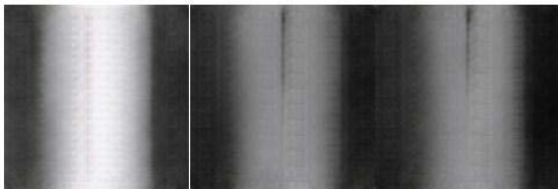


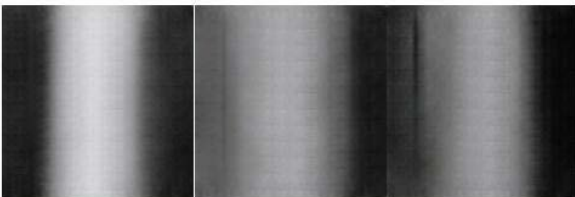
图6 训练时G_loss和D_loss

当使用Adma进行训练时,判别器D在训练第600次的D_loss,生成器G在训练第800次的G_loss均高于使用AdaBelief优化器。比较可知使用AdaBelief优化器对WGAN模型进行优化时,有着比Adma优化器更快的训练速度和更好的泛化能力。

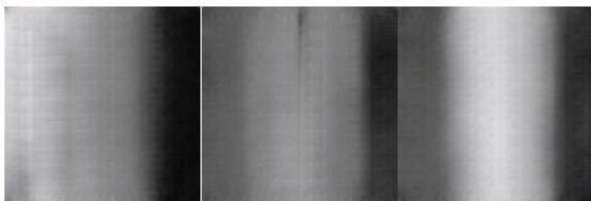
本文WGAN模型在测试集上的表现如图7所示。此模型在小样本的情况下有着较强的图像生成能力,可以生成与焊缝缺陷图像语义相似的高质量图像。



(a) 条形缺陷生成样本



(b) 未焊透生成样本



(c) 圆形缺陷生成样本

图7 生成的样本

通过计算FID来评估本算法的生成图像质量,不同类型缺陷生成的图像质量如表1所示。进一步证明本模型所生成样本的有较高质量和较好的多样性。从表1可以看出,在生成条形缺陷上以FID作为评判标准表现很好,FID为17.56。生成的圆形缺陷是三种缺陷中需要进一步优化的。由此可见该模型生产的条形缺陷和未熔合缺陷样本有着较高的图像质量,圆形缺陷质量较好。

表1 不同焊缝类型生成图像的FID

缺陷类型	FID
条形缺陷	17.5687
圆形缺陷	34.8041
未熔合	22.4758

结语

针对压力容器焊缝样本少、质量差,难以满足深度学习模型训练要求的问题,本文提出基于生成对抗网络完成焊缝X射线样本生成方法,使用AdaBelief优化器代替Adam优化器对WGAN模型进行优化。本文通过使用GDX-ray数据集进行样本生成的实验证明该模型的有效性,通过计算生成样本与原始样本的FID验证生成的样本质量,为解决焊缝样本不均衡问题提供帮助。由于在实验过程中发现WGAN生成的图像仍可以再次优化,接下来的工作将在如何进一步提高焊缝缺陷的生成能力展开,进一步提升样本质量和模型稳定性。

参考文献:

[1]Sabini M,Rusak G . Painting Outside the Box : Image Outpainting with GANs: [J].Computervision and Pattern Recognition, 2018.

[2]Liu L, Cao D, Wu Y, Wei T. Defective Samples Simulation through Adversarial Training for Automatic Surface Inspection. Neurocomputing, 2019, 360: 230-245

[3]黄旭丰. 基于深度迁移学习的焊接质量在线监测方法研究[D]. 南宁: 广西大学, 2019.

[4]谷静, 王琦雯, 张敏, 等. 基于DenseNet网络的焊缝缺陷检测识别[J]. 传感器与微系统, 2020, 39(09): 129-131.

[5]Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Networks[J]. 2014: 2672-2680.

[6]Arjovsky M, Chintala S, Bottou L. Wasserstein GAN[J]. 2017.

[7]Zhuang J, Tang T, Tatikonda S, et al. AdaBelief Optimizer: Adapting Stepsizes by the Belief in Observed Gradients[J]. 2020.

基金项目: 重庆科技学院硕士研究生创新计划项目 “基于生成对抗网络的焊缝缺陷X射线图像识别方法研究”(编号: YKJXC2020421)

重庆科技学院硕士研究生创新计划项目 “基于深度学习的工业现场不规范行为快速检测算法研究”(编号: YKJXC2020417)

作者简介:

王雪(1971.9-), 男, 汉, 四川自贡人, 副教授, 研究方向: 信号处理、故障诊断;

黄旭(1996.12-), 男, 汉, 安徽淮南人, 硕士研究生, 资源与环境, 研究方向: 图像处理、目标检测;

邹懿(1996.7-), 男, 汉, 重庆人, 硕士研究生, 资源与环境, 研究方向: 目标检测、图像处理。