

分类号\_\_\_\_\_

编 号\_\_\_\_\_

U D C\_\_\_\_\_

密 级\_\_\_\_\_



南方科技大学  
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

# 本科生毕业设计（论文）

题 目： 垃圾分类任务的数据增广方法

姓 名： 黄宇海

学 号： 12011507

院 系： 计算机科学与工程系

专 业： 计算机科学与技术

指导教师： 史玉回讲席教授

2024 年      6 月      7 日

# 诚信承诺书

1. 本人郑重承诺所呈交的毕业设计（论文），是在导师的指导下，独立进行研究工作所取得的成果，所有数据、图片资料均真实可靠。

2. 除文中已经注明引用的内容外，本论文不包含任何其他人或集体已经发表或撰写过的作品或成果。对本论文的研究作出重要贡献的个人和集体，均已在文中以明确的方式标明。

3. 本人承诺在毕业论文（设计）选题和研究内容过程中没有抄袭他人研究成果和伪造相关数据等行为。

4. 在毕业论文（设计）中对侵犯任何方面知识产权的行为，由本人承担相应的法律责任。

作者签名：黄宇海

2024 年 6 月 7 日

# 垃圾分类任务的数据增广方法

黄宇海

(计算机科学与工程系 指导教师：史玉回讲席教授)

**[摘要]**：在许多领域，受限于数据获取难度大，标注成本高等原因，往往难以获得充足的训练数据，这样训练得到的深度学习模型往往存在过拟合的问题，进而导致模型泛化能力差，测试精度不高等。本论文以垃圾分类领域为例进行研究，旨在通过图像增广技术解决该领域中存在的数据稀缺和样本不均衡问题，提升垃圾分类器。提出了一种基于生成对抗网络的图像增广方案，通过引入双向生成对抗网络模型，生成了大量具有多样性的垃圾图像样本，并将其与原始数据集合并，构建了一个更加丰富和全面的数据集。实验结果表明，经过 BiGANs (Bidirectional Generative Adversarial Network) 扩增的数据集训练出的分类器，相比于仅使用原始数据集训练的模型，准确率显著提升了约 22%，特别是在少数类别的识别上表现更为突出。此外，通过定性和定量评估，生成样本展示了高度的现实感和类别一致性，证实了方法的有效性和泛化能力。这项研究推动了垃圾分类领域的发展，并为应用数据增强技术克服深度学习中的数据稀缺性挑战提供了宝贵的见解。

**[关键词]**：图像增广；双向生成对抗网络；AlexNet；LeNet-5

**[ABSTRACT]:** In many fields, it is often difficult to obtain sufficient training data due to the difficulty of data acquisition and the high cost of labeling, so that the trained deep learning model often has the problem of overfitting, which in turn leads to poor model generalization ability and low test accuracy. This paper takes the field of garbage classification as an example, aiming to solve the problems of data scarcity and sample imbalance in this field through image augmentation technology, and improve the garbage sorter. By introducing the Bidirectional Generative Adversarial Network (BiGANs) model, we generated a large number of diverse garbage image samples and merged them with the original dataset to construct a richer and more comprehensive dataset. Through experimental verification, we find that the accuracy of the classifier trained on the dataset amplified by BiGANs is significantly improved by about 22% compared with the model trained only with the original dataset, especially in the recognition of a few categories. In addition, through visual and quantitative evaluation, the generated samples demonstrated a high degree of realism and categorical consistency, confirming the validity and generalizability of the method. This research advances the field of waste sorting and provides valuable insights into the application of data augmentation techniques to overcome the challenge of data scarcity in deep learning.

**[Keywords]:** data augmentation; BiGANs; AlexNet; LeNet-5

# 目录

1. 引言. ....	7
2. 理论与方法基础. ....	9
2.1 基于基本图像操作的图像增广.....	9
2.1.1 翻转(Flipping).....	9
2.1.2 色彩空间 (Color).....	9
2.1.3 剪裁(Cropping).....	10
2.1.4 旋转(Rotation).....	10
2.1.5 移位(Translation).....	10
2.1.6 噪声注入(Noise).....	10
2.2 基于生成建模的图像增广.....	11
2.2.1 变分自编码器 (Variational Auto-encoder).....	11
2.2.2 生成对抗网络 (Generative Adversarial Networks).....	12
2.3 两种方式的对比.....	12
3. 基于双向生成对抗网络 (Bidirectional Generative Adversarial Network) 的图像增广方法.....	14
3.1 编码器.....	14
3.2 生成器和辨别器.....	16
3.3 实现过程.....	17
3.3.1 预处理.....	17
3.3.2 初始化与构建模型.....	17

3.3.3 训练.....	18
3.4 改进方法与创新.....	19
<b>4. 实验. ....</b>	<b>20</b>
4.1 数据集.....	20
4.2 评价指标.....	20
4.3 BiGAN 模型.....	20
4.3.1 参数设置.....	20
4.3.2 优化器及损失函数.....	21
4.3.3 模型构建及训练.....	21
4.4 垃圾分类器.....	22
4.4.1 网络结构.....	22
4.4.2 参数设置.....	22
4.4.3 优化器及损失函数.....	23
4.4.4 模型训练.....	23
4.5 对比实验.....	23
4.6 结果与分析.....	24
4.6.1 视觉角度.....	24
4.6.2 数据角度.....	25
<b>5. 总结. ....</b>	<b>29</b>
<b>参考文献. ....</b>	<b>20</b>
<b>致谢. ....</b>	<b>32</b>

## 1. 引言

在当今社会，随着人工智能技术的飞速进步，深度学习模型因其强大的数据处理能力和模式识别能力，在诸多领域内展现出巨大潜力，如自动驾驶、医疗诊断、图像识别等。然而，深度学习模型的卓越性能往往建立在海量标注数据的基础之上，这在很多实际应用场景中构成了一个显著的挑战。特别是像垃圾分类这样的领域，数据获取的困难尤为突出，主要体现在两个方面：一是实际收集各类垃圾图片的成本高且耗时，二是不同类别垃圾的分布往往极不均匀，导致某些类别样本稀缺。这种“数据荒”和样本不均衡问题直接限制了模型的训练质量和最终性能，易引发过拟合现象，进而导致模型泛化能力差，最终测试精度不尽人意。

利用传统的图像增广技术生成的样本缺乏创新性和多样性，尽管在一定程度上能提升模型的性能，但对于复杂场景和细节丰富的图像效果有限，仍难以解决这种“数据荒”和样本不均衡的问题，此外，直接应用标准生成对抗网络虽然能生成新样本，但可能难以保证生成样本类别的一致性，且训练过程相对不稳定。本论文的研究聚焦于如何有效应对垃圾分类领域的这一难题，提出了利用双向生成对抗网络进行图像增广的创新策略。双向生成对抗网络作为一种结合了生成对抗网络与变分自编码器优势的混合模型，不仅能够生成高度逼真的图像样本，而且通过编码器的引入，实现了对潜在空间的有监督学习，使得生成过程能够更好地匹配特定的类别标签。通过双向生成对抗网络生成的大量多样化垃圾图像样本，不仅扩充了原有的数据集，还特别增强了稀有类别的数据量，从而有效地缓解了数据稀缺和样本不均衡的问题。通过实验验证，发现经过双向生成对抗网络扩增的数据集训练出的分类器，相比于仅使用原始数据集训练的模型，准确率显著提升了约 22%，特别是在少数类别的识别上表现更为突出。生成的样本经过视觉和定量评估，证明具有高度的现实感和类别一致性，进一步证实了所提方法的有效性和泛化能力。

本文的主要贡献在于：（1）探讨了双向生成对抗网络在图像增广领域的应用潜力，特别是在垃圾分类这一具体场景下的创新实践；（2）实现了一套基于双向生成对抗网络的图像增广流程，详细阐述其生成器、判别器及编码器的设计；

(3) 通过广泛的实验验证，展示增广后数据集对于提升垃圾分类模型准确率与泛化能力的显著效果。

论文结构安排如下：第二章将详细介绍图像增广的基本方法及双向生成对抗网络框架的构成要素，包括生成器、辨别器和编码器的设计原理；第三章概述基础理论，为后续的实验与分析奠定理论基础；第四章展示实验过程以及分析实验结果，包括数据集选取、模型训练，参数设置，评估模型性能提升的具体数值与效果等；第五章对本论文工作进行总结。



## 2. 理论与方法基础

图像增广指的是通过对现有数据集进行一系列的变换或处理，从而生成新的数据样本，以扩展原始数据集的数量和多样性。这些变换或处理可以包括旋转、缩放、平移、翻转、裁剪、变形等操作，也可以包括添加噪声、调整亮度、对比度、色彩等操作。图像增广旨在提高机器学习模型的泛化能力，减少过拟合，增加模型对于不同变换、干扰和噪声的鲁棒性，从而提高模型在真实场景中的性能表现。本章将从基于基本图像操作的图像增广和基于生成建模的图像增广两方面进行介绍。

### 2.1 基于基本图像操作的图像增广

本节介绍基于几何变换和许多其他图像处理功能的不同增广方法。下面讨论的增广类别可以通过它们的易实现性来进行特征化。了解这些转换将为进一步研究图像增广技术提供有用的基础。

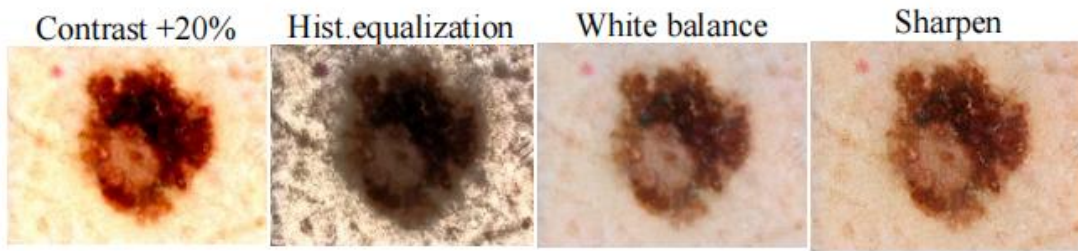


图 1: Mikolajczyk 和 Grochowski 在黑色素瘤分类领域颜色增强的例子<sup>[1]</sup>

#### 2.1.1 翻转 (Flipping)

翻转是图像增广的一种简单方法，它可以通过水平翻转或垂直翻转来增加数据样本的多样性。水平翻转是指将图像沿着垂直中轴线进行镜像翻转，而垂直翻转则是将图像沿着水平中轴线进行镜像翻转。水平翻转比垂直翻转更普遍。这种增强方法是最容易实现的之一，并且在如 CIFAR-10<sup>[2]</sup> 和 ImageNet<sup>[3]</sup> 等数据集上已被证明是有效的<sup>[4]</sup>。但在涉及文本识别的数据集，如 MNIST<sup>[5]</sup> 或 SVHN<sup>[6]</sup>，这不是一个安全的转换，比如数字“6”进行翻转后可变成数字“9”，他们的标签类别发生变换。

#### 2.1.2 色彩空间 (Color)

数字图像数据通常被编码为三维张量（高度×宽度×颜色通道）。在颜色通

道空间执行增广是另一种非常实用的策略（图 1）。非常简单的颜色增广包括隔离单个颜色通道，例如 R、G 或 B。通过隔离该矩阵并从其他颜色通道添加两个零矩阵，图像可以快速转换为其在一个颜色通道中的表示<sup>[7]</sup>。此外，RGB 值可以通过简单的矩阵操作轻松地进行操作，以增加或减少图像的亮度。更高级的颜色增广来自于生成描述图像的颜色直方图。更改这些直方图中的强度值会导致光照变化，就像在照片编辑应用程序中使用的那样<sup>[4]</sup>。

### 2.1.3 剪裁 (Cropping)

裁剪图像可以作为一种实用的处理步骤，用于处理具有不同高度和宽度尺寸的图像数据，方法是从每张图像中裁取一个中心区域。此外，随机裁剪也可以用来提供类似于平移的效果。随机裁剪与平移之间的区别在于，裁剪会减小输入图像的尺寸，例如从 (256, 256) 变为 (224, 224)，而平移则保持图像的空间尺寸不变<sup>[8]</sup>。根据选择的裁剪程度，这种裁剪可能不会保留图像的标签信息，即可能导致裁剪后的图像不再属于原来的类别。

### 2.1.4 旋转 (Rotation)

旋转增强是通过在  $1^\circ$  和  $359^\circ$  之间的轴上左右旋转图像来完成的。旋转增强的安全性在很大程度上取决于旋转度参数的决定。在 1 到 20 或 -1 到 -20 之间的轻微旋转可能对 MNIST 等数字识别任务有用，但随着旋转程度的增加，数据的标签在转换后不再保留<sup>[9]</sup>。

### 2.1.5 移位 (Translation)

将图像向左、右、上或下移动是一种非常实用的转换手段，可以用来避免数据中的位置偏差。例如，如果数据集中的所有图像都是居中的，这在面部识别数据集中很常见，这就要求模型也要在完全居中的图像上进行测试<sup>[9]</sup>。当原始图像沿某个方向平移位时，剩余的空间可以用恒定值填充，如全 0 或全 255，或者填充随机值或高斯分布噪声。这种填充方式在图像增广后保持了图像的空间维度不变。

### 2.1.6 噪声注入 (Noise)

噪声注入 (Noise Injection) 是指在数据中加入一个由随机值构成的矩阵，这些随机值通常是从高斯分布（也称为正态分布）中抽取的。Moreno-Barea 等人

<sup>[10]</sup>在来自 UCI 存储库<sup>[11]</sup>的九个数据集上测试了噪声注入。向图像中添加噪声可以帮助卷积神经网络学习更加鲁棒的特征。

几何变换是解决训练数据中位置偏差的很好的解决方案。训练数据与测试数据分布不同可能产生多种偏差来源。如果在人脸识别数据集中存在位置偏差，也就是说，每张脸都没有完美地居于框架中，那么可以考虑使用几何变换作为解决方案。除了强大地克服位置偏差的能力之外，几何变换之所以有用，还因为它们易于实现。许多图像处理库使得水平翻转和旋转等操作变得轻松入手。几何变换的一些缺点包括额外的内存消耗、变换计算成本和额外的训练时间。一些几何变换，如平移或随机裁剪，必须手动观察以确保它们没有改变图像的标签。在许多应用领域中，比如医学图像分析，训练数据与测试数据之间的偏差比位置和平移变化更加复杂。因此，几何变换可以应用的范围和时间相对有限。

## 2.2 基于生成建模的图像增广

生成建模是指从数据集创建人工实例的做法，以便它们保留与原始集合相似的特征<sup>[12]</sup>。生成对抗网络和变分自编码器都属于生成建模的方式。

### 2.2.1 变分自编码器（Variational Auto-encoder）

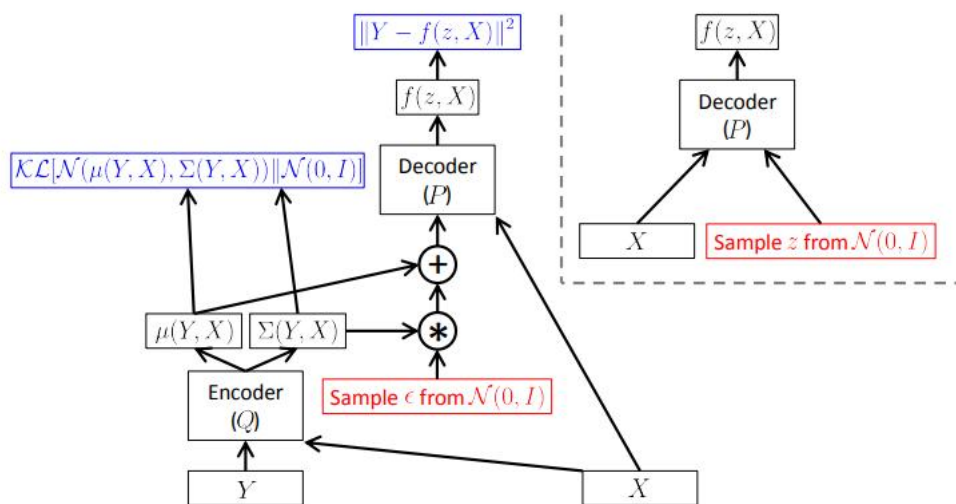


图 2：变分自编码器结构图<sup>[13]</sup>

变分自编码器包括两部分：编码器和解码器。编码器负责将输入数据（如图像）转化为一个低维的隐向量，这个向量代表了输入数据的一种压缩表示。解码

器则尝试根据这个隐向量重构原始数据。在训练过程中，变分自编码器通过最大化数据的似然来学习参数，同时确保隐向量服从一个简单的先验分布（通常是高斯分布），这有助于生成多样化的样本。从隐向量的先验分布中采样，然后将这些采样的隐向量输入解码器，就能得到新的与训练数据类似但不完全相同的图像。但生成图像的质量通常不如专门设计的生成对抗网络。

## 2.2.2 生成对抗网络（Generative Adversarial Networks）

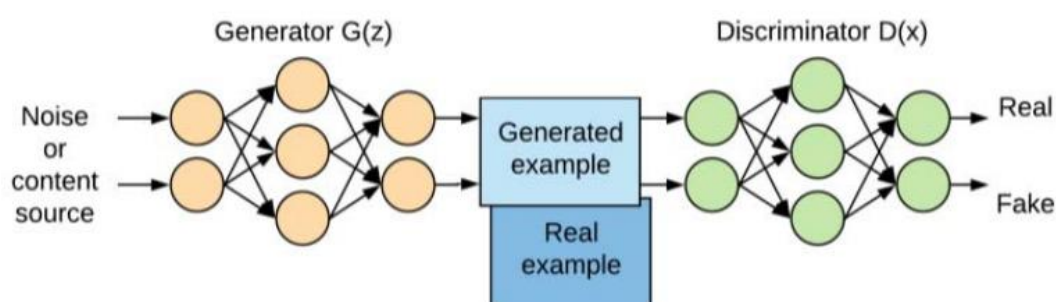


图 3：由 Mikolajczyk 和 Grochowski 提供的生成对抗网络概念图<sup>[1]</sup>

生成对抗网络由两个核心组件构成：生成器（Generator）和判别器（Discriminator）。生成器负责创建新的数据实例，试图模仿真实数据的分布；判别器则负责区分这些生成的实例和真实的训练数据，判断它们是否来自真实数据集。两个网络通过一种对抗过程相互竞争和学习，最终生成器能够产出越来越难以被判别器辨认的样本，即接近真实数据的新数据。

## 2.3 两种方式的对比

**图像质量与逼真度：**VAEs（Variational Auto-encoder）生成的新数据实例往往在视觉质量上不如 GANs（Generative Adversarial Networks）所生成的图像清晰和逼真。GANs 通过其对抗训练机制，其中生成器努力欺骗判别器，促使生成的图像尽可能接近真实数据分布，因此在许多情况下能够产生更精细、细节更丰富的图像。

**模式多样性：**GANs 倾向于捕捉数据分布中的多种模式，生成多样化的样本，而 VAEs 有时会生成较模糊或过于平滑的图像，缺乏多样性。这是因为 VAEs 在优化过程中通常倾向于最小化重构误差，这可能导致生成的样本集中在数据分布的均值附近，而非探索数据的全部范围。

训练稳定性: GANs 的训练相对不稳定, 需要精心调整的超参数和训练策略以避免模式塌陷(mode collapse)等问题, 即生成器总是生成同一类型的少数几个样本。

### 3. 基于双向生成对抗网络（BiGANs）的图像增广方法

由第 2 章可知道通过变分自编码器生成的图像缺乏多样性，但训练稳定，生成对抗网络生成的图像质量更高，但训练不稳定。而有一种名为双向生成对抗网络的模型结合了生成对抗网络与变分自编码器各自的优势，不仅能够生成高度逼真的图像样本，而且通过编码器的引入，实现了对潜在空间的有监督学习，使得生成过程能够更好地匹配特定的类别标签。本文将使用这个模型对垃圾分类器的数据集进行图像增广，本章详细介绍双向生成对抗网络。

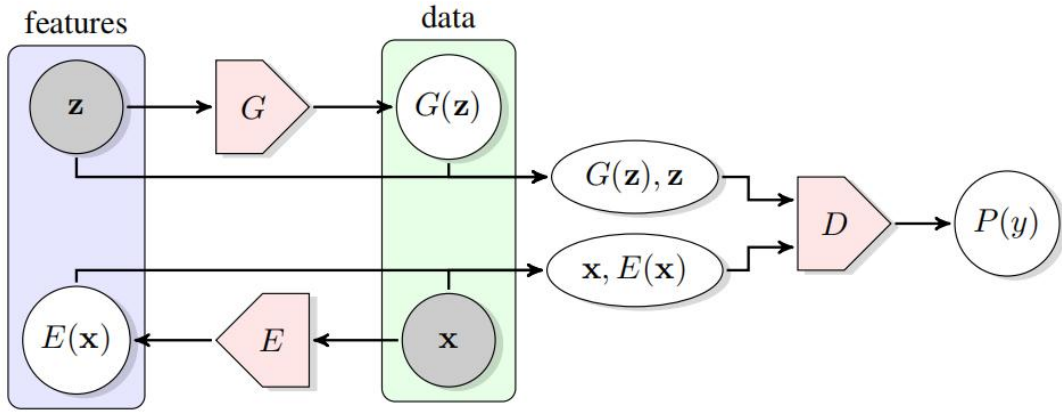


图 4：双向生成对抗网络的结构图<sup>[14]</sup>

除了来自标准 GANs 的生成器  $G$  外，还包括了一个编码器  $E$ ，它将数据  $x$  映射到潜在表示  $z$ ，判别器  $D$  不仅用于区分真实样本  $x$  和生成样本  $G(z)$ ，还被训练来预测真实样本  $x$  对应的潜在向量  $E(x)$ ，生成样本  $G(z)$  对应的潜在向量  $z$ 。这意味着判别器需要同时具备生成数据真伪判别能力以及逆映射能力，即从数据空间映射回潜在空间的能力。由于 BiGANs 编码器学习预测给定数据  $x$  的特征  $z$ ，并且之前对 GANs 的研究表明这些特征捕获了数据的语义属性，因此假设经过训练的 BiGANs 编码器可以作为相关语义任务的有用特征表示，就像训练用于预测给定图像的语义“标签”的完全监督视觉模型作为相关视觉任务的强大特征表示一样。在这种情况下，潜在表示  $z$  可以被认为是  $x$  的“标签”，但它是“免费”的，不需要监督。学习从数据到潜在表示的逆映射的另一种方法是直接建模  $p(z|G(z))$ ，预测给定生成的数据  $G(z)$  的生成器输入  $z$ 。这种方法称为潜在回归变量。

#### 3.1 编码器

编码器 E 架构遵循 AlexNet<sup>[15]</sup>通过第五个也是最后一个卷积层（Conv5），移除局部响应归一化层，批量归一化 (Batch Normalization) 和非线性 Leaky ReLU 应用与每个卷积层的输出。AlexNet<sup>[15]</sup>是一个经典的深度卷积神经网络，原本用于图像分类任务。选择 AlexNet<sup>[15]</sup>作为编码器的基础，是因为其在图像特征提取方面的高效性和广泛的成功应用。通过移除局部响应归一化层并加入批量归一化，可以提高训练速度和模型的泛化能力，而 LeakyReLU 相较于传统 ReLU，允许负值的微小梯度流动，有助于缓解梯度消失问题。编码器通过这些设计改进，能更有效地学习图像中的高层次特征，从而在潜在空间中编码这些特征，潜在空间的组织和编码效率直接影响生成样本的多样性和类别匹配度。

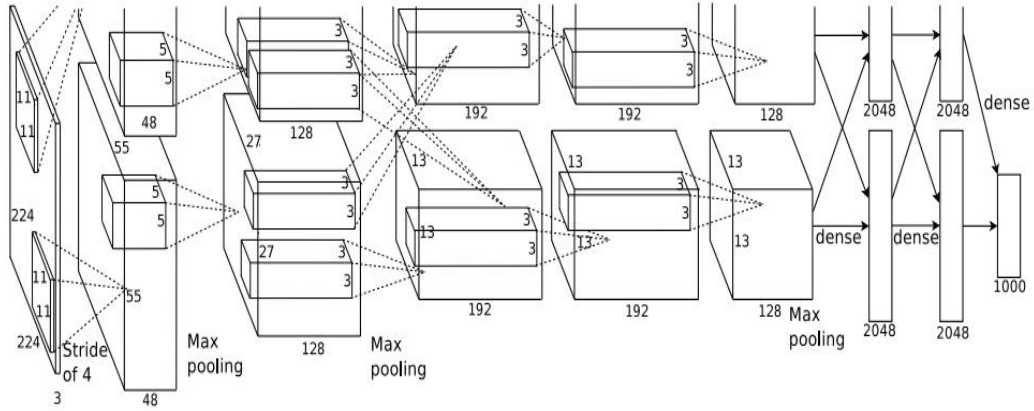


图 5： AlexNet 结构图<sup>[15]</sup>

卷积层 C1：使用 96 个核对输入图像进行滤波，卷积核大小为  $11 \times 11 \times 3$ ，步长为 4。

卷积层 C2：使用卷积层 C1 的输出作为输入，并使用 256 个卷积核进行滤波，核大小为  $5 \times 5 \times 48$ 。

卷积层 C3：有 384 个核，核大小为  $3 \times 3 \times 256$ ，与卷积层 C2 的输出相连。

卷积层 C4：有 384 个核，核大小为  $3 \times 3 \times 192$ 。

卷积层 C5：有 256 个核，核大小为  $3 \times 3 \times 192$ 。

其中，卷积层 C3、C4、C5 互相连接，中间没有接入池化层或归一化层。移除 C3 至 C5 之间的池化层，虽然牺牲了空间尺寸的缩小，但保留了更多空间信息，有助于生成更精确的图像细节。

同时，编码器 E 将真实数据映射到潜在空间，这一映射过程是通过有监督的



方式进行的。但它不同于以往显式的有监督训练通过使用传统的交叉熵函数来实现，上面简述的框架流程中也无法直接体现它的有监督性。它的监督性主要来源于两个方面：目标函数的引导作用和类别信息的隐含指导。

**目标函数的引导作用：**在训练过程中，编码器不仅仅要学习如何将数据映射到潜在空间，还要确保这个映射结果能够被判别器  $D$  正确地逆映射回去，即判别器不仅要判断真实样本和生成样本，还要能够从生成样本反推出潜在向量，并且这个潜在向量应当反映数据的真实类别。因此，编码器和判别器之间形成了一个有反馈的、相互促进的关系，其中编码器的输出在某种程度上被期望与已知的类别标签相匹配。

**类别信息的隐含指导：**虽然 BiGANs 在训练时并没有直接给编码器提供类别标签作为输入或损失函数的一部分，但是通过与判别器的对抗性训练，编码器生成的潜在向量实际上隐含了输入数据的类别信息。这可以理解作为一种软监督，因为它依赖于判别器对潜在空间中不同类别分布的区分能力，这与监督学习中直接使用类别标签指导模型学习不同类别的特征表达相似。

### 3.2 生成器和判别器

判别器  $D$  和生成器  $G$  的架构都是 Radford 等人（2016）使用的<sup>[16]</sup>。生成器  $G$  包括四个部分跨越卷积层，每层随后进行批归一化（加速模型训练过程，提高模型的稳定性和生成样本的质量）和校正非线性（引入非线性并促进模型的表达能力）。

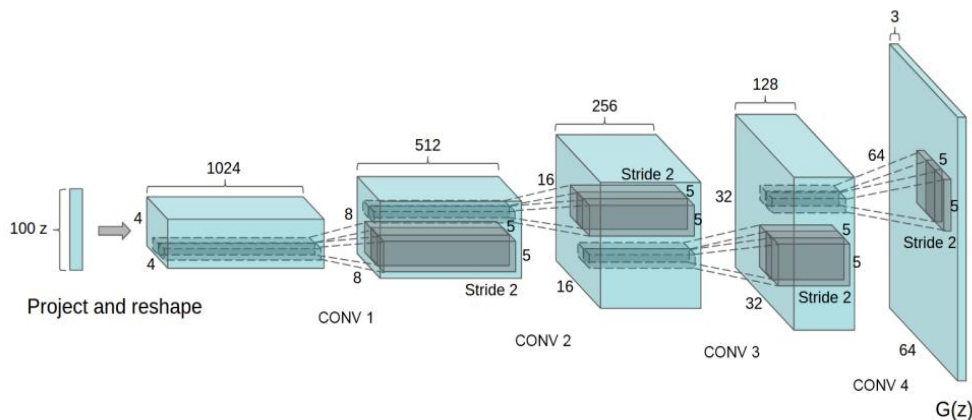


图 6: 生成器  $G$  结构图<sup>[16]</sup>

将 100 维均匀分布  $z$  投影到具有许多特征图的小空间范围卷积表示。然后，



经过一连串的 4 个  $5 \times 5$  部分跨越卷积层，步幅为 2，然后将“反卷积”后的这种高级表示转换为  $64 \times 64$  像素的图像。注意，没有使用全连接或池化层。

辨别器则使用一连串的卷积层，捕捉数据的局部特征并逐渐抽象出全局特征。与生成器相反，辨别器的卷积层通常减少特征图的尺寸，通过池化层或者步长大于 1 的卷积来实现，从而对输入数据进行降采样并提取高层特征。

### 3.3 实现过程

#### 3.3.1 预处理

将数据集的图像按照一定的比例调整大小，使其较短的边长度为 72。然后从调整大小后的图像中作  $64 \times 64$  大小的随机剪裁，剪裁后的图像水平翻转的概率为  $1/2$ ，并将其像素值缩放为  $[-1, 1]$ 。最后将所有数据以图 7 样式储存。

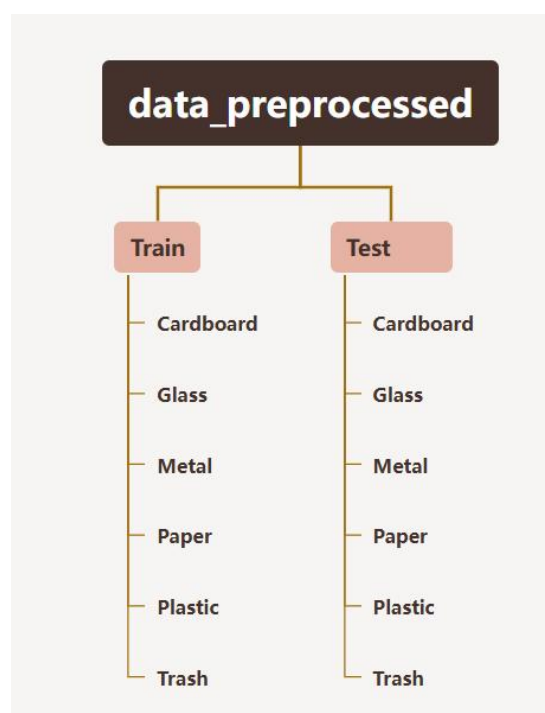


图 7：数据集储存样式

#### 3.3.2 初始化与构建模型

- 1) 初始化：定义图像的尺寸（ $64 \times 64 \times 3$ ），潜在向量的空间维度（ $4 \times 4 \times 256$ ），并设置优化器为 Adam。
- 2) 构建辨别器：辨别器接受两部分输入：一部分是潜在向量  $z$ ，另一部分是图像。为了将两者合并处理，首先对潜在向量  $z$  进行零填充使其形状与图像相同，然后通过 Concatenate 函数沿着深度轴直接拼接。然后根据 3.2 所提的辨别器结构进行构建，通过一系列的卷积层处理复合输入，每个卷积层输出

接 LeakyReLU 激活函数（带有一个斜率，用于非线性变换且更好处理梯度消失问题）和 BatchNormalization 层（加速模型的训练过程，提高模型稳定性）。最后通过一个具有 Sigmoid 激活函数的全连接层，输出一个介于 0 和 1 之间的值，代表输入对是这真实数据的概率，接近 1 表示模型认为输入对是真实的，接近 0 则认为是伪造的。

- 3) 构建生成器：生成器只有服从高斯分布的噪声  $z$  作为输入，然后根据 3.2 所提出的生成器结构进行构建，通过一系列的转置卷积层将低纬特征图上采样至更高的分辨率，同时增加特征通道数，逐步构建出更加精细的图像结构，除了最后一个转置卷积层，其他的之后接 LeakyReLU 激活函数（带有一个斜率，用于非线性变换且更好处理梯度消失问题）和 BatchNormalization 层（加速模型的训练过程，提高模型稳定性，提升生成样本的质量），最后一个转置卷积层使用 Tanh 函数约束输出范围  $[-1, 1]$  与预处理后的图像相对应。
- 4) 构建编码器：编码器只有图像作为输入，然后根据 3.1 所提出的编码器结构进行构建，通过一系列的卷积层将高分辨率的图像转换为低纬度的潜在表示，最后输出的潜在表示与噪声  $z$  大小相同。
- 5) 构建联合模型：联合生成器，判别器，编码器构建联合模型，以[真实图像，噪音]作为输入，最后输出图像标签对的结果。

### 3.3.3 训练

- 1) 加载预处理好的数据集

- 2) 循环训练，每个 epoch 执行以下步骤：

- 2.1) 生成样本：从潜在空间中随机抽取噪声  $z$ ，通过生成器生成伪图像。

- 2.2) 编码真实图像：随机抽取数据集的真实图像，通过编码器转换成潜在向量。

- 2.3) 训练判别器的辨别能力：使用二元交叉熵损失函数预测与实际标签的差距（真实图像配对标记为 1，生成图像配对标记为 0），优化目标是最大化这一损失，即尽可能准确地地区分真实数据和生成数据。

- 2.4) 训练编码器和生成器以及判别器的逆映射能力：从简单的分布中提取随机噪声  $z$ ，从训练集中随机采样真实图像，之后将[真实图像，

$z$ ]输入联合模型中，在联合模型中，生成器根据随机噪声  $z$  生成样本图像，编码器将真实图像转换为潜在空间表示  $z_{\text{r}}$ ，判别器分别对[样本图像,  $z$ ]，[真实图像,  $z_{\text{r}}$ ]进行预测，最终得到损失函数  $G_{\text{loss}}$ ，根据  $G_{\text{loss}}$  进行梯度计算更新模型参数。

2.5) 周期性采样与保存：在每个指定周期内，模型生成一些样本图像，并保存当前训练阶段生成的图像用于可视化模型的进展。

3) 训练结束保存模型的生成器、判别器和编码器，以便后续使用。

### 3.4 改进方法与创新

相比于 GANs，BiGANs 引入了 VAEs 中的编码器，它不仅在一定程度上让模型的训练更稳定，还在潜在空间进行有监督学习，确保生成样本类别的准确性。同时，首次将 BiGANs 框架应用于垃圾分类，通过图像增广技术缓解了数据稀缺与不平衡问题；显著提升模型在复杂场景及少数类识别上的性能，验证了方法的实用性。

## 4. 实验

本章从 BiGAN 模型增广以及传统数据增广方法两个方面进行实验，其中使用了翻转，噪声注入，色彩空间这三种传统数据增广方法。详细描述了每种方法的具体参数设置、优化器和损失函数的使用及使用原因、预处理流程、如何使用 Python 的库构建模型的网络结构、模型训练所使用的主要函数以及流程。以垃圾分类器的分类准确率作为评价指标，进行多方面的对比实验，结果分析。如果使用增广数据的分类器准确度提高了，则说明增广的数据是有效的。

### 4.1 数据集

TrashNet 数据集<sup>[17]</sup>包括六个类：Glass, Paper, Cardboard, Plastic, Metal, and Trash.一共有 2527 张图像，其中 501 张 Glass，594 张 Paper，403 张 Cardboard，482 张 Plastic，410 张 Metal，137 张 Trash。

### 4.2 评价指标

本实验采用分类准确率来评估基于 BiGAN 模型增广与传统数据增广方法对垃圾分类器性能的影响。

分类准确率是最直接且广泛使用的评价指标之一，特别是在监督学习任务中，如本论文中的图像分类任务。分类准确率的定义为模型正确分类的样本数占总测试样本数的比例。其计算公式为：准确率 = 正确分类的样本数 / 总测试样本数 \* 100%。

对于垃圾分类器而言，准确率高意味着模型能够更有效地识别并区分不同类别的垃圾，是衡量模型分类性能的核心标准。在本次实验中，通过比较增广前后的准确率变化，直观反映了数据增广技术对垃圾分类模型泛化能力的提升效果。若采用增广数据后，分类准确率显著提高，这直接证明了增广数据的有效性，即通过引入更多样化的训练样本，模型能够学习到更加鲁棒的特征表示，从而在未见过的数据上表现得更好。

### 4.3 BiGAN 模型实现

#### 4.3.1 参数设置

输入图像规格：图像尺寸设定为  $64 \times 64$ ，具备 3 个颜色通道（RGB），确保模型能够处理彩色图像数据。

潜在空间维度：选择  $4 \times 4 \times 256$  的潜在空间维度是为了与编码器中 AlexNet<sup>[15]</sup>网络结构的输出对应。

训练周期 (Epochs)：模型训练设定为 4000 轮迭代，确保模型有充分机会遍历整个训练集，学习到数据的复杂模式。

批次大小 (Batch\_size)：每一轮迭代处理 32 个样本，这是一种常见的选择，既保证了计算效率，又能较好地估计梯度。

采样间隔 (Sample\_interval)：每 40 个批次保存一次生成样本，用于监控训练进度和模型性能，同时便于视觉检查生成样本的质量。

#### 4.3.2 优化器及损失函数

Adam<sup>[18]</sup> 优化器：Adam 优化器结合了动量项和 RMSProp 的优点，能动态调整每个参数的学习率，这对于非平稳目标函数特别有效。设置的学习率为 0.0002，Beta1（动量项的指数衰减率）为 0.5，Beta2（二次矩的指数衰减率）为 0.999。

损失函数：二元交叉熵损失广泛应用于分类任务，因为它能够衡量模型预测概率分布与真实标签之间的差异。在这里，它用于判别器，以优化其区分真实样本与生成样本的能力，以及在联合模型中，确保生成样本的类别准确性。

#### 4.3.3 模型构建及训练

使用 Keras 构建模型：通过 Sequential API，根据上述结构逐层搭建生成器 G，判别器 D，编码器 E。每层的定义清晰，便于理解和调试。Model.summary() 方法用于展示模型的结构概览，包括各层的参数量和模型总参数量。

训练：使用 Train\_on\_batch() 进行损失函数，模型参数梯度的计算，并使用 Adam 优化器根据计算梯度更新参数。通过构建生成器、判别器、编码器的联合模型来更新生成器和编码器的参数。其中，通过编码器 E 和生成器 G 的联合训练，潜在空间被组织得更为有序，每个区域倾向于对应特定的类别特征。这意味着从特定区域采样的潜在向量通过 G 生成的图像将很可能属于该类别，从而在类别生成上具有高度的一致性。

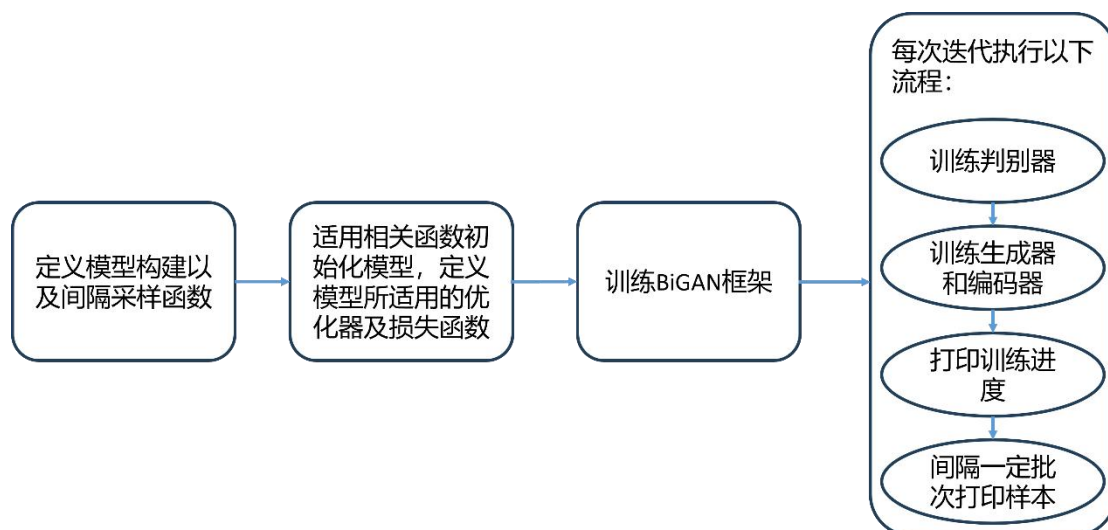


图 8：BiGAN 框架训练流程图

## 4. 4 垃圾分类器

### 4. 4. 1 网络结构

本实验采用的垃圾分类器网络模型在架构上借鉴了经典的 LeNet-5<sup>[19]</sup> 结构，但做了一些微调，具体包括 7 层网络结构（不含输入层）：其中包括 2 个卷积层、2 个降采样层（池化层）、2 个全连接层和输出层。

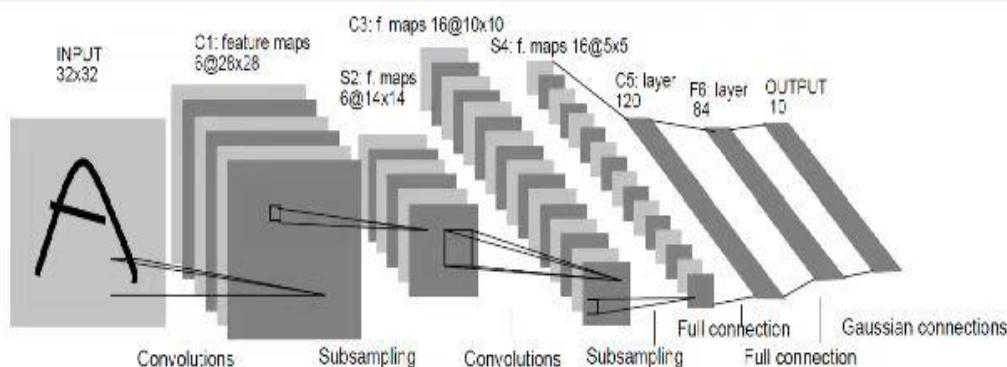


图 9：LeNet-5 结构图

### 4. 4. 2 参数设置

**输入图像大小：**将每个输入图像预处理为  $32 \times 42 \times 3$ ，确保与网络结构兼容。

**训练周期（Epochs）：**设置为 50，充分迭代，确保模型充分学习数据分布并观察收敛趋势。

**批量大小（Batch\_size）：**每批次处理 6 张图像，平衡内存占用与速度，每批更新频率，确保梯度量。

采样间隔（`Sample_interval`）：每 49 个批次计算当前模型的测试及准确率，方便观察模型性能。

#### 4.4.3 优化器及损失函数

优化器：采用随机梯度下降优化器<sup>[20]</sup>，其中学习率（用于控制每次参数更新的步长，即参数沿着梯度方向更新的幅度）为 0.001，动量 **Momentum**（用于加速优化过程，帮助参数在更新时保持一定的惯性，避免陷入局部最优解）为 0.9，平衡快速收敛与避免过拟合同时优化效率。

损失函数：交叉熵损失函数，适合分类任务，衡量预测概率分布与真实标签差距，促进类别区分。

#### 4.4.4 模型训练

在每次迭代开始前，通过 `optimizer.zero_grad()` 清空梯度，准备记录新的梯度量。输入图像到模型进项前向传播得到输出，计算输出与标签的损失函数损失，计算损失的梯度，优化器根据执行 `optimizer.step()` 更新模型参数。

### 4.5 对比实验

- 1) 使用原数据集训练垃圾分类器。
- 2) 使用训练好的 BiGANs 模型对原数据集进行图像增广，并将增广后的数据集用于垃圾分类器的训练，对比增广前后数据集在垃圾分类器的测试准确率。
- 3) 对原数据集进行翻转增广（随机水平或垂直翻转），并将增广后的数据集用于垃圾分类器的训练，对比增广前后数据集在垃圾分类器的测试准确率。
- 4) 对原数据集进行噪声注入增广（从高斯分布中取噪声值，使用 `imgaug` 来实现），并将增广后的数据集用于垃圾分类器的训练，对比增广前后数据集在垃圾分类器的测试准确率。
- 5) 对原数据集进行色彩空间增广（使用 `imgaug` 同时对亮度、对比度、饱和度和色调做改变），并将增广后的数据集用于垃圾分类器的训练，对比增广前后数据集在垃圾分类器的测试准确率。
- 6) 对比各种增广方法增广后测试集在垃圾分类器上的准确率。

## 4. 6 结果与分析

### 4. 6. 1 视觉角度

#### 1) BiGAN 模型增广

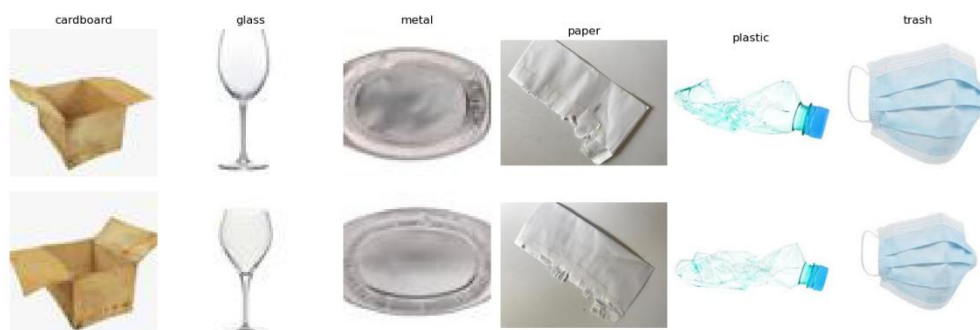


图 10: BiGAN 增广, 原图像 (第一行) 和增广图像 (第二行)

Cardboard 类和 Paper 类的增广图像和原图像差距较大, 颜色、纹理都有所改变, 明显增加了该类数据的多样性; 其他三类在颜色上相对单调且相近, 改变不大。

#### 2) 翻转增广



图 11: 翻转增广, 原图像 (第一行) 和增广图像 (第二行)

Cardboard 类属于垂直翻转增广, 其他五类属于水平翻转增广, 增广前后物体本身并没有发生实质性的变化, 只是方向发生了翻转。这有助于让模型学习到物体的不变性特征, 从而提高其对不同视角下的物体识别能力。

#### 3) 噪声注入增广



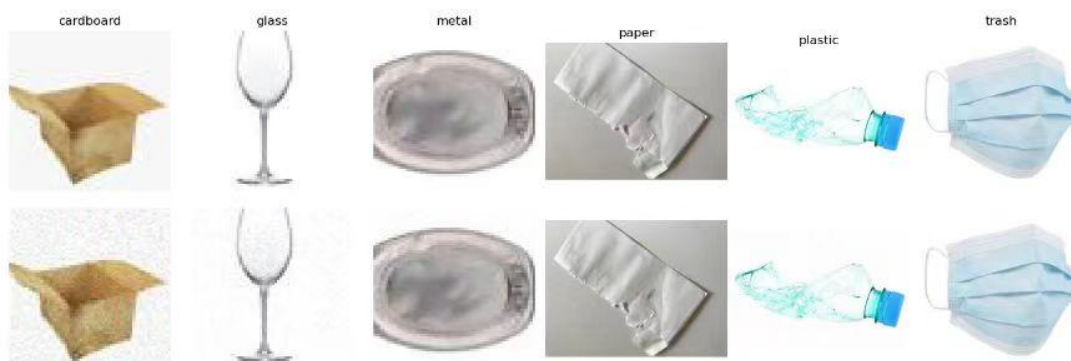


图 12：噪声注入增广，原图像（第一行）和增广图像（第二行）

Cardboard 类和 Glass 类通过高斯噪声注入后，明显变得模糊，其他四类受影响相对较少。但总的来说，这些图像在经过噪声注入后，整体质量都有所下降，部分区域变得模糊不清或者出现噪点。这有助于让模型学习到图像在受到噪声干扰情况下的表现，从而提高其在实际应用中的鲁棒性。

#### 4) 色彩空间

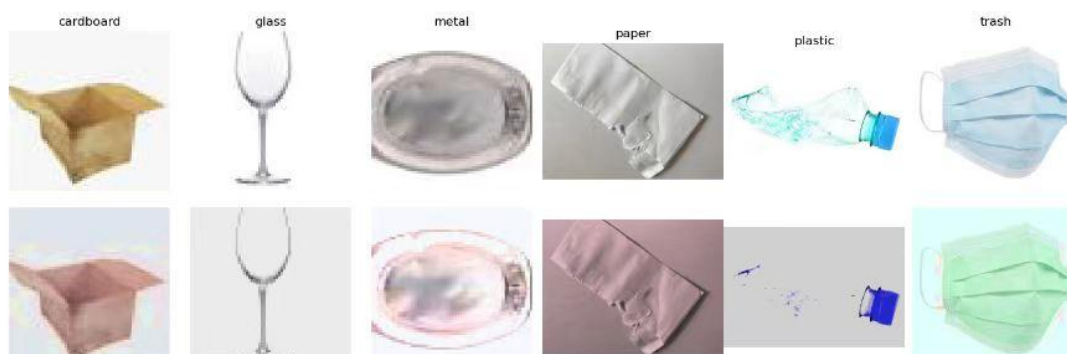


图 13：色彩空间增广，原图像（第一行）和增广图像（第二行）

这些图像经过色彩空间增广后，亮度、对比度、饱和度、色调都有所改变，这有助于让模型学习到不同色彩空间下物体的表现，从而提高其在实际应用中的鲁棒性和泛化能力。

### 4.6.2 数据角度

#### 1) BiGAN 模型

表 1 BiGAN 增广后测试集准确率

类别	原数据集	增广后的数据集
Cardboard	59%	81%
Glass	63%	65%
Metal	52%	56%

Paper	46%	75%
Plastic	44%	47%
Trash	25%	38%
Total	51%	62%

相较于仅使用原始数据集的垃圾分类器，图像增广后的准确率显著提升，尤其在少数类别识别上表现出色。其中，Cardboard 提升了约 37%，Glass 提升了约 3%，Metal 提升了约 8%，Paper 提升了约 63%，Plastic 提升了约 7%，Trash 提升了约 52%，总体提升了约 22%，Paper 类提升最大，纸张可能具有丰富的纹理、颜色和形状，使用图像增广技术可以帮助模型更好地学习这些细微的特征，因此提升最大；Glass 类提升最小，玻璃类别可能相对较简单，其特征可能更加明显和一致，因此图像增广的影响可能相对较小，提升也较小。最终结果表明基于 BiGANs 的图像增广技术大大提升了分类器的泛化性，并在垃圾分类领域能有效缓解样本不均衡和数据稀缺的问题。

## 2) 翻转增广

**表 2 翻转增广后测试集准确率**

类别	原数据集	增广后的数据集
Cardboard	59%	53%
Glass	63%	56%
Metal	52%	56%
Paper	46%	64%
Plastic	44%	37%
Trash	25%	29%
Total	51%	52%

整体来看，翻转增广后，测试集的准确率从 51%略微上升到了 52%，表明翻转增广可能对某些类别有所帮助，但也可能对其他类别产生负面影响，具体的增益或损失因类别而异。从表格中可以看到，Paper 类别的准确率从 46%显著提高到了 64%，而 Plastic 类别的准确率却从 44%下降到 37%。Trash 类别的准确率也仅小幅提高。因此，翻转增广的效果可能因类别而异，需要进一步探索和优化。

## 3) 噪声注入增广

**表 3 噪声注入增广后测试集准确率**

类别	原数据集	增广后的数据集
Cardboard	59%	41%
Glass	63%	65%
Metal	52%	58%
Paper	46%	51%
Plastic	44%	44%
Trash	25%	31%
Total	51%	51%

噪声注入增广后，测试集的准确率基本维持不变，仍为 51%。这意味着噪声注入增广并未对垃圾分类器的性能带来提升。**Cardboard** 类别的准确率从 59% 大幅下降到 41%，可能是由于噪声注入导致模型无法正确识别原本容易辨识的纸板特征。**Glass** 类别的准确率从 63% 略微上升到 65%，可能是因为噪声注入增强了模型对不同光照条件下的玻璃物品的理解。**Metal** 和 **Paper** 类别的准确率都有所提高，分别为 52% 到 58% 和 46% 到 51%，可能是因为噪声注入提高了模型对金属和纸张纹理的鲁棒性。**Plastic** 类别的准确率保持不变，可能是因为噪声注入对该类别的影响相对较小。**Trash** 类别的准确率从 25% 提高到 31%，这可能是因为噪声注入使得模型更好地应对各种复杂的垃圾场景。

#### 4) 色彩空间

**表 4 色彩空间增广后测试集准确率**

类别	原数据集	增广后的数据集
Cardboard	59%	64%
Glass	63%	67%
Metal	52%	54%
Paper	46%	61%
Plastic	44%	38%
Trash	25%	27%
Total	51%	55%

色彩空间增广后，测试集的准确率从 51% 提升到 55%，这表明色彩空间增

广有效地提高了模型的泛化能力。其中除了 Plastic 类，其他五类都有所提升。

总的来说，BiGAN 模型增广图像不论在视觉还是数据方面都优于其他三种传统的增广方式，这是因为 BiGAN 能够自适应地学习数据的潜在分布，意味着它可以根据数据本身的特性动态调整增广策略，而不依赖于手动设计的规则（如特定轴的翻转、固定的色彩调整值等）。在这三种传统的增广方式中，色彩空间增广在该数据集上的表现最好，噪声注入增广最差，但测试集准确率没有降低，意味着并没有降低模型的性能。

## 5. 总结与展望

本论文针对垃圾分类领域中数据稀缺与样本不均衡的挑战，提出了一种基于双向生成对抗网络的图像增广策略。研究通过创新性地应用 BiGANs，成功生成大量多样化的垃圾图像样本，增强了数据集的丰富性和平衡性。实验结果显示，采用此增广数据集训练的分类器，相较于仅使用原始数据集的模型，准确率约提升了 22%，尤其是在少数类别识别上表现出色。生成样本的视觉和量化评估证明了其高保真度和类别一致性，验证了所提方法的有效性和泛化力。

本研究不仅解决了垃圾分类数据获取难的实际问题，还推动了深度学习技术在面临数据不足情况下的应用边界。通过引入编码器实现潜在空间的有监督学习，确保了生成样本与类别标签的高度匹配，这是对传统图像增广和标准 GANs 应用的一大改进。此外，研究通过实际应用展示了 BiGANs 在复杂图像分类任务中的强大潜力，特别是在提升模型对复杂和细节图像的理解能力上。

论文的创新点在于：1) 首次将 BiGANs 框架应用于垃圾分类，通过图像增广技术缓解了数据稀缺与不均衡问题；2) 引入编码器进行潜在空间有监督学习，确保生成样本类别精确性；3) 显著提升模型在复杂场景及少数类识别上的性能，验证了方法的实用性。

未来可进一步探索 BiGANs 在更多类别和精细分类上的应用，如条件化生成；优化增广后数据的选取机制以提高训练效率；以及拓展至多模态图像增广，如视频序列，增强动态垃圾识别能力。

综上所述，本研究通过创新性地运用 BiGANs 图像增广技术，有效应对了垃圾分类中的数据难题，不仅对模型性能提升做出了实质贡献，也为同类研究提供了新视角和方法论参考，促进了深度学习技术在数据受限环境下的应用发展。

## 参考文献

- [1] Mikołajczyk A, Grochowski M. Data augmentation for improving deep learning in image classification problem[C]. 2018 international interdisciplinary PhD workshop (IIPhDW). IEEE, 2018: 117-122.
- [2] Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images[R]. Technical report, University of Toronto, 2009.
- [3] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database[C]. 2009 IEEE conference on computer vision and pattern recognition, IEEE, 2009, 248-255.
- [4] 徐升平. 基于有限样本的人脸识别方法研究[D]. 四川: 电子科技大学, 2020.
- [5] Deng L. The mnist database of handwritten digit images for machine learning research [best of the web][J]. IEEE Signal Processing Magazine, 2012, 29(6): 141-142.
- [6] Netzer Y, Wang T, Coates A, et al. Reading digits in natural images with unsupervised feature learning[C], NIPS workshop on deep learning and unsupervised feature learning. 2011, 2011(5): 7.
- [7] 张家灏. 基于高分辨率神经网络的海洋中尺度涡识别方法研究[D]. 湖南: 国防科技大学, 2021.
- [8] 范黎. 基于生成对抗网络的图像数据增强技术研究及应用[D]. 浙江: 浙江大学, 2022.
- [9] 郭智刚. 基于循环 GAN 网络的细胞图像分割算法研究[D]. 北京: 北京交通大学, 2021.
- [10] Moreno-Barea F J, Strazzera F, Jerez J M, et al. Forward noise adjustment scheme for data augmentation[C]. 2018 IEEE symposium series on computational intelligence (SSCI). IEEE, 2018: 728-734.
- [11] Dua D, Graff C. UCI machine learning repository[OL]. URL <http://archive.ics.uci.edu/ml>, 2017.
- [12] 白杨. 机器学习的隐私关键问题研究[D]. 四川: 电子科技大学, 2022.
- [13] Doersch C. Tutorial on variational autoencoders[R]. arXiv preprint arXiv:1606.05908, 2016.
- [14] Donahue J, Krähenbühl P, Darrell T. Adversarial feature learning[R]. arXiv preprint arXiv:1605.09782, 2016.
- [15] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

- [16] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[R]. arXiv preprint arXiv:1511.06434, 2015.
- [17] Yang M, Thung G. Classification of trash for recyclability status[R]. CS229 Project Report, 2016.
- [18] Da K. A method for stochastic optimization[R]. arXiv preprint arXiv:1412.6980, 2014.
- [19] Hinton G E, Srivastava N, Krizhevsky A, et al. Improving neural networks by preventing co-adaptation of feature detectors[R]. arXiv preprint arXiv:1207.0580, 2012.
- [20] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

## 致谢

随着这篇论文的结束,意味着大学本科的学习生涯就要结束了。四年的时间,经历了很多也学会了很多。从大一开始的课程学习到大三的创新实践初次接触科研,不论是学习还是生活都有很多值得回忆的地方。有过懊悔的地方,也有过成功的喜悦。总的来说,这四年的本科生生涯是我人生中最重要的一段经历,将铭记于心。

四年时间不长也不短,除了美好的回忆,也受到过很多人的帮助。首先,要感谢我的导师史玉回老师和赵琪老师。两位老师不论是在科研还是生活上都给予了我无微不至的指导,关怀和帮助。每当遇到困难时,两位老师总会耐心的指导我,教会我解决困难的办法,在这个过程不断引导着我养成良好的工作习惯,让我在这两年有了飞速的成长和长足的进步。两位老师的教育之恩,学生永远铭记于心。师恩浩荡,如沐春风,也祝福两位老师身体健康,工作顺利。

其次,我想感谢我的同学们。感谢创新实践小组的黄德赐,郭希阳,感谢你们在小组项目遇到困难时的帮助,让初次踏上科研道路时不是孤身一人,因为你们让我体验到了团队协作的快乐。感谢陈孙兵,杨子德,陈建宇,汪昱,李辉日常生活中的关心与陪伴,能与大家一起度过这四年难忘的日子让我感到十分开心和荣幸,这将会是我一生中的难忘的回忆。

最后要感谢我的家人,感谢我的父母对我的含辛茹苦的教育和培养,你们的鼓励与支持让我可以自信地面对困难挑战,勇于攀登。

再次感恩一路上的所有遇见,让我的大学本科时光如此弥足珍贵。再见了南科大,再见了我的学生生涯!我将带着我的收获奔赴人生的下一阶段。