

图像增广

STUDENT 1

Student Number 1

STUDENT 2

Student Number 2

STUDENT 3

Student Number 3

摘 要：深度学习是目前机器视觉的前沿解决方案，而海量高质量的训练数据集是深度学习解决机器视觉问题的基本保障。收集和准确标注图像数据集是一个极其费时且代价昂贵的过程。随着机器视觉的广泛应用，这个问题将会越来越突出。所幸的是，图像增广的出现提供了大量可能性。图像增广技术是一种有效解决深度学习在少量或者低质量训练数据中进行训练的一种技术手段，该技术不断地伴随着深度学习与机器视觉的发展。本文对目前深度学习中的图像数据增广方法进行研究综述，梳理了目前深度学习领域为缓解模型过拟合问题而提出的各类图像增广方法，包括单数据变形、多数据混合、学习数据分布和学习增广策略等方法，接着以图像数据为主要研究对象，对各类算法进一步按照核心思想进行细分，并对方法的原理、适用场景和优缺点进行比较分析。最后，对现有图像增广方法进行了总结。相信未来在该领域的研究有更广阔的发展前景。

关键词： 图像增广；深度学习；机器视觉；生成对抗网络

目录

1 绪论.....	1
1.1 课题研究背景.....	1
1.2 本文的研究内容.....	1
2 单数据增强.....	2
2.1 几何变换.....	2
2.2 色域变换.....	2
2.3 清晰度变换.....	3
2.4 噪声注入.....	4
2.5 局部擦除.....	5
2.6 本章小结.....	5
3 多数据混合.....	6
3.1 图像空间的数据混合.....	6
3.2 特征空间的数据混合.....	7
3.4 本章小结.....	7
4 学习数据分布.....	8
4.1 基于生成对抗网络的图像增强方法.....	8
4.1.1 DCGAN	8
4.1.2 BAGAN	9
4.1.3 DAGAN	9
4.1.4 CycleGAN.....	10
4.2 基于自动编码器和生成对抗网络组合的图像增强方法.....	10
4.3 挑战.....	11
4.4 小结.....	12
5 智能图像增广.....	13
5.1 元学习.....	13
5.2 强化学习.....	13
5.3 在线超参数优化技术.....	14
5.4 特点.....	14
5.5 小结.....	14
6 总结.....	15
参考文献.....	18

1 绪论

1.1 课题研究背景

图像增广技术，就是对现有的数据进行一些图形学或者几何学上的图像变换。图像增广技术是在有效训练数据受限的情况下解决深度学习模型训练问题的一种有效方法。大量的增广技术和方法被提出来丰富和增广训练数据集，提升神经网络的泛化能力^[1]。本文就图像增广的具体方法进行了系统的讨论，并指出方法对应的实际应用场景是什么。

图像增广的提出是为了解决训练过程中的过拟合问题，这个方法旨在从过拟合问题的根源——训练样本不足，来解决该问题^[2]。

1.2 本文的研究内容

如今的图像增广技术，可以细分为“单数据增强”、“多数据混合”、“学习数据分布”、“学习增强策略”四个方向^[2]。我们将基于这四个方向，具体讨论我们现有的图像增广技术的具体实现方法，对各类方法的原理、适用的场景以及优缺点进行分析。最后，我们会对数据增强方法进行总结，并列举他们的应用场合。

2 单数据增强

单数据增强指以单个数据自身为操作对象，通过各种变换操作改变原始数据的表现形态，以产生不同于原始数据的新数据。

2.1 几何变换

① 定义

几何变换是最常见的图像增广方法，包括：旋转、镜像、平移、剪切、缩放、扭曲等。

② 应用场景

需要在数据集上增加视角、位置方面的鲁棒性的场景。

③ 注意事项

要根据数据的特点进行几何变换，否则适得其反。

例如手写字体识别数据集 MNIST，轻度的旋转、平移、缩放可以不改变图像的语义标注，但如果将数字“6”上下翻转则会完全改变图像原本语义标注，给模型引入错误信息导致精度下降。这种情况同样适用于 CIFAR-10、ImageNet 这种拍摄显示生活物体、场景的数据集。而如果对于遥感器拍摄地面的图像，任意角度的旋转、任意方向的翻转以及轻微的平移、缩放、裁剪都可以保留图像的语义信息。

④ 实验证明

Taylor 等人^[3]在包含 101 类物体的图像数据集 Caltech101 上的实验结果显示，图像裁剪带来了 13.82% 的精度提升，远大于旋转和镜像，由此可见对于生活中的物体图像，由于旋转和镜像的程度有限，可增加的图像数量有限，相比之下，裁剪可以得到更多的图像，并且可以为模型引入物体在不同视角位置的更多信息。

2.2 色域变换

① 定义

在图像各通道上进行亮度变换，对数据集增加各种的光照亮度偏差，生成

新的样本。

② 应用场景

在数据集上增加在不同光照下的鲁棒性。

③ 方法

1) 分段线性变换。将 R、G、B 三个通道分离，分别与两个 0 值矩阵组合，呈现红、绿、蓝单一颜色的三幅图像

2) 基于颜色直方图变换

3) 颜色抖动。通过几种颜色组合模拟出大范围内多色彩模式的图像增广

4) PCA 抖动。通过给协方差矩阵增加噪音，实现一种图像在视觉表现上的滤镜效果。对原图像进行主成分分析 PCA，求得协方差矩阵，然后对主成分的特征值施加一个均值为 0 的随机抖动，再变换回去。在 AlexNet^[4]中施加的是均值为 0、方差为 0.1 的高斯扰动。

5) 其他色彩空间表示。HSV(hue, saturation, value)，YUV(Y 为亮度，UV 为存储色度)，CMY(cyan, magenta, yellow)。

④ 注意事项

色域变换与几何变换存在相同问题，需要按照数据的特征进行变换，否则会失去重要的颜色信息，进而改变图像原有的语义信息。

⑤ 实验证明

Taylor^[3]等学者的图像增广实验也做了色域变换的方式，实验证明，对于图像分类任务，空间几何信息比色彩信息更重要。

表 2.1 不同数据增强下的实验结果^[3]

数据增广方法		Top-1 精度/%	Top-5 精度/%
基准线（无增广）		48.13±0.42	64.5±0.65
几何变换	镜像	49.73±1.13	67.36±1.38
	旋转	50.80±0.63	69.41±0.48
	裁剪	61.95±1.01	79.10±0.80
	颜色抖动	49.57±0.53	67.18±0.42
色域变换	边缘增强	49.29±1.16	66.49±0.84
	PCA 抖动	49.41±0.84	67.54±1.01

2.3 清晰度变换

① 定义

通过核滤波器^[5]（ $n \times m$ 的矩阵）对图像进行卷积操作，实现图像的锐化和模

糊处理，改变图像视觉清晰度，产生新的样本。

② 应用场景

在数据集上增加在不同清晰度下的鲁棒性，例如运动模糊等，同时也更加突出物体的细节信息。

③ 方法

1) 高斯模糊滤波器。使得图像变得模糊。

2) 高对比度的水平或垂直的边缘滤波器。使图像边缘锐化。

3) PatchShuffle 正则化^[6]。对图像内部中局部信息进行置换，打乱局部空间排列。

④ 注意事项

因为操作和 CNN 非常类似，所以一般直接集成于卷积网络中，不单独处理。

2.4 噪声注入

① 定义

在图像上叠加噪声，噪声为服从某分布的随机矩阵，生成新样本。

② 应用场景

在数据集上增加在不同质量图像的鲁棒性。

③ 方法

高斯噪声、瑞利噪声、伽马噪声、均匀噪声、椒盐噪声。

④ 注意事项

仅适用于数据过拟合的情况，如果模型在训练数据时欠拟合，那么噪声注入的图像并不能为模型带来新的有效信息

⑤ 实验证明

Ma^[7]等学者在 30 类遥感图像场景数据集 (aerial image dataset, AID) 上，采用高斯噪声、椒盐噪声等噪声注入方式对训练数据进行数据增强。实验结果表明，噪声注入对于遥感场景分类这样的复杂任务，并没有带来明显的精度提升。

⑥ 目前发展

对抗攻击：某些旨在干扰机器学习模型决策的噪声叠加在图像上，形成“对抗样本”，使得图像虽然看起来与原来图像几乎没区别，但是模型会高置信度错分图像，这种输入微小扰动使得分类器分类错误的行为成为“对抗攻击”。

为防止对抗攻击，可以利用噪声注入数据增广的方法进行“对抗训练”，Li^[8]

等学者在实验中发现，对抗训练虽然不能提高模型对于普通样本的测试精度，但是可以显著提高对于对抗样本的测试精度。

2.5 局部擦除

① 定义

将图像局部区域的像素值信息丢弃，生成新的样本。通过局部遮挡的方式，迫使模型学习更宽广的具有描述性质的特征，防止过拟合于特定的视觉特征。

② 应用场景

在数据集上增加在有遮挡情况下的鲁棒性。

③ 方法

Zhong^[9]等学者受到 dropout 的启发，采用随机或人为设计方式丢失局部图像信息。

④ 注意事项

局部擦除可能会篡改图像原有的语义，例如手写字体识别 MNIST 中的数字“7”可能由于局部信息丢失变为“1”，导致图像无法被正确识别。局部擦除的数据增强方法需要人为干预来保证有效性

2.6 本章小结

本章主要介绍了单数据增强的常见五种类型以及各种不同类型的定义、应用场景、方法、注意事项、实验证明和目前发展等。其中单数据增强主要包括五种类型：几何变换、色域变换、清晰度变换、噪声注入以及局部擦除。

3 多数据混合

多数据混合是从图像空间或特征空间将多幅图像的信息进行混合以产生新的训练数据。

3.1 图像空间的数据混合

① 定义

对多幅图像进行线性叠加或非线性混合。

② 应用场景

提高模型的泛化能力，减少模型对错误标签的记忆。

③ 方法

1) 基于线性叠加方式

以 SamplePairing^[10]为例，从训练集中随机抽取两幅图像，分别进行单数据变形增广操作（如裁剪、翻转等），然后用取像素均值的方式叠加两幅图像形成新的样本，标签为这两幅图像中的其中一个。对于图像的随机抽取，使用 unlimited 类别的抽取方式得到的效果最好。

另外 Zhang^[11]等学者提出 mix up 方式改进 SamplePairing，图像叠加采用以下公式：

$$\hat{x} = \lambda x_i + (1 - \lambda)x_j$$

$$\hat{y} = \lambda y_i + (1 - \lambda)y_j$$

其中 $\lambda \in [0, 1]$ ， x 为图像表示向量， y 为 one-hot 编码方法表示的图像标签。他们采用 mix up 的方式在 5 种数据集上进行实验，实验结果显示，mixup 数据增广方法不仅可以提高深度神经网络模型的泛化能力，而且可以有效减少模型对错误标签的记忆，增加模型对于对抗样本的鲁棒性，甚至可以稳定生成对抗网络的训练。

图像的叠加在人的角度是不合理的，但 Tokozume^[12]等学者受声音信号合成的启发，从“波形混合”的角度解释了图像线性叠加对机器而言是合理的。并且，他们将声音识别的类间学习方法应用到图像，按照随机的比例混合两幅图像，并且训练模型输出这个混合的比例，相比简单地混合图像，这种方法训练的模型可以取得更好的表现。

2) 随机裁剪拼接方式

除了多图像线性叠加混合方式，还有一些随机裁剪拼接的数据增广方式，

也取得非常好的效果。

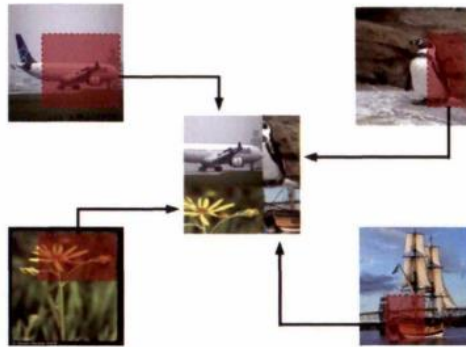


图 3.1 随机裁剪拼接方式

3.2 特征空间的数据混合

① 定义

在 CNN 提取的图像特征基础上进行数据增广。

② 应用场景

为小样本类合成新的样本，解决类别样本数量不均衡的问题。

③ 方法

采用 SMOTE 算法来定义特征空间，将每个样本对应到特征空间中的某一点。对于小样本类中的样本 x ，利用欧氏距离计算 K 个最相邻样本，随机选取其中一个 x_i ，在样本 x 和样本 x_i 的连线段上随机选取一点作为新的样本点，重复该步骤直到类别数据平衡。

⑤ 注意事项

在特征空间中生成新的样本特征向量难以解释，此外，Wong^[13]等学者的研究发现，在数据空间的变换优于在特征空间的变换，这使得特征空间中进行数据混合的方法很少被采用。

3.4 本章小结

本章主要介绍了多数据混合的两种类型：图像空间的数据混合和特征空间的数据混合。接着对这两种类型的多数据混合进行了展开，介绍了其定义、应用场景、方法以及注意事项。

4 学习数据分布

单数据变形类和多数据混合类的数据增广方法操作的对象主要都是单幅图像或者多幅图像，产生新图像时可以利用的先验知识很少，仅仅是图像本身的信息和图像对的互信息。而机器学习中的生成式方法，可以通过训练，学习数据集的潜在概率分布，在数据分布中进行过采样生成新数据，由于将整个数据集作为先验知识，这种数据增广方法在理论上是一种更加优秀的方法。

从数据自身的特点出发，搜索适合不同特点数据集的数据增强策略能够从体系结构搜索的角度重新定义了一种数据增强的新模式^[14]。

4.1 基于生成对抗网络的图像增强方法

生成对抗网络，即 GAN(Generative Adversarial Network)。它包含两个模型，生成模型和判别模型。生成模型的任务是生成看起来自然真实的、和原始数据相似的实例。而判别模型的任务判断给定的实例看起来是自然真实的还是人为伪造的。

在训练过程中，生成网络的目标就是尽量生成真实的图片去欺骗判别网络。而判别网络的目标就是尽量把生成网络生成的图片和真实的图片分别开来。这样，生成网络和判别网络构成了一个动态的“博弈过程”。这样的模型在训练的最后会达到纳什均衡，纳什均衡是指生成模型恢复了训练数据的分布（即伪造出了和真实数据一模一样的样本），判别模型再也判别不出来结果，此时它判别正确的准确率为 50%，约等于乱猜。此时双方网路都得到利益最大化，不再改变自己的策略，也就是不再更新自己的权重^[15]。

研究人员在原始生成对抗网络框架的基础上又提出了多种不同的改进方案，通过设计不同的神经网络架构和损失函数等手段不断提升生成对抗网络的变体的性能。这些变体网络在图像增广领域取得了比较好的效果。下面介绍一些最典型的变体。

4.1.1 DCGAN

① 基本原理

DCGAN 模型在 GAN 的理论基础上，开创性地将 CNN 和 GAN 相结合以实现对图像的处理，并提出了一系列对网络结构的限制以提高网络的稳定性。

它的核心思想有：步幅卷积、全局平均池化、批归一化、使用 Leaky Relu 激活函数^[16]。它的具体原理不是本文的重点，不再详细介绍。我们只需要知道这个方法对于图像数据增强取得了非常不错的效果。

② 效果

Frid-Adar 等学者使用 DCGANs 实现肝脏的数据的增广。它们 DCGANs 在少量标准的样本数据集中合成大量的带标注图像，能够有效地扩大肝部异常检测的训练数据集。

实验表明，训练数据集中加入 DCGANs 合成的图像样本，应用模型在测试数据集中灵敏度和特异度分别提升 7.1 个百分点和 4 个百分点^[17]。

医学领域是一个典型的缺少数据并且标注成本较高的行业。因此，GAN 一般在医学领域的使用比较广泛。

4.1.2 BAGAN

① 基本原理

BAGAN 提出使用已有 3D 模型来生成前景数据，结合前景数据不仅可以降低 GAN 生成图像的复杂度，还可以提升图像的分类质量。它解决了两个主要问题：样本数量较少时，如何更好地训练一个 GAN 网络；如何解决 ACGAN 的两个 loss 函数自相矛盾的问题^[18]。

② 效果

实验结果表明 BAGAN 在 MNIST、CIFAR-10、Flowers 和 GTSRB 四个数据集中，BAGAN 的分类准确率都是高于普通 GAN 网络和 ACGAN 的。所以 BAGAN 可以更好地生成指定类别的样本，实验结果如图 4.1 所示。

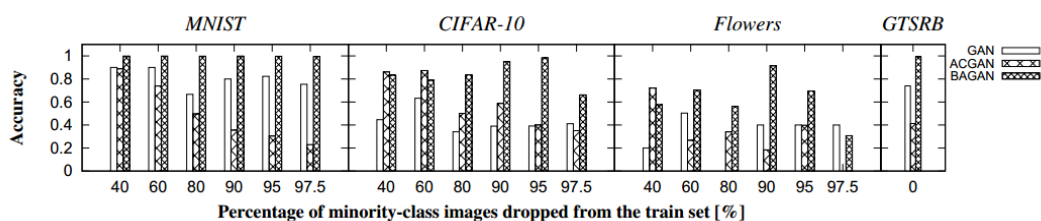


图 4.1 BAGAN 的实验结果

可见，在数据集丢弃比例达到 90% 时，通过 BAGAN 网络图像增强后的图像经过训练后，仍然能达到非常好的效果。

4.1.3 DAGAN

① 基本原理

DAGAN 是一种自监督方案，自动从人脸视频中生成稠密 3D 几何信息，不需要任何标注数据；依据此信息，进一步估计稀疏人脸关键点，用于捕获人头

重要移动；深度信息也用于学习 3D 跨模态（外观与深度）attention 机制，引导生成用于扭曲原图的运动场，依此来获取微表情变化^[19]。

② 效果

DAGAN 可生成高逼真人脸，且在未见过人脸上取得不错效果。

4.1.4 CycleGAN

① 基本原理

风格迁移，也可以说“图到图的翻译”，可以视为一种广义上的图像变换，是一类针对图像的领域迁移问题。本质上是建立一种不同数据分布之间的相互映射。

举个形象的例子，在一张人脸上生成他对应的各种各样不同的表情，就是图像风格迁移的典型应用^[2]。

CycleGAN 的输入只需要两种域的数据，而不需要他们有严格对应关系。这个网络主要目的是实现域适应(Domain Adaptation)。

我们以风景照片和梵高画作为例，假设现在有两个数据集 X 和 Y 分别存放风景照片和梵高画作。我们希望训练出一个生成器 G_1 ，它吃一个风景照，吐出一个梵高画作；同时，我们还希望训练出另一个生成器 G_2 ，它吃一个梵高画作，吐出一个风景照。

为了达到这个目的，我们还需要训练两个判别器 D_1 和 D_2 ，分别判断两个生成器生成图片的好坏：如果生成器产生的图片 y' 不像数据集 Y 里的图片 y ，此时判别器 D_2 应给它低分；反之给他高分。此外，判别器 D_2 应该永远给真实图片 y 高分^[20]。

对于判别器 D_1 也是同理。

② 效果

在人脸表情识别上，可以借助 CycleGAN 进一步生成同一个身份不同表情的图像，实现各类表情数据的增强。在医疗影像处理上，也可以实现同一个人体器官 MR 影像和 CT 影像之间的转换。在遥感领域中，实现同一遥感场景下 SAR 和红外影像与可见光影像之间的转换^[21]。

可见，这种数据增广方法同样具有很大的应用价值。

4.2 基于自动编码器和生成对抗网络组合的图像增强方法

自动编码器通过将其网络结构的一般用于编码，获得图像的低维向量表

示，将网络结构的另一半用于解码，获得根据低维向量表示重新构造的图像数据，这种编码解码的方式能够实现训练数据样本和噪声数据样本的容量扩充，实现了利用数据增强技术提高神经网络的泛化能力和鲁棒性。

本小节介绍一种最经典的模型：CVAE-GAN。

① 基本原理

CVAE-GAN 首先使用编码器将给定的训练图像数据和类别标签编码为符合给定概率分布的隐变量，再通过生成器将从隐变量中采样得到的数据和对应的类别标签生成图像数据。

再将该图像数据输入到分类器和判别器中，从而输入分类标签和判别标签，生成器和判别器构成了一个生成对抗网络，其中生成器尝试通过已经学会了区分真实样本和虚假样本的判别器提供的梯度来学习真实数据分布^[21]。

② 效果

CVAE-GAN 方法能够在多种图像任务中取得较好的性能，包括图像生成任务、图像修复任务等，但在生成位置类别的样本方面还存在一定的可探索性。

该网络能够根据给定的图像在不改变其分类结果的前提下生成与原图像不一样的图像，比如说给定一个人的肖像照，让工具生成不同的这个人的不同表情，虽然表情不同但都是同一个人；或者是同一种鸟类的不同姿态等等。

4.3 挑战

基于“学习数据分布”的图像增广方法需要面对以下挑战。

① 模式坍塌

模式坍塌发生时，生成的样本冗余度大，质量低和样本的差异性小，此时无法达到训练数据增广的目标。尽管有学者做出了很多努力，但是我梦仍然无法从根本上解决这个问题。

② 训练困难

以上模型的训练过程存在梯度突变和梯度消失的问题，训练过程极其不稳定，虽然目前已有部分研究能够缓解该问题，但仍需要结合大量的训练技巧才能获得理想的训练结果。

③ 庞大的计算量。

以上模型的训练过程极其复杂且需要庞大的计算量，限制了其在大尺寸图像数据集上广泛应用。以目前的硬件计算速度而言，计算代价较高。

④ 实用性有待验证。

以上模型都比较复杂，复杂程度甚至超过了原本的任务。过于复杂的模型和训练过程限制了以上模型的图像增广方法不能和基本图像变换增广方法一样作为预处理应用^[1]。

毕竟我们也不希望一个一天就能执行完毕的任务需要花费一个月的时间去进行图像增广的工作。

4.4 小结

本章介绍了“学习数据分布”中常见的模型，并描述了他们的应用场景。用基于本章介绍的模型生成接近真实的虚拟图像样本的方式进行训练集图像增广，为图像增广提供了新的思路。更重要的是，虚拟图像从噪声图像中生成，比真实的图像更具有随机性和多样性。最后，提出了这个方向发展的阻碍，即未来需要攻克的难关。

表 4.1 汇总了当前基于“学习数据分布”图像增广的具有代表性的研究及其效果。

表 4.1 基于“学习数据分布”的图像增广方法汇总

方法	效果
DCGAN	训练肝部扫描图像数据集中加入 DCGANs 合成的图像样本，应用模型在测试数据集中灵敏度和特异度分别提升 7.1 个百分点和 4 个百分点。
BAGAN	在 MNIST、CIFAR-10、Flowers 和 GTSRB 四个数据集中，分类准确性表现比 ACGANs 更优秀
DAGAN	在 Omniglot 数据集，准确率从 69%提升到 82%，提升幅度为 13 个百分点
CycleGAN	摆脱了传统模型需 pix2pix 需要精准配对训练样本的困境，只需要构建两个图像风格领域各自的样本集，可以使用非配对的样本进行训练，极大降低了构建训练样本集的难度
CVAE-GAN	在图像生成任务、图像修复任务取得了比较好的效果

5 智能图像增广

我们在前文中介绍了这么多图像增广的方法，但是应该如何去选择呢？当然不是直接凭直觉乱选，我们可以通过训练一个模型去自适应地选用最优的数据增广策略，来实现模型性能提升的最大化。

在实际应用中，并非所有的增广方式都适用当前的训练数据。

常见的智能图像增广策略有两种。即“元学习”和“强化学习”。

5.1 元学习

元学习希望使得模型获取一种“学会学习”的能力，使其可以在获取已有“知识”的基础上快速学习新的任务。

在机器学习中，训练单位是一条数据，通过数据来对模型进行优化；数据可以分为训练集、测试集和验证集。在元学习中，训练单位分层级了，第一层训练单位是任务，也就是说，元学习中要准备许多任务来进行学习，第二层训练单位才是每个任务对应的数据^[22]。

简单来说，就是我们要在训练过程中找到最佳的增广策略。下图 5.1 给出了基于元学习实现的寻找最佳增广策略的模型图例。

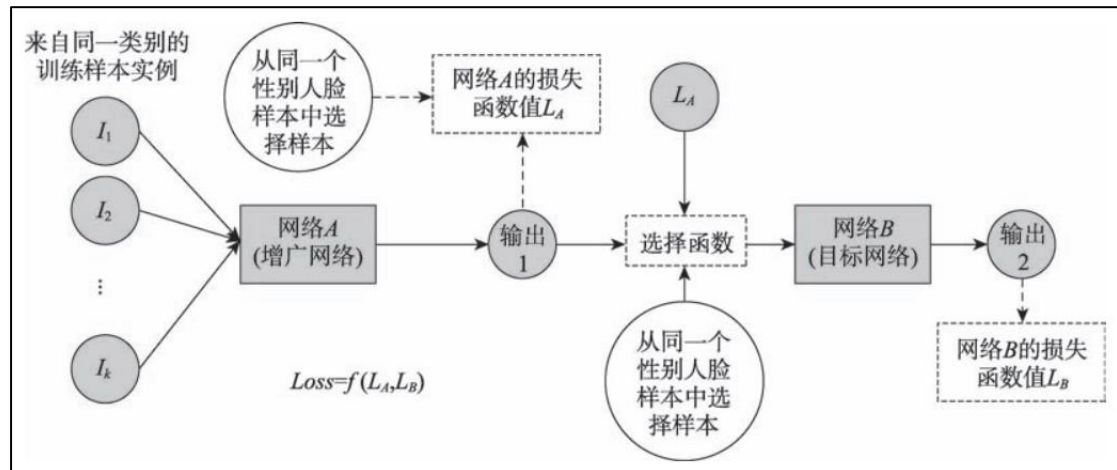


图 5.1 智能图像增广策略：基于元学习

5.2 强化学习

强化学习主要由智能体、环境、状态、动作、奖励组成。智能体执行了某

个动作后，环境将会转换到一个新的状态，对于该新的状态环境会给出奖励信号（正奖励或者负奖励）。随后，智能体根据新的状态和环境反馈的奖励，按照一定的策略执行新的动作。

除了借助神经网络学习增广策略，另一种思路是从给定的图像变换和混合方法中，搜索最优的组合策略^[2]。于是便可以通过强化学习算法来实现这个穷举搜索的过程，直到找到最佳的图像增广方法组合。

5.3 在线超参数优化技术

Lin 等学者在基于元学习的图像数据增广方法的基础上，提出在线超参数优化技术（online hyper-parameter learning for auto-augmentation, OHL-Auto-Aug），该方法将每种增广策略划分成不同的增广幅度，强、中、弱，从而将策略的连续搜索空间变成离散搜索空间。该方法极大地降低了搜索空间，从而加速了智能增广策略的训练效率^[23]。

5.4 特点

智能图像增广有如下的两个弊端，

首先是增广网络和目标网络同时训练，GPU 计算代价庞大；其次训练及复现困难。迭代 1 000 多次仍未收敛，获得的数据增广策略未必是全局最优。

5.5 小结

本章主要介绍了两种智能图像增广的策略，包括原学习和强化学习。此外，还介绍了一种可以加速智能增广策略训练效率的方法，即在线超参数优化技术。最后，简要介绍了智能图像增广的两个特点。

6 总结

第一章为绪论，主要介绍了课题研究背景和本文的研究内容。

第二章主要介绍了单数据增强的常见五种类型以及各种不同类型的定义、应用场景、方法、注意事项、实验证明和目前发展等。其中单数据增强主要包括五种类型：几何变换、色域变换、清晰度变换、噪声注入以及局部擦除。

第三章主要介绍了多数据混合的两种类型：图像空间的数据混合和特征空间的数据混合。接着对这两种类型的多数据混合进行了展开，介绍了其定义、应用场景、方法以及注意事项。

第四章介绍了“学习数据分布”中常见的模型，并描述了他们的应用场景。用基于本章介绍的模型生成接近真实的虚拟图像样本的方式进行训练集图像增广，为图像增广提供了新的思路。更重要的是，虚拟图像从噪声图像中生成，比真实的图像更具有随机性和多样性。最后，提出了这个方向发展的阻碍，即未来需要攻克的难关。

第五章主要介绍了两种智能图像增广的策略，包括原学习和强化学习。此外，还介绍了一种可以加速智能增广策略训练效率的方法，即在线超参数优化技术。最后，简要介绍了智能图像增广的两个特点。

最后我们将不同数据增广方法优缺点比较总结如下的表格。

① 单数据增强

表 6.1 “单数据”增广方法优缺点比较

类别	代表方法	优点	缺点
几何变换	旋转、镜像、平移、裁剪、缩放和扭曲等	简单易操作，可增加数据集的空间集合信息，提高模型在不同视角、位置方面的鲁棒性	增加的信息量有限，对数据重复记忆，不恰当的操作可能会改变图像原本的语义标注
色域变换	颜色抖动和 PCA 抖动	简单易操作，可增加数据集的色彩信息，提高模型在不同光照条件下的鲁棒性。	增加的信息量有限，对数据重复记忆，可能改变图像中的重要颜色信息。
清晰度变换	高斯模糊、边缘锐化和 Patch Shuffle 等	可提高模型对运动模糊的鲁棒性，可突出物体细节信息	使用滤波方式实现，与 CNN 内部机制重复

续表 6.1 “单数据” 增广方法优缺点比较

类别	代表方法	优点	缺点
噪声注入	高斯噪声、瑞利噪声、伽马噪声、均匀噪声和椒盐噪声等	增强模型对噪声干扰和冗余信息的过滤能力，提高模型对不同质量图像的识别能力	不能增加新的有效信息，对模型精度提升效果不明显
局部擦除	随机擦除、Cutout、Hide-and-Seek 和 GridMask 等	可以增加模型在遮挡条件下的鲁棒性，使模型去学习图像中更加具有描述性质的特征，关注到整幅图像的全局信息	可能篡改原图的语义信息，重要局部信息擦除后可能无法识别图像

② 多数据混合

表 6.2 “多数据” 增广方法优缺点比较

类别	代表方法	优点	缺点
图像维度数据混合	SamplePairing、mixup、Between-Class、非线性图像混合和随机裁剪拼接等	混合了多幅图像的像素值信息	不合常理，缺乏可解释性
特征空间数据混合	SMOTE、特征空间外插值等	融合了多幅图像的特征信息	特征向量难以理解

③ 学习数据分布

表 6.3 “学习数据分布” 增广方法优缺点比较

类别	代表方法	优点	缺点
图像维度数据混合	生成对抗网络	方法和应用众多，主要基于 conditional GAN 生成带标注样本和基于 DCGAN 生成图像	在拟合的数据分布中采样，可无限量生成样本
特征空间数据混合	图像风格迁移	方法和应用众多，主要基于 CycleGAN 在保留图像内容的同时实现风格迁移	可实现相同内容不同模态的图像之间的相互转换，可帮助解决许多领域内的特殊问题

④ 学习增强策略

表 6.4 学习增强策略方法优缺点比较

类别	代表方法	优点	缺点
元学习	OHL-Auto-Aug	用神经网络去替代确定的数据增广方法，训练模型学习更好的增广策略	引入额外的网络，需要额外的训练开销

续表 6.4 学习增强策略方法优缺点比较

类别	代表方法	优点	缺点
强化学习	AutoAugment	组合现有的数据增广方法，搜索最优策略	策略搜索空间大，训练复杂度 高，计算开销大

参考文献

- [1] 林成创,单纯,赵淦森,杨志荣,彭璟,陈少洁,黄润桦,李壮伟,易序晟,杜嘉华,李双印,罗浩宇,樊小毛,陈冰川.机器视觉应用中的图像数据增广综述[J]计算机科学与探索,2021,15(4):583-611
- [2] GB/T 7714 马崇昇,唐娉,赵理君,张正.深度学习图像数据增广方法研究综述[J]中国图象图形学报,2021,26(3):487-502
- [3] Taylor L and Nitschke G. 2017. Improving deep learning using generic data augmentation [EB/OL]. 2017-08-20 [2020-03-03]. <https://arxiv.org/pdf/1708.06020.pdf>
- [4] Krizhevsky A, Sutskever I and Hinton G E. 2017. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6): 84-90 [DOI: 10.1145/3065386]
- [5] Shorten C and Khoshgoftaar T M. 2019. A survey on image data augmentation for deep learning. Journal of Big Data, 6 (1): 1-48 [DOI: 10.1186/s40537-019-0197-0]
- [6] Kang G L, Dong X Y, Zheng L and Yang 2017. PatchShuffle regularization [EB/OL]. 2017-07-22 [2020-03-03].<https://arxiv.org/pdf/1707.07103.pdf>
- [7] Ma D G, Tang P and Zhao L J. 2019. SiftingGAN: generating and sifting labeled samples to improve the remote sensing image scene classification baseline in vitro. IEEE Geoscience and Remote Sensing Letters, 16(7): 1046-1050 [DOI: 10.1109/LGRS.2018.2890413]
- [8] Li S T, Chen Y K, Peng Y L and Bai L. 2018. Learning more robust features with adversarial training [EB/OL]. 2018-04-20[2020-03-03]. <https://arxiv.org/pdf/1804.07757.pdf>
- [9] Zhong Z, Zheng L, Kang G L, Li S Z and Yang Y. 2017. Random erasing data augmentation[EB/OL]. 2017-08-16 , 2020-03-03. <https://arxiv.org/pdf/1708.04896.pdf>
- [10] Inoue H. 2018. Data augmentation by pairing samples for images classification [EB/OL]. 2018-01-09 [2020-03-03]. <https://arxiv.org/pdf/1801.02929.pdf>
- [11] Zhang H Y, Cisse M, Dauphin Y N and Lopez-Paz D. 2017. Mixup: beyond empirical risk minimization [EB/OL]. 2017-10-25 [2020-03-03]. <https://arxiv.org/pdf/>

1710.09412.pdf

- [12]Tokozume Y, Ushiku Y and Harada T. 2018. Between-class learning for image classification/ZProceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 5486-5494 [DOI: 10.1109/CVPH.2018.00575]
- [13]Wong S C , Gatt A, Stamatescu V and McDonnell M I). 2016. Understanding data augmentation for classification: when to warp?//Proceedings of 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA). Gold Coast, Australia: IEEE: 1-6 [DOI: 10.1109/DICTA.2016.7797091]
- [14]基于图像的数据增强方法发展现状综述, [EB/OL], 2012, <https://zhuanlan.zhihu.com/p/486197438>
- [15]Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.
- [16]Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. arXiv preprint arXiv:1511.06434, 2015.
- [17]RID- ADAR M, DIAMANT I, KLANG E, et al. GAN based synthetic medical image augmentation for increased CNN performance in liver lesion classification[J]. Neuro computing, 2018, 321: 321-331.
- [18]Mariani G, Scheidegger F, Istrate R, et al. Bagan: Data augmentation with balancing gan[J]. arXiv preprint arXiv:1803.09655, 2018.
- [19]Antoniou A, Storkey A, Edwards H. Data augmentation generative adversarial networks[J]. arXiv preprint arXiv:1711.04340, 2017.
- [20]Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2223-2232.
- [21]Bao J, Chen D, Wen F, et al. CVAE-GAN: fine-grained image generation through asymmetric training[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2745-2754.
- [22]一文入门元学习, [EB/OL], 2020, <https://zhuanlan.zhihu.com/p/136975128>.
- [23]LIN C, GUO M, LI C M, et al. Online hyper- parameter. learning for auto-augmentation strategy[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Oct 27-Nov 2, 2019. Piscataway: IEEE, 2019: 6579-6588.