

谨以此论文献给我的导师和亲人！

_____XXX

论文题目

学位论文答辩日期： _____

指导教师签字： _____

答辩委员会成员签字： _____

独 创 声 明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含未获得 _____（注：如没有其他需要特别声明的，本栏可空）或其他教育机构的学位或证书使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

学位论文作者签名： 签字日期： 年 月 日

学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，并同意以下事项：

1. 学校有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许论文被查阅和借阅。
2. 学校可以将学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。同时授权清华大学“中国学术期刊(光盘版)电子杂志社”用于出版和编入 CNKI《中国知识资源总库》，授权中国科学技术信息研究所将本学位论文收录到《中国学位论文全文数据库》。

(保密的学位论文在解密后适用本授权书)

学位论文作者签名：

签字日期： 年 月 日

导师签字：

签字日期： 年 月 日

论文题目

摘 要

中文摘要

关键词：关键词 1， 关键词 2

English Title

Abstract

English Abstract

Keywords: Keyword1, Keyword2

目 录

1 绪论	1
1.1 研究背景和意义	1
1.2 国内外研究现状	2
1.2.1 生成对抗网络研究现状	2
1.2.2 图像转译研究现状	2
1.2.3 水下图像合成研究现状	3
1.3 课题来源	3
1.4 论文内容和安排	3
2 基于生成对抗网络的图像转译方法	5
2.1 生成对抗网络	5
2.1.1 生成模型	5
2.1.2 生成对抗网络原理	5
2.2 生成对抗网络问题和发展	6
2.2.1 cGAN	6
2.2.2 DCGAN	7
2.2.3 LSGAN	8
2.2.4 ACGAN	9
2.3 基于生成对抗网络的图像转译	10
2.3.1 成对图像转译方法	11
2.3.2 非成对图像转译方法	12
2.3.3 图像多模态转译	13
2.4 本章小结	17
3 水下图像合成研究	18
3.1 水下图像成像原理及存在问题	18
3.2 基于生成对抗网络的水下图像研究	19
3.2.1 水下图像数据集	19
3.2.2 水下图像评价指标	20
3.2.3 经典水下图像合成方法	21
3.3 基于分解的水下图像研究	22
3.4 本章小结	22

4 水下图像多模态转译模型设计	24
4.1 水下图像多模态转译问题定义	24
4.2 水下图像多模态转译问题分析	24
4.3 建立模型算法和损失函数	24
4.3.1 模型创新点	24
4.3.2 目标函数和算法	24
4.3.3 网络结构	24
4.4 本章小结	25
5 水下图像多模态转译的实验与分析评测	26
5.1 experiment	26
6 总结与展望	27
6.1 全文总结	27
6.2 后续工作展望	27
7 第 6 章	28
7.1 第 6.1 节	28
7.1.1 第 6.1.1 节	28
致 谢	31
个人简历、在学期间发表的学术论文与研究成果	32

1 绪论

1.1 研究背景和意义

计算机视觉领域中，数据是解决一切问题的基础；缺乏海量的语音、图像、视频等数据来进行研究，是现在面临的最大挑战。针对匮乏的数据，具有真实性的数据生成问题一直是遏制计算机视觉发展和亟需解决的问题。从生成对抗网络问世开始，计算机视觉领域打开了一个全新的局面，愈来愈多的工作基于生成对抗网络开展。在图像处理领域，基于生成对抗网络，可以通过训练简单的对抗模型合成大量现实中不存在但是具有真实性的图像数据。

在生成对抗网络快速发展的过程中，图像到图像转译被单独划分成一类问题针对进行解决。原始的生成对抗网络图像转译是指将输入图像“转译”成一副相应的输出图像，输入图像与输出图像之间具有相同的内容但是具有不同的场景特征。类似我们谷歌翻译一句话，翻译器不会改变我们输入的内容和意思，但是可以翻译成英语、法语等其他语种。图像到图像的转译简称为图像转译，目的是将一个图像从当前场景转译到其他场景中，比如转译成一副对应内容的边缘信息图、语义标签图或者特定场景风格图。我们将这种转译视为从源域映射到目标域的转译，输入和输出分别属于两个不同的域。

在研究图像转译问题中，根据可获得的数据集是否包含成对的训练图像可以将解决方法划分为成对的图像转译子问题和非成对的图像转译子问题。当数据集满足包含成对的训练图像时，成对的图像转译问题就可以通过条件生成对抗网络等有监督的方法来进行解决。现实情况更多的是不满足获取到大量的成对数据，或者像水下场景等获取成对训练数据特别困难、代价高昂，非成对图像数据的无监督转译模型被提出。另外，源域和目标域的训练图像数量不平衡，往往一个域中图像数据数量远远超过另一个域图像数据数量时给图像转译带来巨大的难度。

类似谷歌翻译器，不同语种之间的翻译很多情况下并不是一一对应的，往往一个语义内容可以对应多个翻译结果。在图像转译中，需要成对训练数据的有监督图像转译往往是一对一的转译，给定源域输入有真实的目标域生成图像数据；不成对训练数据的无监督图像转译可以进行一对多的转译，网络无法固定学习到特定的映射，因此可以产生多样输出。这个现象说明，一方面无法直接得到特定的输出这方面有缺陷，另一方面意味着我们可以使用一些手段得到多样化的输出结果，实现多模态输出。

图像到图像的转译，目前已经有大量的应用。现实中的很多问题都可以转换

成图像转译问题进行解决，图像转译具有强大的应用价值。可以应用于边框图上色、素描画上色等任务，极大的解放了人力；可以应用于人脸表情漫画化、人体姿态图像编辑等，实现特定风格的图像生成和控制，已经应用于图像美化等商业化软件中；还可以应用于图像超分辨重建任务，提升图像质量；以及在医学图像、水下图像等多种特定学科的常规手段难以直接进行解决的问题中得到应用，给其他学科的图像问题解决提供了新的思路和方法。

在水下环境，一方面数据采集耗费巨大、成本昂贵，另一方面由于水介质存在光的吸收和散射，导致很难获取到大量高质量的清晰水下图像数据。进行水下图像研究时，以水下图像增强和复原为例，许多方法都先通过图像转译模型合成对的水下图像数据，再进行水下图像的增强或者复原操作。可见，同时合成多种水下环境条件这些模态的高质量图像数据对水下图像研究领域意义重大，可以为水下视觉领域提供大量的训练数据，推动水下研究工作的发展和进步。随着深度学习的广泛发展，生成对抗网络开始越来越多的

1.2 国内外研究现状

1.2.1 生成对抗网络研究现状

Goodfellow 在 2014 年提出生成对抗网络 (Generative Adversarial Network, GAN)^[1] 后，计算机视觉领域开始乘风向上。生成对抗网络由生成器和判别器两个神经网络组成，生成器试图生成可以欺骗过判别器的真实性结果，判别器负责区分输入的样本是真实样本还是生成器生成的结果。这样处于对抗关系中的两个神经网络彼此不断优化，使得该网络的合成结果能够生成超越变分自动编码器 (VAE)^[2]、自回归模型 (AR 模型) 等以前生成模型。越来越多的科研人员开始深入研究生成对抗网络，近年深度卷积生成对抗网络 (DCGAN)、条件生成对抗网络 (CGAN) 等许多基础生成对抗网络的变种工作成功在文本到图像生成、图像到图像转译、视频预测等领域应用都发挥了独特而强大的功能。然而，生成对抗网络也有其缺点，譬如存在训练不稳定、会产生模式崩塌等问题，后续的大量变种工作针对这些问题在生成对抗模型上进行解决。

1.2.2 图像转译研究现状

基于生成对抗网络的图像到图像的转译 (Image-to-Image Translation) 首次被 Isola 等在 2017 年的工作 pix2pix (Image-to-Image Translation with Conditional Adversarial Networks, pix2pix)^[3] 中提出，依托条件为图像的条件生成对抗网络的思想，训练图像使用尺寸相同的成对数据，是最典型的有监督图像转译方法。同年，朱

俊彦等提出 cycleGAN (Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, cycleGAN)^[4], 基于生成对抗网络提出的循环一致性损失, 该方法解决了更普遍没有成对训练数据的无监督图像转译问题。自此, 图像转译开始被应用于各种具体问题的解决。

无监督图像转译方法中, UNIT^[5] 将问题定义——实现两个域之间的转译是满足两个域的联合分布, 假设输入到生成器的潜在编码共享, 通过编码器将两个域的图像数据编码到潜在空间中。收分解表达学习的影响, MUNIT^[6] 在 UNIT 提出共享潜在编码的基础上假设潜在编码包括内容编码和风格编码, 可以合成相同内容不同风格的多模态结果。DRIT^[7] 与 MUNIT 思路一致, 在于内容编码和风格编码连接方式上有所区别。以上基于分解的图像转译方法广泛应用于风格转换和多种多模态图像转译任务中。

以上的图像转译方法都是在两个域之间, 若要进行多个域之间的转译需训练多个域间生成器一一对应, StarGAN^[8] 用一个生成器结构学习多个域之间的映射从而实现多个域的转译。StarGAN v2^[9] 将 StarGAN 多域转译工作拓展至多模态, 可以实现多域多模态的图像转译任务。在 DRIT 工作基础上, DRIT++^[10] 被提出, 将两个域之间的转译也拓展至多个域。

1.2.3 水下图像合成研究现状

目前水下图像转译

1.3 课题来源

国家自然科学基金面上项目“类别不平衡条件下海洋浮游生物图像精细识别及其原位应用研究”(批准号: 61771440)、国家自然科学基金面上项目“海洋中小型浮游生物原位光学观测关键技术研究”(批准号: 41776113)。

1.4 论文内容和安排

本工作主要内容和章节安排如下:

第一章为绪论部分, 该部分主要针对基于生成对抗网络的图像转译和水下图像转译的研究背景和意义、国内外研究现状以及课题来源进行论述。

第二章介绍了生成对抗网络的基本概念、基于生成对抗网络的主流算法和基于生成对抗网络应用于图像转译的方法。详细对算法概念和原理、网络模型结构、生成对抗网络及主流算法优缺点以及在图像转译场景上进行了具体介绍。

第三章介绍了水下图像研究。

第四章主要描述了本工作提出的基于生成对抗网络的图像多模态转译模型，随算法流程、网络模型设计思路以及具体的模型实现细节做了详细展现。

第五章介绍了。

第六章对本工作进行贡献总结和讨论，分析研究问题的特点以及提出方法的优点和缺点，对研究过程中固有问题进行分析和展望。

2 基于生成对抗网络的图像转译方法

2.1 生成对抗网络

2.1.1 生成模型

生成模型 (Generative Model) 描述的是这样一类模型：从分布 p_{data} 取样的若干样本构成训练集，经过训练，生成模型可以学习并模拟该概率分布 p_{model} 。有些生成模型的可以通过显式概率分布函数表达出来，而有些通过从分布中采样得到数据等隐式表达。越来越多的情况下，我们需要生成模型 p_{model} 生成以假乱真的高质量目标样本。

实际应用中，主要用到的三种生成模型：一是自回归网络 (Auto-Regressive Networks, AR 模型) 又叫 Fully-visible Bayes networks (FVBN)；二是变分自动编码器 (Variational Autoencoder, VAE)；三是生成对抗网络 (Generative Adversarial Networks, GANs)。这些模型各有其特点，自回归模型在采样过程中效率较低切不容易为图像提供简单的低维编码，变分自动编码器倾向于生成稍微模糊的样本，生成对抗网络训练难以稳定和优化。

本工作研究内容基于生成对抗网络，后续章节将对生成对抗网络的进行详细分析和论述。

2.1.2 生成对抗网络原理

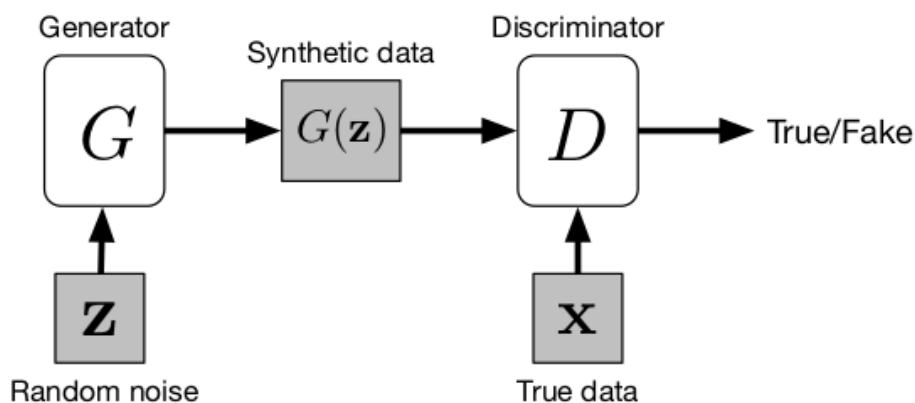


图 2-1 生成对抗网络示意图。

生成对抗网络是 Goodfellow 在 2014 年提出的，生成对抗网络够生成比以往的

生成模型更好的合成图像，自此成为最热门的研究领域之一。生成对抗网络的提出采用二人零和博弈思想，两个神经网络在游戏框架下相互博弈，不断彼此优化。生成对抗网络（GAN）网络结构如图 2-1 所示。由两个彼此独立的神经网络组成：一个生成器（Generator），输入随机噪声向量 z ，输出合成数据结果 $G(z)$ ；一个判别器（Discriminator），输入真实数据 x 或者合成数据结果 $G(z)$ ，判别器试图区分真实数据和合成数据。

$$\min_G \max_D \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (2-1)$$

在训练过程中，生成器和判别器分别单独交替训练，目标函数如式 2-1 所示。优化生成器时判别器固定，生成样本当作真实数据优化， \min_G 最小化生成样本和真实样本之间的差异；优化判别器时固定生成器，把生成样本当作虚假数据进行处理， \max_D 尽可能的让判别器最大化的判别出样本来自真实数据还是生成数据。这是博弈得以进行的关键之处，理想情况下生成分布会拟合于真实分布。

生成对抗网络相比较于之前的生成模型的优势有两点。一是因为不依赖任何先验假设，而传统的方法会假设数据服从某一分布，然后用极大似然去估计分布；二是生成类似真实样本的方法简单，直接利用生成器前向传播进行生成。然而，生成对抗网络也有其不可忽视的缺点。训练中有时不收敛，网络不稳定难以训练；模式坍塌问题，生成的缺少多样性的结果。

2.2 生成对抗网络问题和发展

本节将对生成对抗网络发展过程中具有里程碑意义的变种模型进行介绍，主要创新点针对网络模型和目标函数的改进，以及解决了传统生成对抗网络的问题。

2.2.1 cGAN

原始生成对抗网络不能生成具有特定属性的图像结果，针对这个问题，Mirza 等人提出条件生成对抗网络（Conditional Generative Adversarial Nets, cGAN）^[11]。如图 2-2 所示，核心创新点在于将属性信息 y 融入生成器 G 和判别器 D 中，属性 y 可以是任何标签信息，例如图像的类别信息或者图像本身等。

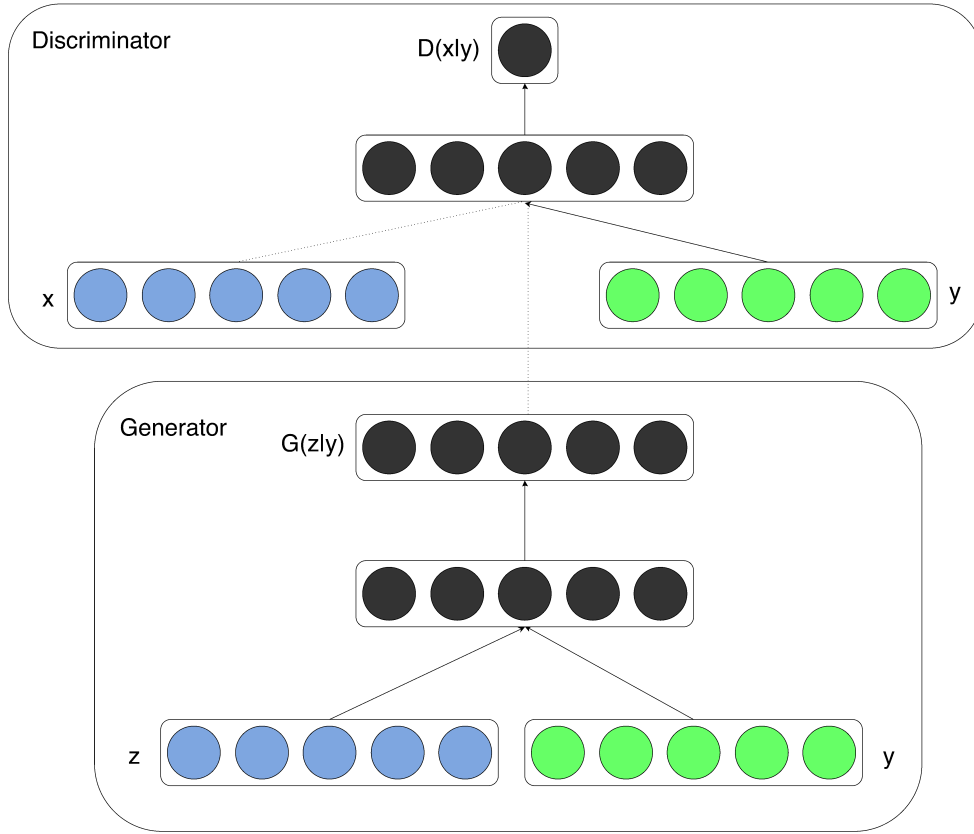


图 2-2 条件生成对抗网络示意图。

$$\min_G \max_D \mathbb{E}_{x \sim p_{data}(x)} [\log D(x|y)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z|y)))] \quad (2-2)$$

引入条件 y 后，在 GAN 基础上的目标函数如式 2-2 所示。当 y 是特定类别的话，合成结果为特定类别的数据，因此 cGAN 可以看作从从无监督学习 GAN 到有监督学习的改进。cGAN 生成的图像虽有很多缺陷，譬如图像边缘模糊，生成的图像分辨率太低等，但是它为后续图像风格转换任务中对属性特征的处理提供了方法思路。

2.2.2 DCGAN

从 Alex Krizhevsky、Ilya Sutskever 和 Geoffrey Hinton 创造大型深度卷积神经网络 AlexNet^[12] 赢得 2012 年 ImageNet^[13] 大规模视觉识别挑战赛 ILSVRC 开始，卷积神经网络 (CNN) 在计算机视觉领域成功亮相。

2016 年，Radford 等提出了深度卷积生成对抗网络 (Deep Convolutional Generative Adversarial Networks, DCGAN)^[14]，使用深度卷积神经网络代替原始 GAN 中

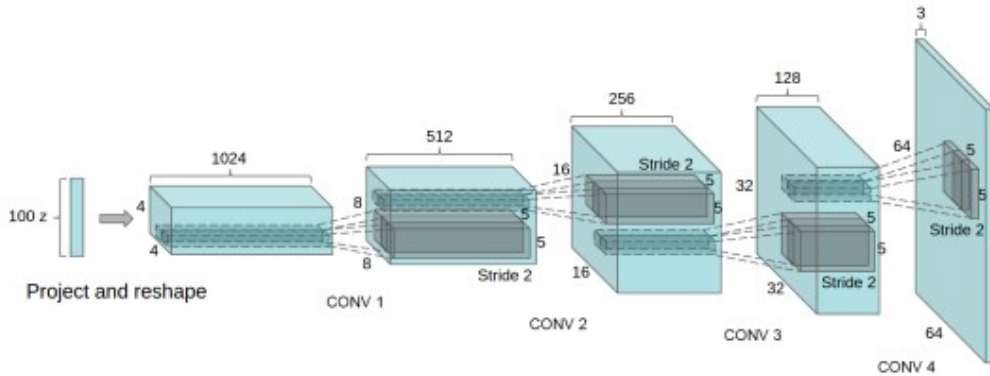


图 2-3 深度卷积生成对抗网络示意图。

全链接层的线性感知，将网络拓展至层数更深、参数更多的任务中，进而提高了生成样本的质量和收敛的速度。

深度卷积生成对抗网络对生成对抗网络的模型结构做了一些改变，如图 2-3 所示。创新点如下：

（1）在网络结构中取消全链接层，使用全卷积网络。生成器和判别器对称存在，极大的提升了 GAN 训练稳定和生成结果质量。

（2）将对抗网络中的池化层取消，使用带步长的卷积层和反卷积层进行上/下采样操作，更好的提取图像特征。

（3）在生成器和判别器中使用批归一化（BatchNorm），为了网络稳定，在生成器输出层和判别器的输入层取消使用批归一化。

（4）生成器中，除了最后一层使用 Tanh 激活函数外，其余都是 ReLU 激活函数。

（5）在判别器中，所有层都使用 LeakyReLU 激活函数，从而防止梯度稀疏。

深度卷积生成对抗网络对生成对抗网络所做的创新，虽然一定程度上稳定了训练，生成较高质量的结果，但是并没有从根本上解决生成对抗网络训练不稳定的问题，在训练过程中仍需要注意生成器和判别器的平衡问题以及模式坍塌问题的出现。

2.2.3 LSGAN

LSGAN^[15] 针对 GAN 生成图像质量不高以及训练过程不稳定的问题进行改进，具体来说就是将 GAN 的目标函数由交叉熵损失换成最小二乘损失。

准确的决策边界会穿过真实数据点，把决策边界当作中介，生成图像结果和真实数据之间的距离可以由生成结果和决策边界之间的距离来反映。如图 2-4 所示，蓝色十字形状表示假样本数据，红色空心圆圈是真样本数据，蓝色线条是原始

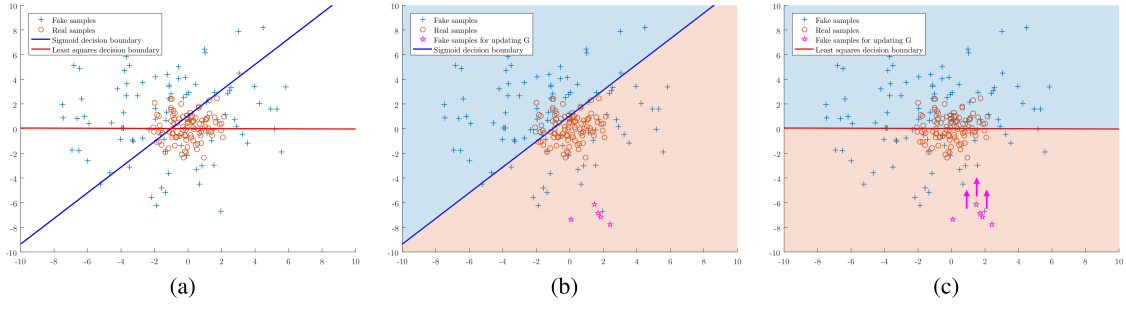


图 2-4 GAN 中 Sigmoid 决策边界 LSGAN 最小二乘决策边界图示。

GAN 中 Sigmoid 决策边界，红色线条是 LSGAN 中最小二乘决策边界。(a) 是两个损失函数决策边界对比图。在 (b) 和 (c) 中，枚红色五星表示更新生成器过程中的加样本数据。如 (b) 中所示，橘色是真实数据区域，蓝色是加样本区域。若使用交叉熵作为损失，当交叉熵损失很小时，生成器不会再优化那些被生成器识别为真实图片的生成结果，事实上此时这些生成结果距离决策边界也就是真实数据分布依旧很远，生成器却也不再进行优化。如 (c) 当损失函数为最小二乘损失时，要想最小化最小二乘损失，生成器必然会将距离决策边界也就是真实数据分布的生成结果拉向决策边界。

LSGAN 目标函数定义如式 2-3 和 2-4，其中参数设置 $a = c = 1$ ， $b = 0$

$$\min_D \frac{1}{2} \mathbb{E}_{x \sim p_{data}(x)} [(D(x) - b)^2] + \frac{1}{2} \mathbb{E}_{z \sim p_z(z)} [D(G(z)) - a]^2 \quad (2-3)$$

$$\min_G \frac{1}{2} \mathbb{E}_{z \sim p_z(z)} [D(G(z)) - c]^2 \quad (2-4)$$

2.2.4 ACGAN

CGAN 通过结合标签信息来提高生成数据的质量，ACGAN^[16] 除此之外，还通过重建标签信息来提高生成数据的质量。模型结构如图 2-5 所示，生成器的输入是分类标签和固定的噪声分布，判别器的输出除了指示真假的对抗损失外，还出增加一个辅助分类器判别分类标签，输入的标签当作目标进行反馈，提高判别网络的学习效果。

目标函数包含两个部分：判别真实性的对抗损失 L_S 如式 2-5 和类别判别损失 L_C 如式 2-6。其核心贡献对于真实图片 X_{real} 和生成器伪造的图片 X_{fake} ，辅助分类器应该能够预测它所属的类别。噪声分布 Z 独立于类别标签 C ；判别器训练中试图最大化 $L_S + L_C$ ，而生成器最大化 $L_C - L_S$ 。

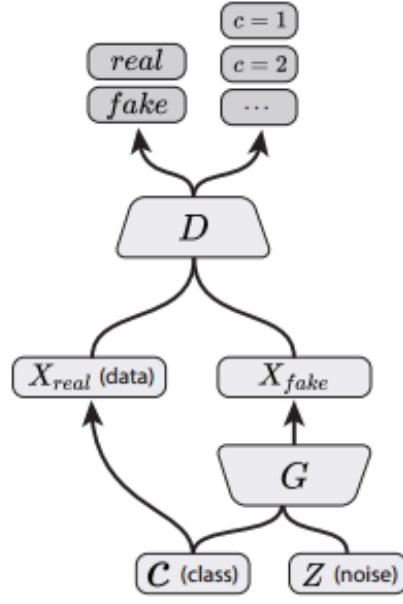


图 2-5 ACGAN 结构示意图。

$$L_S = \mathbb{E}[\log P(S = \text{real}|X_{\text{real}})] + \mathbb{E}[\log P(S = \text{fake}|X_{\text{fake}})] \quad (2-5)$$

$$L_C = \mathbb{E}[\log P(C = c|X_{\text{real}})] + \mathbb{E}[\log P(C = c|X_{\text{fake}})] \quad (2-6)$$

ACGAN 的一个作用就是弥补数据的不足，让计算机拥有了能够模仿现有数据从而生成独特数据的能力，用途相当广泛。

2.3 基于生成对抗网络的图像转译

图像翻译旨在通过设计端到端的模型将源域图像转换到目标域图像，通常源域提供图像的内容，目标域提供图像的“风格”（可以是图像属性或图像风格），在源域内容下实现目标域的“风格”化，从而实现源域图像到目标域图像的转换。说的通俗点图像翻译可以是标签图到场景图的转换、线条轮廓到色彩图像转换、图像的风格转换，春夏场景的变换，人脸的属性变换，也可以是白昼交替的转换。只要符合上述端到端转换的任务，都可以通过图像翻译实现。根据训练中是否需要成对图像即源域图像和目标域图像是否一一对应，可以分为成对图像转译和非成对图像转译。

2.3.1 成对图像转译方法

2017 年 Isola 等人提出 pix2pix^[3]，是典型的配对训练图像的图像转译方法。本质上是基于 cGAN，将输入图像作为条件，学习从输入图像到输出图像之间的映射，从而得到指定的输出图像。pix2pix 等成对图像转译方法可以比较清晰的语义图到真实场景图、灰度图上色、日景到夜景、草图到图像等图像的转换结果。

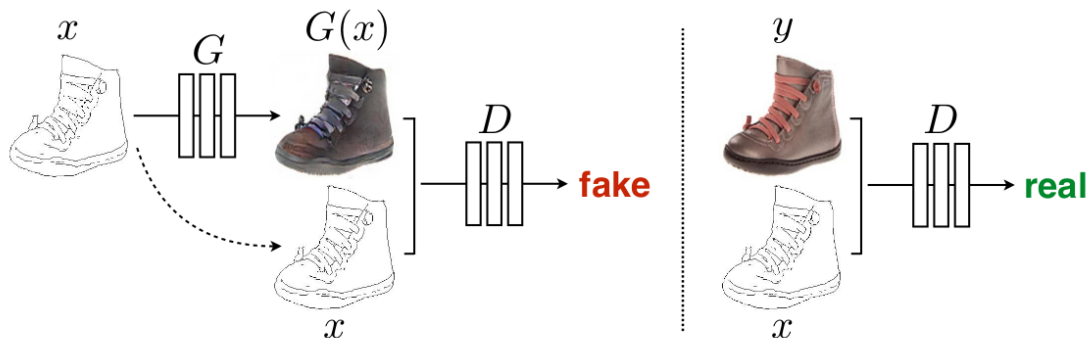


图 2-6 pix2pix 训练示意图。

以草图到图像的转译为例，训练过程如图 2-6所示。网络结果基于 cGAN，将图片 x 作为条件， y 是跟 x 配对的真实结果，需要输入生成器 G 和判别器 D 中。判别器试图区分合成图像 $\{x, G(x)\}$ 和真实图像 $\{x, y\}$ ，生成器试图生成具有真实性的结果以迷惑判别器。跟无条件 GAN 不同，pix2pix 的生成器和判别器都需要输入条件 x 。

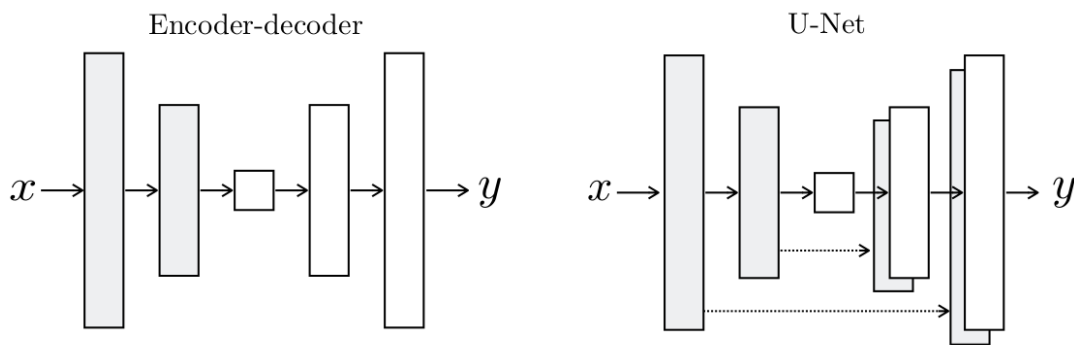


图 2-7 pix2pix 生成器 U-net 网络结构示意图。

pix2pix 的生成器和判别器网络结构在 DCGAN 的深度卷积的基础上做了一些改进。生成器采用 U-Net 全卷积结构，跟 Encoder-Decoder 网络结构先下采样到低维度再上采样到原始分辨率有所区别。如图 2-7所示，左侧是 Encoder-Decoder 网络结构，右侧是 U-Net^[17] 结构，区别是加入 skip-connection，编码过程中对应的特

征图和解码之后同样尺寸的特征图按照通道拼接在一起，目的是用来保留不同分辨率下像素级的细节信息。为了更好的对图像局部做判断，判别器利用马尔可夫性的判别器 PatchGAN 结构。其工作原理是将图像中独立的 $N \times N$ 个 patch 真假判别，而非将整个图像真假判别，这种设计使得参数更少、运行更快，且能够产生更为锐利的结果。

$$L_{cGAN}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))] \quad (2-7)$$

$$L_{L1}(G) = \mathbb{E}_{x,y,z}[\|y - G(x, z)\|_1] \quad (2-8)$$

$$\min_G \max_D L_{cGAN}(G, D) + \lambda L_{L1}(G) \quad (2-9)$$

基于 cGAN 的目标函数为 2-7， L_1 损失为 2-8，总目标函数为 2-9。 L_1 损失可以恢复图像低频部分让生成图像跟真实训练数据尽量相似，对抗损失额可以恢复图像的高频部分细节。

2.3.2 非成对图像转译方法

很多任务中，由于真实数据限制，我们无法得到成对的源域和目标域图像来进行训练，比如将照片转译成艺术作品风格、人脸转译成漫画风格等源域和目标域图像没有关联性甚至毫不相干的风格转换任务。由此可见，非成对图像转译应用场景更加广泛。

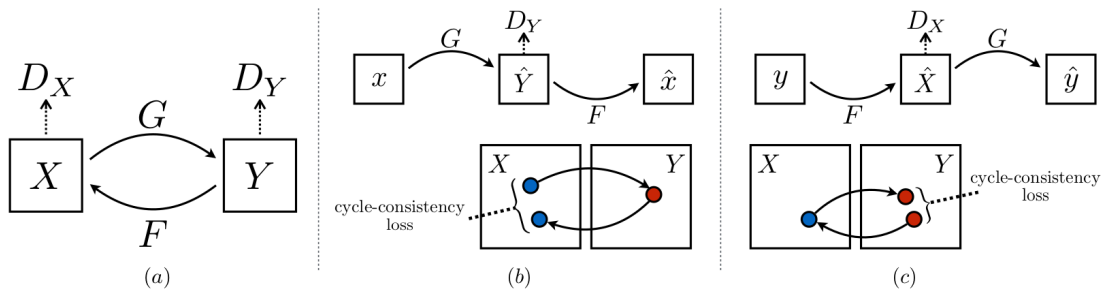


图 2-8 CycleGAN 网络结构示意图^[4]。

CycleGAN^[4] 是一种新颖的无监督循环生成结构，主要思路是训练两对生成器和判别器模型将图像从一个域转换到另一个域，过程中要求循环一致性。实质上，CycleGAN 在每个方向上都是单向 GAN，共享两个生成器，各自带一个判别器，总共两个生成器两个判别器。网络结构如图 2-8 所示，假设有 X 和 Y 两个域，生成器 G 基于 X 域图像生成 Y 域图像；生成器 F 基于 Y 域图像生成 X 域图像，这两个生成器是相反的过程，通过图 2-8(b)(c) 中的循环一致性损失进行约束。CycleGAN 的两个判别器 D_X 和 D_Y 分别用来判断输入 X 和 Y 域的图像真假。因此，CycleGAN 可以看作两个 GAN 的融合，一个 GAN 由生成器 G 和判别器 D_X 构成，实现从 X 到 Y 域图像的生成和判别；另一个 GAN 由生成器 F 和判别器 D_Y 构成，实现从 Y 域到 X 域的图像生成和判别，两个网络构成循环过程。

$$L_{cyc}(F, G, X, Y) = \mathbb{E}_{x \sim p_{data}(x)} [\|G(F(x)) - x\|_1] + \mathbb{E}_{y \sim p_{data}(y)} [\|G(F(y)) - y\|_1] \quad (2-10)$$

CycleGAN 创新提出的循环一致性损失函数如 2-10 所示，映射 $G(F(y)) \approx y$ 和 $G(F(x)) \approx x$ 在训练过程中学习到。也就是 X 域图片转译到 Y 域中，还能再逆转回来。生成器 G 对应方向判别器 D_X 和生成器 F 对应方向判别器 D_Y 分别可以定义一个 GAN 对抗损失，最终目标函数 2-11 由两个方向上的对抗损失加上循环一致性损失三部分组成。

$$L = L_{GAN}(F, D_Y, X, Y) + L_{GAN}(G, D_X, X, Y) + \lambda L_{cyc}(F, G, X, Y) \quad (2-11)$$

2.3.3 图像多模态转译

从 UNIT^[5] 以共享潜在空间为假设看成图像求联合概率分布，实现无监督的两个域之间分解表达的图像转译。包括 UNIT 在内，当时图像转译无论是有监督还是无监督方法，大都是一对一，即输入一张图像只能生成一种风格，缺乏生成结果多样性。比如，同一个场景不同光照条件就是一个模式，不仅仅是白天和黑夜风格，也有傍晚、清晨等风格。在 UNIT 分解表达学习基础上，MUNIT 和 DRIT 进一步提出多模态图像转译方法。

MUNIT 将潜在空间的潜在编码分为内容编码 c 和风格编码 s 。不同域的图像共享内容编码空间 C 而独享风格编码空间 S 。如图 2-9 (a) 部分所示，域 X_1 的风

格编码空间为 s_1 内容编码空间为 c_1 ，域 X_2 的风格编码空间为 s_2 内容编码空间为 c_2 。内容控制图像中低维信息，风格描述图像中高维属性如颜色、纹理、样式等。网络架构清晰，两个编码器 $E_1 E_2$ 分别生成域 X_1 和域 X_2 的内容编码和风格编码；两个生成器 $G_1 G_2$ 分别生成域 X_1 和域 X_2 的图像结果；两个判别器 $D_1 D_2$ 分别判别域 X_1 和域 X_2 的图像真假。

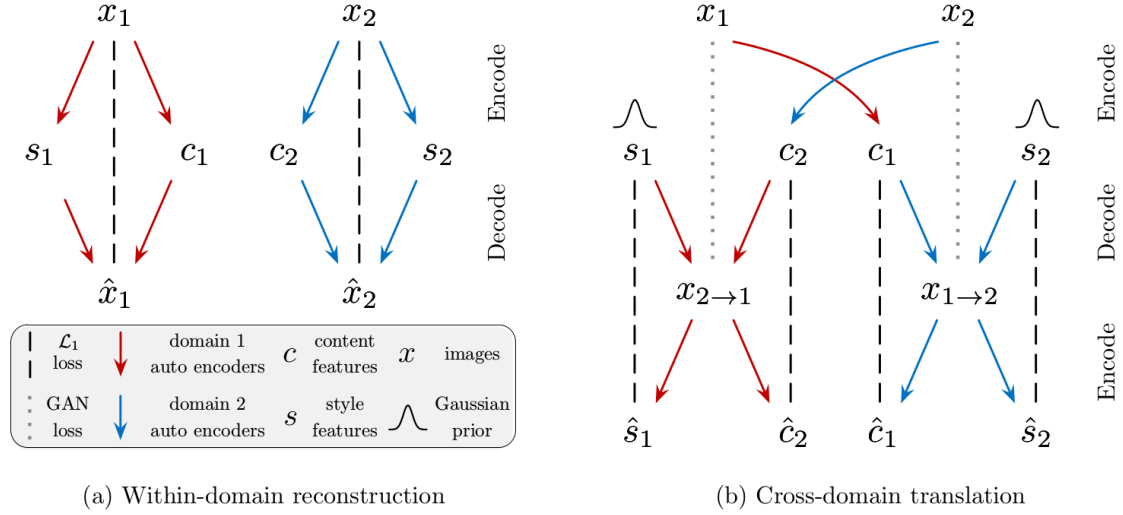


图 2-9 MUNIT 网络结构示意图^[6]。

整个网络的训练包括两个部分，如图 2-9(a) 所示的域内重建和如图 2-9(b) 所示的跨域转译。

域内重建时，输入图像 $x_1 \in X_1$ 经过编码器 E_1 可以得到内容编码 c_1 和风格编码 s_1 ，即 $c_1, s_1 = E_1(x_1)$ ，再输入到域 X_1 的生成器 G_1 ，形成图像 $\hat{x}_1 = G_1(c_1, s_1)$ 。同样的，输入图像 $x_2 \in X_2$ 经过编码器 E_2 可以得到内容编码 c_2 和风格编码 s_2 ，即 $c_2, s_2 = E_2(x_2)$ ，再输入到域 X_2 的生成器 G_2 ，形成图像 $\hat{x}_2 = G_2(c_2, s_2)$ 。

跨域转译时，输入图像 $x_1 \in X_1$ 经过编码器 E_1 可以得到内容编码 c_1 和风格编码 s_1 ，即 $c_1, s_1 = E_1(x_1)$ ，保留内容编码 c_1 与满足正态分布的随机风格编码 s_2 同时输入到生成器 G_2 中，生成具有域 X_2 风格的图像 $x_{1 \rightarrow 2} = G(c_1, s_2)$ ；同样的，输入图像 $x_2 \in X_2$ 经过编码器 E_2 可以得到内容编码 c_2 和风格编码 s_2 ，即 $c_2, s_2 = E_2(x_2)$ ，保留内容编码 c_2 与满足正态分布的随机风格编码 s_1 同时输入到生成器 G_1 中，生成具有域 X_1 风格的图像 $x_{2 \rightarrow 1} = G(c_2, s_1)$ 。

MUNIT 结构包括生成结构和判别结构，生成结构包括编码器和解码器。编码器包括内容编码器和风格编码器，解码器采用 AdaIN (Adaptive Instance Normalization)^[18] 仿射变换将目标域风格转译；判别器使用多尺度判别器^[19] 结构，能帮助实现高分辨率图像的判别。

$$L_{recon}^{x_1} = \mathbb{E}_{x_1 \sim p(x_1)} [\| G_1(E_1^c(x_1), E_1^s(x_1)) - x_1 \|_1] \quad (2-12)$$

由此，可以构建重建损失函数包括图像重建损失函数 (如式 2-12)，内容重建损失 (如式 2-13) 和风格重建损失函数 (如式 2-14)。

$$L_{recon}^{c_1} = \mathbb{E}_{c_1 \sim p(c_1), s_2 \sim q(s_2)} [\| E_2^c(G_2(c_1, s_2)) - c_1 \|_1] \quad (2-13)$$

$$L_{recon}^{s_2} = \mathbb{E}_{c_1 \sim p(c_1), s_2 \sim q(s_2)} [\| E_2^s(G_2(c_1, s_2)) - s_2 \|_1] \quad (2-14)$$

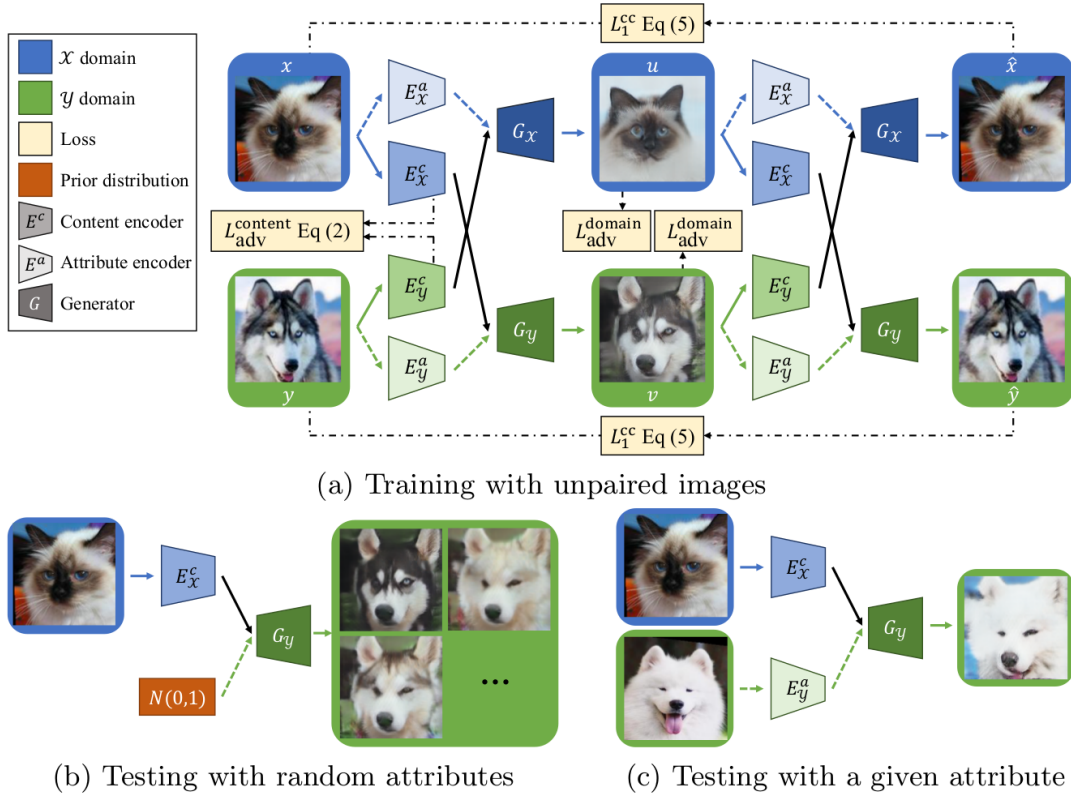
目标函数 (如式 2-16) 是对抗损失 (如式 2-15) 以及重建损失函数之和。

$$\begin{aligned} L_{GAN}^{x_2} = & \mathbb{E}_{x_2 \sim p(x_2)} [\log D_2(x_2)] \\ & + \mathbb{E}_{c_1 \sim p(c_1), s_2 \sim q(s_2)} [\log(1 - D_2(G_2(c_1, s_2)))] \end{aligned} \quad (2-15)$$

$$\begin{aligned} \min_{E_1, E_2, G_1, G_2} \max_{D_1, D_2} L(E_1, E_2, G_1, G_2, D_1, D_2) = & L_{GAN}^{x_1} + L_{GAN}^{x_2} \\ & + \lambda_x (L_{recon}^{x_1} + L_{recon}^{x_2}) \\ & + \lambda_c (L_{recon}^{c_1} + L_{recon}^{c_2}) \\ & + \lambda_s (L_{recon}^{s_1} + L_{recon}^{s_2}) \end{aligned} \quad (2-16)$$

DRIT^[7] 跟 MUNIT 在思路完全一致，都是共享内容空间，独享属性/风格空间。网络训练如图 2-10所示同样由域内重建和跨域转译两个部分组成，跨域转译用来生成交换风格厚的图像，域内重建被编码后的潜在编码。用 X, Y 表示两个域，网络由两个内容编码器 E_X^c 和 E_Y^c 和两个属性编码器 E_X^a 和 E_Y^a ，两个生成器 G_X 和 G_Y 以及两个判别器 D_X, D_Y 组成。

跨域转译和域内重建跟 MUNIT 一致，在此不再赘述。在内容编码和属性编码结合时，DRIT 和 MUNIT 有一定的差异，MUNIT 将风格编码使用 AdaIN 内嵌到生成器（解码器）的中间层，而 DRIT 采用共享两个编码器 E_X^c, E_Y^c 最后一层和两


 图 2-10 DRIT 训练示意图^[7]。

个生成器 G_X , G_Y 第一层的权值。共享权值的情况下, 为了区分两个域的编码文中提出使用判别损失 2-17。

$$\begin{aligned}
 L_{adv}^{content} = & \mathbb{E}_x \left[\frac{1}{2} \log D^c(E_X^c(x)) + \frac{1}{2} \log(1 - D^c(E_X^c(x))) \right] \\
 & + \mathbb{E}_y \left[\frac{1}{2} \log D^c(E_Y^c(y)) + \frac{1}{2} \log(1 - D^c(E_Y^c(y))) \right]
 \end{aligned} \quad (2-17)$$

损失函数中也加入 CycleGAN 中提出的循环一致性损失 2-18, 其中 $u = G_X(E_Y^c(y), E_X^a(x))$, $v = G_Y(E_X^c(x), E_Y^a(y))$, 还有重建损失, 对抗损失等。总的目标函数如 2-16所示。

$$\begin{aligned}
 L_1^{cc}(G_X, G_Y, E_X^c, E_Y^c, E_X^a, E_Y^a) = & \mathbb{E}_{x,y} [\| G_X(E_Y^c(y), E_X^a(x)) - x \|_1 \\
 & + \| G_Y(E_X^c(x), E_Y^a(y)) - y \|_1]
 \end{aligned} \quad (2-18)$$

$$\begin{aligned} \min_{G,E,E^a} \max_{D,D^c} & \lambda_{adv}^{content} L_{adv}^c + \lambda_1^{cc} L_1^{cc} + \lambda_{adv}^{domain} L_{adv}^{domain} \\ & + \lambda_1^{recon} L_1^{recon} + \lambda_1^{latent} L_1^{latent} + \lambda_{KL} L_{KL} \end{aligned} \quad (2-19)$$

除了 MUNIT 和 DRIT 这些经典的多模态转译工作外，近期发表的工作 StarGAN v2^[9] 是在 StarGAN^[8] 多域一对一转译任务的基础上，进一步实现了多个域彼之间的多模态转译。DRIT++^[10] 则是在 DRIT 两个域多模态转译任务基础上，进一步实现多域之间转译。越来越多的研究工作，将多模态转译任务拓展至更广泛的应用场景。

2.4 本章小结

本章深入分析介绍了基于生成对抗网络的多模态转译方法。

第一节介绍了生成模型以及各自特点，引入生成对抗网络结构和原理，分析了生成对抗网络的优缺点。

第二节介绍了生成对抗网络的研究历史，详细阐述了在生成对抗网络的发展过程中，针对生成对抗网络中存在的问题进行解决和具有里程碑意义的经典模型。

第三节在前两节铺垫的基础上，引入基于生成对抗网络的多模态转译概念，展示了经典成对图像转译和非成对图像转译经典方法，仔细研究了常用的多模态转译模型以及后续发展的前沿工作。为后续我们基于生成对抗网络的水下图像多模态转译问题奠定理论基础。

3 水下图像合成研究

海洋包含着未知的生物和巨大能源，对地球上生命延续起着重要的作用。自二十世纪中叶以来，高科技海洋勘测一直备受瞩目，视觉技术因能传达高密度信息而被应用在海洋环境中。研究人员致力于为各种水下应用捕获高质量的水下图像，包括水下机器人、水下救援、生态检测、海洋生物跟踪和实时导航等。

3.1 水下图像成像原理及存在问题

水下环境的特殊物理和化学特性严重影响了水下图像的质量和数量，产生许多比在地面成像的更难克服的问题。如图 3-1 所示，水下成像存在的问题可以进行详尽分析。

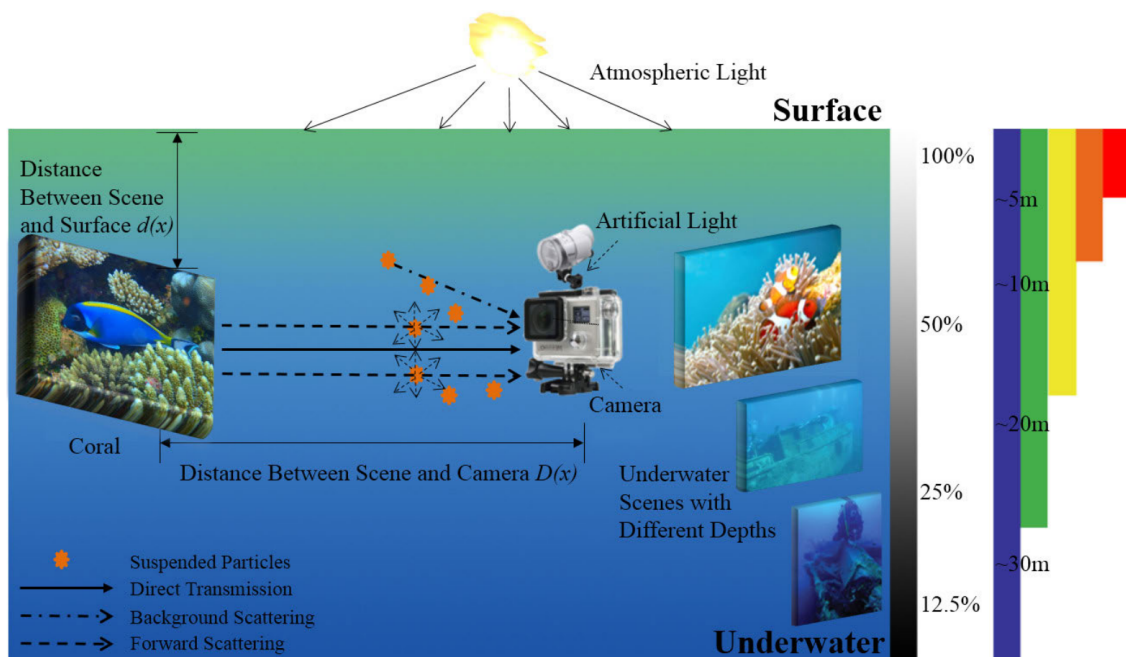


图 3-1 水下光学成像原理图。

水下图像存在色偏的问题，图像视觉上会呈现蓝色、绿色和蓝绿色。这是由于不同颜色的光在水中的衰减特性造成的，红光波长较长，蓝色和绿色光波长较短，在水介质中光的传播过程中波长较长的光吸收速度更快，因此水下图像呈现蓝绿色调。

水下采集到的图像会存在对比度低，模糊的问题。吸收和散射，即悬浮在水中的颗粒吸收了传播中的大部分光，并将水下场景的反射光在到达相机前改变了

光的方向，水的浑浊严重降低了图像质量。人为光线除了吸收和散射，还会出现照明不均匀以及照明不足的问题，会造成阴影。

此外，由于水下环境的成像需要耗费巨大的人力和物力资源，水下图像数据量远远小于非水下自然场景的图像数据量，高质量的图像结果更难能可贵。缺少水下环境中的高质量图像给水下视觉研究造成巨大的困难，这也是研究人员针对水下图像增强和复原进行大量工作的原因。

3.2 基于生成对抗网络的水下图像研究

3.2.1 水下图像数据集

真实水下场景的常用水下数据集数量有限。

UIEB^[20]是具有多样性场景、数量大以及具有高质量参考图像的成对真实场景数据集。原始图像共有 950 张，具有参考图像配对结果的 890 张，其余 60 张没有对应的参考图像。采用 12 种图像增强方法来进行图像增强，针对每张原始图像 50 个志愿者来选出令人满意的图像增强结果作为真值，令半数以上人不满意的 60 结果作为挑战图像没有真值。

RUIE^[21]是共四千多的大规模真实水下数据集。针对可见度增强的性能进行评估的水下图像质量子集，包含五种质量的水下图像其中每种 726 张，用来测试水下图像增强效果；评估在不同照明和色偏算法的子集，包含蓝色、绿色以及蓝绿色三种色偏的水下图像每种 100 张；对高级计算机视觉任务如分类和检测的子集，包含对扇贝、海参和海胆三种海洋生物的标注每种 100 张。

水下图像数据集相比较于非水下场景（空中）图像数量非常少，进行水下视觉图像研究时，研究人员会按照任务需要将空中图像进行水下环境条件模拟。

UWCNN 论文中提出的从地面图像得到相应水下图像数据集。通过 RGB-D NYU-v2 indoor 数据集生成 10 种类型的水下图像。RGB-D NYU-v2 indoor 共 1449 张图像，其中的 1000 张用于训练，剩下的 449 张用于测试，数据集提供的 10 类水图像也是共 1449 张图像。对于每张室内图像，根据生成的随机均匀全球大气光和深度（从 0.5 米到 15 米，每张图像生成 5 张水下图像。RGB-D NYU-v2 indoor 数据集中的图像作为 ground truth（真图），和合成的水下图像（假图）配对。

EUVP 是 UGAN^[22]中提出数据集的基础上更加齐全的数据集。UGAN 中，ground truth 是真实的相对不失真的水下图像（真图），配对的是生成的失真的水下图像（假图）。从 ImageNet 中选择一些不失真的水下图像，组成 X 域，共 6143 张图像，再从 ImageNet 中选择一些有明显水下图像特点的水下图像，组成 Y 域，共 1817 张图像。用 CycleGAN 实现风格迁移，即将 X 域中不失真的图像转成 Y 域中

失真水下图像的风格（主要是颜色的变化）。最终用于训练的图像对是将 X 域中的图像都转译成失真水下图像。EUVP 中，分为 paired 和 unpaired 两种。Paired 文件夹中包含 3 个子文件夹：Underwater Dark、Underwater ImageNet 和 Underwater Scenes，其中 Underwater Dark 有 5550 个图像对，570 张验证图像，共 11670 张；Underwater ImageNet 有 3700 个图像对，1270 个验证图像，共 8670 张图像；Underwater Scenes 有 2185 个图像对，130 个验证图像，共 4500 张图像。Unpaired：poor quality 中有 3195 张，good quality 中有 3140 张，validation 中有 330 张，共 6665 张图像。

SUN 数据集共包含 908 个类别和 131067 张图像，用其中采样良好的 397 个类别去评估场景识别算法。水下图像只是其中一类，一些论文中会用到 SUN 数据集去做水下增强和复原。仅关注水下图像，有 5 类：coral reef、ocean deep、ocean shallow、pool、wreck。

3.2.2 水下图像评价指标

UCIQE^[23] 和 UIQM^[24] 是无参考/无真值评价水下图像质量的常用指标，在图像增强和图像复原任务中被广泛应用。

$$UCIQE = c_1 \times \sigma_c + c_2 \times con_l + c_3 \times \mu_s \quad (3-1)$$

UCIQM 是色彩浓度，饱和度和对比度为测量分量的线性组合，可以定量的评价水下增强结果的色偏、模糊和对比度情况。属于无参考/无真值的图像质量评价指标，首先将水下图像从 RGB 颜色空间转换到 CIELab 颜色空间，这样更符合人类视觉感知，然后计算个测量分量，具体计算如式 3-1 所示。其中， σ_c 为色度的标准方差， con_l 为亮度的对比度， μ_s 是饱和度的平均值， c_1 ， c_2 和 c_3 分别为线性组合的权重值。

$$UIQM = c_1 \times UICM + c_2 \times UISM + c_3 \times UIConM \quad (3-2)$$

UIQM 针对水下图像的退化机理和成像特点，采用色彩测量 UICM，清晰度测量 UISM 和对比度测量 UIConM 作为评价水下图像质量的依据。UIQM 属于无参考/无真值的图像评价指标，通过测量分量的线性组合来表征图像的视觉质量，如式 3-2 所示。其中， c_1 ， c_2 和 c_3 分别为线性组合的权重值，权重值的设定需要视具体任务而定，评价水下图像的颜色偏差修正结果时，需设定色度测量分量 UICM

更大的权重因子；当评价对比度和清晰度时，需要设定清晰度测量分量 UISM 和对比度测量分量 UIConM 更大的权重因子。

尽管，两种无参考的指标一定程度上给出定量的图像评价分数，但缺点也很明显。首先，水下图像质量不佳的视觉类型较多，仅参考几个测量分量的线性组合对于图像评价是片面的，缺少客观性和准确性；其次，测量分量权重的设置具有主观性，难以公正的评价水下图像质量；最后，评价结果域视觉效果不一致，会出现指标指示评价较好但视觉上没有直观的展现图像的优越性。

3.2.3 经典水下图像合成方法

目前现有基于生成对抗网络的水下图像处理模型，大多是利用海洋光学成像原理/无监督图像转译方法将空中图像转译成相对应的成对水下图像，目的都是为图像增强和图像恢复提供配对图像数据，图像处理模型训练有监督的网络结构实现图像增强/图像恢复。代表性的工作如下，

UWCNN^[20] 是一种基于卷积神经网络的经典图像增强模型。在 NYU v2^[25] 深度数据集上，将海洋光学成像原理利用深度图应用于合成十种不同类型的水下图像，针对每种水下矮图像类型分别训练多个 UWCNN 模型，能够模拟各种退化的水下图像以进行水下数据增强。训练集使用 NYU v2 数据集合成配对的十种类型水样，测试集选择网络上的水下真实图像测试训练好的每种水样类型的 UWCNN 效果。

UGAN^[22] 使用 CycleGAN 来生成水下图像，从而为进行颜色矫正提供数据。在整个过程中，需要在水下和陆地两个单独的域中拍摄物体图像。使用生成对抗网络作为生成模型，构造将水下图像的真实外观估计变成成对图像转译问题，使用来自两个域的图像作为输入和真值。使用 ImageNet 的子集来训练和评估网络性能。UGAN-P 是在 UGAN 基础上加入梯度差异损失，直接惩罚生成器中图像梯度预测的差异来增强预测。

WaterGAN^[26] 提出无监督的水下图像色彩矫正模型。RGB-D 的地面图像数据集和水下图像样本集作为输入使用 WaterGAN 合成与 RGB-D 对齐的相应水下结果，用生成的水下结果和地面结果作为配对数据训练图像复原模型，测试使用真实的水下图像，输出矫正后的图像和相对深度图。

UWGAN^[27] 将彩色 RGB 图像及其深度图作为输入，然后通过生成对抗性训练学习参数，从而基于水下光学成像模型合成不同类型的水下真实图像。提出用于水下图像恢复和增强的 U-net 体系结构，比较了 U-net 中不同损失函数的影响，在此基础上提出了最适合水下图像恢复的损失函数。

Underwater-GAN^[28] 基于生成对抗网络，将 Wasserstein GAN 与梯度惩罚项作为基本网络框架的图像复原模型，对抗损失和感知损失作为生成器损失。恢复水下图像由于复杂的水下成像环境和恶劣的光照条件而产生的退化。数据集不存在成对的清晰图像以及相应的退化水下图像，从网上收集 3000 张清晰图像作为真值，使用清晰图像来模拟 17 中不同衰减进行训练和测试。

Uplavikar 等提出水下增强模型^[29]，问题是水类型的多样性使得难以用一个模型来完美实现全部水类型的增强，该工作提出一个 Encoder-Decoder 卷积网络结构，将不同水类型视为不同域消除相应干扰来学习图像的内容特征。将 NYU v2 深度数据集成 10 种 Jerlov 风格水样^[20] 作为训练数据，UIEB 作为测试数据。Encoder 将图像编码到潜在空间时，还进行了水样类型判别，帮助 Decoder 重建出不带水样类型的增强图像结果。

在进行图像增强和图像复原工作中，往往需要成对的数据，在有监督的网络结构中进行训练，因此水下图像合成方法主要是为了后续研究提供成对的合成数据结果。在上述提到的水下图像合成方法中，主要分成两种类型的合成。一种是利于水下光学成像原理，利用深度图或者传输图作为水下条件将空中图像模拟合成相对应的水下场景图像，如 WaterGAN，UWCNN，UWGAN，Underwater-GAN 和 Uplavikar 提出的方法。依赖这种方法始终无法摆脱物理模型的控制，需要深度图/传输图作为指引；且生成水下图像类型固定，若要生成多样化结果需要训练多个模型依次对应。另一种是利用无监督一对一映射的图像转译方法 CycleGAN，使用不成对的数据即可进行水下图像合成，如 UGAN 和 MLFcGAN。

这些基于深度学习的水下图像增强和水下图像复原模型，训练需要成对的地面和水下数据进行训练。在缺少成对地面和水下真实数据集情况下，需要使用 RGB 地面真实图像和深度图/传输图利用海洋光学原理模拟生成不同类型的水下图像结果。这样的方法始终无法摆脱物理模型的控制，需要深度图/传输图作为指引；且生成水下图像类型固定，若要生成多样化结果需要训练多个模型依次对应。

3.3 基于分解的水下图像研究

3.4 本章小结

本章主要对水下图像进行研究，为后续研究水下图像转译提供理论基础和数据支撑。

第一小节，介绍了水下图像的成像原理和存在问题。这是现有水下图像增强和水下图像复原任务的出发点和解决目标。

第二小节，对水下任务中常用数据集进行简要介绍，对水下图像评价指标进

行研究和分析其原理和合理性，最后是介绍水下图像处理常用的经典方法。

4 水下图像多模态转译模型设计

本章主要介绍水下图像多模态转译问题以及相应的改进算法。

4.1 水下图像多模态转译问题定义

4.2 水下图像多模态转译问题分析

基于上述几章的分析，我们了解到水下图像合成的方法，一种是基于水下光学物理模型将给定图像合成水下场景图像，另一种是利用 CycleGAN 这种无监督图像转译模型，实现图像从当前场景到水下场景的转译。

这些问题给水下图像转译问题带来巨大的困难。将有限的给定输入图像，转译成多种水下环境条件的图像结果就显得尤为重要。我们将这种给定图像转译到多种水下环境的图像转译称作水下图像多模态转译。基于生成对抗网络，我们的网络结构可以在无监督条件下，实现水下图像的多模态转译

定义、选择的 baselines 与问题之间的关系

4.3 建立模型算法和损失函数

4.3.1 模型创新点

4.3.2 目标函数和算法

4.3.3 网络结构

表 4-1 RUIE 数据集上多模态转译定量结果对比

Metrics	CycleGAN	MUNIT	DRIT	Ours
FID↑	243.4	139.4	179.2	0
LPIPS↓	0.634	0.452	0.575	0

表 4-2 UWCNN 数据集上多模态转译定量结果对比

Metrics	CycleGAN	MUNIT	DRIT	Ours
FID↑	80.7	232.1	290.7	0
LPIPS↓	0.490	0.647	0.668	0

4.4 本章小结

5 水下图像多模态转译的实验与分析评测

本章将针对水下图像多模态转译问题设计实验、确定评价指标，验证我们设计的网络模块的有效性，并域其他方法进行对比和分析。

首先，

5.1 experiment

6 总结与展望

6.1 全文总结

6.2 后续工作展望

7 第 6 章

7.1 第 6.1 节

7.1.1 第 6.1.1 节

参考文献

- [1] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. *Advances in neural information processing systems*, 2014. 2672–2680.
- [2] Kingma D P, Welling M. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013..
- [3] Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017. 1125–1134.
- [4] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. *Proceedings of the IEEE international conference on computer vision*, 2017. 2223–2232.
- [5] Liu M Y, Breuel T, Kautz J. Unsupervised image-to-image translation networks. *Advances in neural information processing systems*, 2017, 30:700–708.
- [6] Huang X, Liu M Y, Belongie S, et al. Multimodal unsupervised image-to-image translation. *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018. 172–189.
- [7] Lee H Y, Tseng H Y, Huang J B, et al. Diverse image-to-image translation via disentangled representations. *Proceedings of the European conference on computer vision (ECCV)*, 2018. 35–51.
- [8] Choi Y, Choi M, Kim M, et al. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018. 8789–8797.
- [9] Choi Y, Uh Y, Yoo J, et al. Stargan v2: Diverse image synthesis for multiple domains. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. 8188–8197.
- [10] Lee H Y, Tseng H Y, Mao Q, et al. Drit++: Diverse image-to-image translation via disentangled representations. *International Journal of Computer Vision*, 2020. 1–16.
- [11] Mirza M, Osindero S. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014..
- [12] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 2017, 60(6):84–90.
- [13] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database. *2009 IEEE conference on computer vision and pattern recognition. Ieee*, 2009. 248–255.
- [14] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015..
- [15] Mao X, Li Q, Xie H, et al. Least squares generative adversarial networks. *Proceedings of the IEEE international conference on computer vision*, 2017. 2794–2802.

- [16] Odena A, Olah C, Shlens J. Conditional image synthesis with auxiliary classifier gans. International conference on machine learning. PMLR, 2017. 2642–2651.
- [17] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention. Springer, 2015. 234–241.
- [18] Huang X, Belongie S. Arbitrary style transfer in real-time with adaptive instance normalization. Proceedings of the IEEE International Conference on Computer Vision, 2017. 1501–1510.
- [19] Wang T C, Liu M Y, Zhu J Y, et al. High-resolution image synthesis and semantic manipulation with conditional gans. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018. 8798–8807.
- [20] Anwar S, Li C, Porikli F. Deep underwater image enhancement. arXiv preprint arXiv:1807.03528, 2018..
- [21] Liu R, Fan X, Zhu M, et al. Real-world underwater enhancement: challenges, benchmarks, and solutions. arXiv preprint arXiv:1901.05320, 2019..
- [22] Fabbri C, Islam M J, Sattar J. Enhancing underwater imagery using generative adversarial networks. 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018. 7159–7165.
- [23] Yang M, Sowmya A. An underwater color image quality evaluation metric. IEEE Transactions on Image Processing, 2015, 24(12):6062–6071.
- [24] Panetta K, Gao C, Agaian S. Human-visual-system-inspired underwater image quality measures. IEEE Journal of Oceanic Engineering, 2015, 41(3):541–551.
- [25] Nathan Silberman P K, Fergus R. Indoor segmentation and support inference from rgb-d images. ECCV, 2012.
- [26] Li J, Skinner K A, Eustice R M, et al. Watergan: Unsupervised generative network to enable real-time color correction of monocular underwater images. IEEE Robotics and Automation letters, 2017, 3(1):387–394.
- [27] Wang N, Zhou Y, Han F, et al. Uwgan: Underwater gan for real-world underwater color restoration and dehazing. arXiv preprint arXiv:1912.10269, 2019..
- [28] Yu X, Qu Y, Hong M. Underwater-gan: Underwater image restoration via conditional generative adversarial network. International Conference on Pattern Recognition. Springer, 2018. 66–75.
- [29] Uplavikar P M, Wu Z, Wang Z. All-in-one underwater image enhancement using domain-adversarial learning. CVPR Workshops, 2019. 1–8.

致 谢

衷心感谢导师对本人的悉心指导，以及在学习和生活上的支持、鼓励和帮助！

个人简历、在学期间发表的学术论文与研究成果

个人简历

xxxx 年 xx 月 xx 日出生于 xx 省 xx 市。

xxxx 年 xx 月考入 xx 大学 xx 学院 xx 专业，xxxx 年 xx 月本科毕业并获得 xx 学士学位。

xxxx 年 xx 月考入 xx 大学 xx 学院 xx 专业，xxxx 年 xx 月硕士毕业并获得 xx 硕士学位。

xxxx 年 xx 月考入 xx 大学 xx 学院 xx 专业攻读博士学位至今。

发表的学术论文

- [1] XXX X, XXX X, XXX X, et al., Title XXX, IEEE XXX, vol. X, pp. XXX-XXX, 20XX. (SCI/EI XXX)
- [2] XXX X, XXX X, XXX X, et al., Title XXX, IEEE XXX, vol. X, pp. XXX-XXX, 20XX. (SCI/EI XXX)
- [3] XXX X, XXX X, XXX X, et al., Title XXX, IEEE XXX, vol. X, pp. XXX-XXX, 20XX. (SCI/EI XXX)

研究成果

- [1] XXX, XXX, XXX 等. XXX 挑战赛 X 等奖, 20XX 年. 主办方: XXX, XXX, XXX.