基于无监督生成对抗网络的

水下图像合成算法

**Underwater image synthesis algorithm**

**based on unsupervised generative confrontation network**

作 者 姓 名：

学 科 专 业： 计算机科学与技术

学 号：

大连海事大学

Dalian Maritime University

1 引言

1.1 背景介绍

许多领域依靠配备有成像传感器的水下机器人平台来提供海底的高分辨率视图。例如，许多海洋学家使用马赛克拼图来研究水下物体和城市，海洋科学家使用珊瑚礁系统调查来跟踪随时间推移的白化事件。虽然近几十年来水下平台的视觉能力有了很大的进步，但海底环境对感知提出了陆地上不存在的独特挑战。范围相关的照明效果（例如衰减）会导致成像场景和相机之间的光线指数衰减。这种衰减在波长上以不同的速率起作用，并且对于红色通道最强。因此，与在空气中成像时场景的真实颜色相比，原始水下图像显得相对蓝色或绿色。同时，光线通过散射效应被添加回传感器，在整个场景中造成雾霾效应，从而降低了有效分辨率。近几十年来，立体相机一直处于解决这些挑战的最前沿。此外，神经网络可以实现复杂非线性系统的端到端建模。 然而，深度学习在海底应用并没有像陆地应用那样普遍。一个挑战是许多深度学习结构需要大量训练数据，通常与标签或相应的地面实况传感器测量值配对。在深海环境中收集具有深度信息的大量水下数据具有挑战性；获取自然海底场景真实色彩的地面实况也是一个悬而未决的问题。使用校准的立体对，高分辨率图像可以与深度信息对齐，以计算大规模的马赛克图或度量精确的 3D 重建。然而，由于距离相关的水下照明效果导致的图像退化仍然阻碍了这些方法，并且水下图像的恢复涉及复杂物理过程的逆向效应，并且具有特定调查地点的水柱特征的先验知识。

神经网络可以实现复杂非线性系统的端到端建模。然而，深度学习在海底应用并没有像陆地应用那样普遍。许多基于深度学习的图像处理方法需要大量的训练数据，但在深海环境中，收集具有深度信息的大量水下数据具有挑战性；获取自然海底场景真实色彩也是一个悬而未决的问题。

现有公开水下图像数据集由于不存在 ground truth，因此无法用于神经网络模型的训练。

本次实验针对水下图像训练数据集不足的问题，基于生成对抗网络，合成水下图像数据集。

1.2 相关工作

原始GAN由两个有机中整体构成——生成器 G 和判别器 D，生成器的目的就是将随机输入的高斯噪声映射成图像（"假图"），判别器则是判断输入图像是否来自生成器的概率，即判断输入图像是否为假图的概率。

GAN的训练也与CNN大不相同，CNN是定义好特定的损失函数，然后利用梯度下降及其改进算法进行优化参数，尽可能用局部最优解去逼近全局最优解。但是 GAN 的训练是个动态的过程，是生成器 G 与判别器 D 两者之间的相互博弈过程。通俗点讲，GAN 的目的就是无中生有，以假乱真。即要使得生成器 G 生成的所谓的“假图”骗过判别器 D，那么最优状态就是生成器 G 生成的所谓的“假图”在判别器 D 的判别结果为 0.5，不知道到底是真图还是假图。

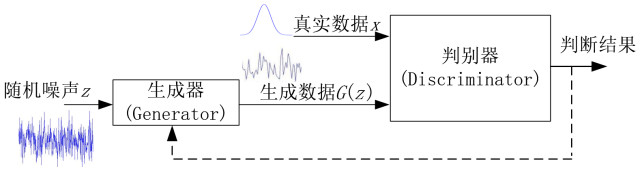


图1：原始GAN网络

我们的实验使用WaterGAN构建为一个生成对抗网络，它同时训练两个网络：生成器G和鉴别器D（图2）。 在标准 GAN中，生成器输入是噪声向量 z，它通过一系列卷积和反卷积层进行投影、重塑和传播。 输出是合成图像 G(z)。鉴别器接收合成图像和单独的真实图像数据集 x 作为输入，并将每个样本分类为真实 (1) 或合成 (0)。 生成器的目标是输出鉴别器分类为真实的合成图像。

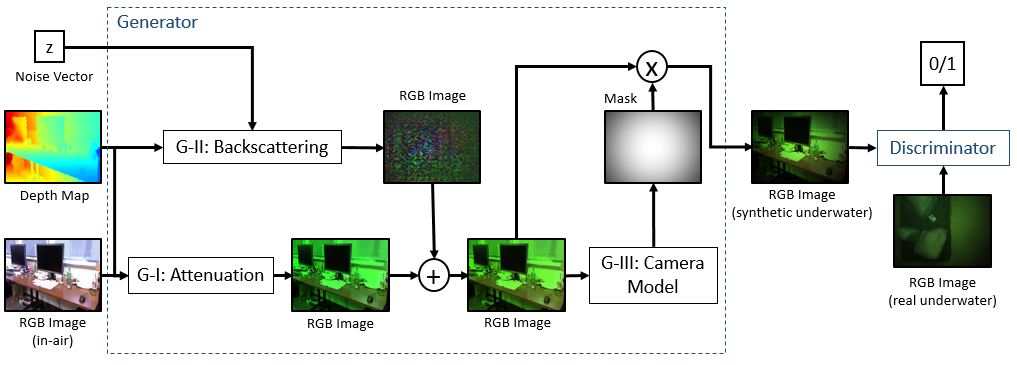


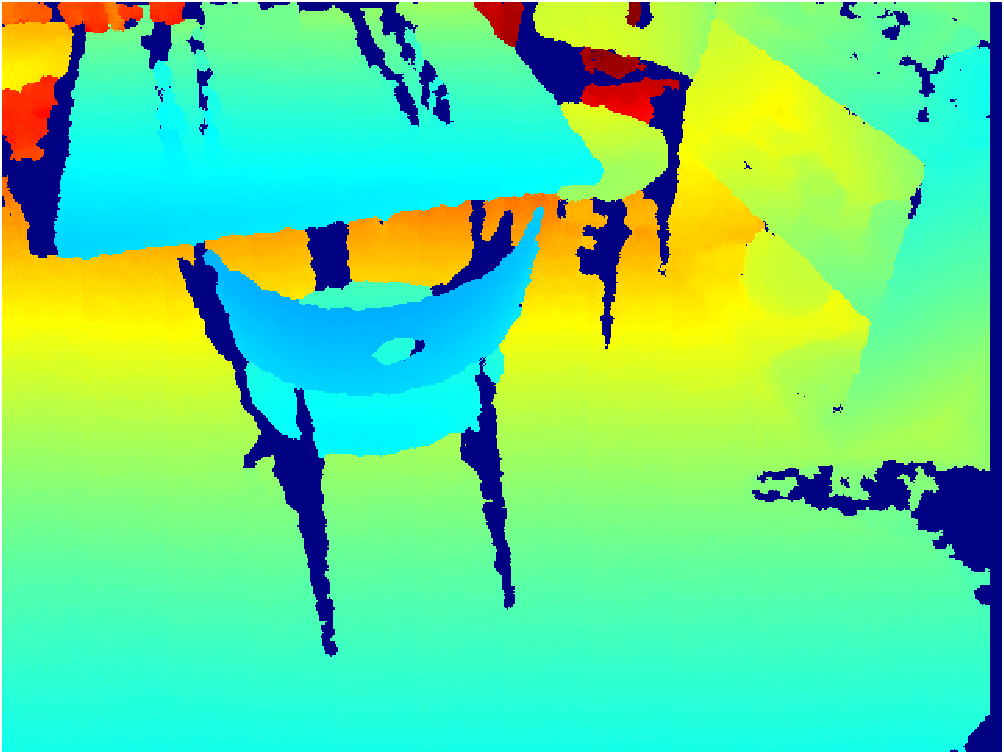
图2：WaterGAN网络

2 算法介绍

2.1 数据集简介

本次实验算法使用RGB-D数据集作为网络训练的输入，RGB-D图像又被叫做深度图像，它有普通三通道RGB彩色图像和与之对应的深度图组成，RGB图和深度图是配准的，像素之间一一对应。深度图包含与视点场景对象表面距离有关信息的图像通道，通道本身类似于灰度图像，每个像素值是传感器测出距离物体的实际距离。

本实验使用的RGB-D数据集为Microsoft 7-Scenes，所有的场景都是由手持 Kinect RGB-D 摄像机以 640×480 的分辨率记录下来的。使用一个 Kinect Fusion 系统的实现来获得地面真实摄像机的轨迹，以及一个密集的3D模型。不同的用户记录每个场景的多个序列，并将其分成不同的训练和测试序列集。数据集提供的每幅彩色图像为24位图，深度图为16位图。示例图片如图3所示，由于16位深度图包含信息丰富，为了更好的显示其表达的信息，在图3第三列对深度图做伪彩色处理。



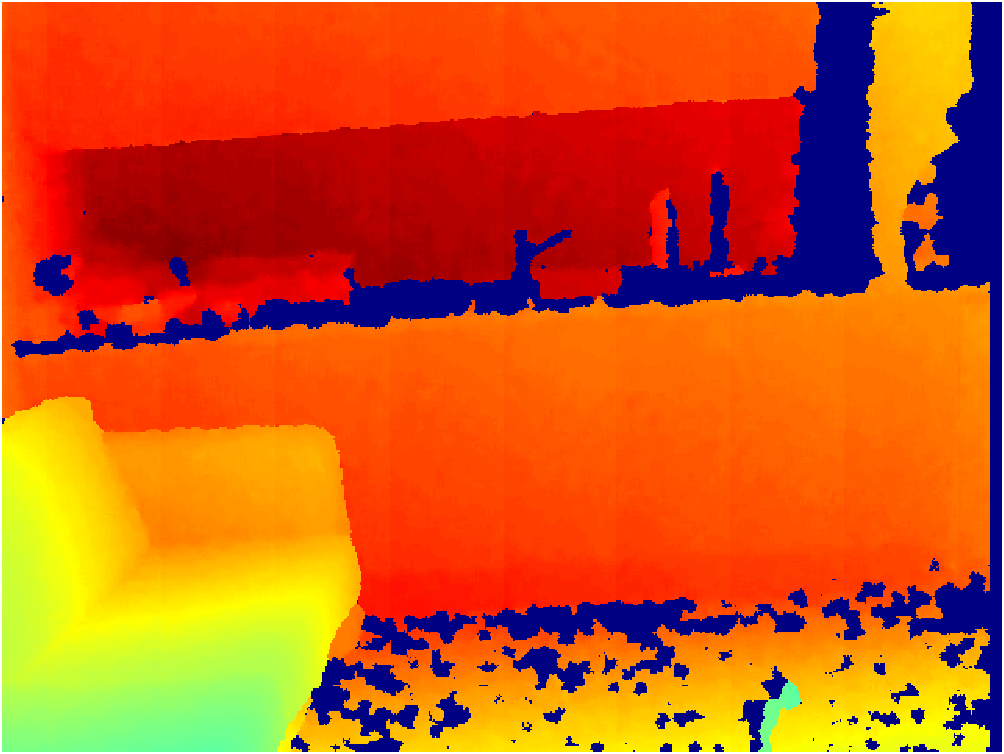


图3：RGB-D数据集示例图片。

第一列为24位RGB彩色图像，第二列为16位深度图，第三列为伪彩色深度图。

2.2 网络模型

网络模型主要分为3个部分，分别为G-I衰减模型、G-II散射模型和G-III相机模型。

G-I衰减模型：简单的修改 Jaffe-McGlamery 的成像模型， 光的衰减依赖于波长与场景深度。为了确保该阶段仅衰减光而不是增加光，并且确保该系数保持在物理范围内，我们将η约束为大于0。对于训练模型参数，所有输入深度图和图像的尺寸均为48×64，保留了全尺寸图像的长宽比。数学描述为：

 （1）

G-II 散射模型: 受水中悬浮粒子的影响，光的散射会使水下图像呈现出特殊的雾度效果。

 （2）

β是取决于波长的标量参数。为了捕获范围相关性，输入48×64深度图和100长度的噪声矢量。噪声矢量被投影，整形并连接到深度图，作为单个通道48×64掩膜。为了捕获与波长有关的效应，将输入复制为内核尺寸为5×5的三个独立的卷积层。将此输出批量标准化，并通过泄漏率为0.2的最终泄漏整流线性单元（LReLU）。不同卷积层的三个输出中的每一个都连接在一起，以创建一个48×64×3维的 Mask 。 由于后向散射将光添加回图像，并确保成像场景的基础结构不会因RGB-D输入而失真，因此将M2添加到G-I的输出中：

 （3）

G-III 相机模型 : 对光晕进行建模，由于镜头的影响，光晕会在图像的边界周围产生阴影图案， 光晕模型：

 （4）

r：图像中心到每个像素的归一化半径，即 r = 0 在图像中心，而 r = 1 在边界处，常数a，b和c 是网络估算的模型参数，输出mask具有输入图像的尺寸，渐晕图像G3：

 （5）

判别器部分使用卷积神经网络，其网络结构如图4所示。

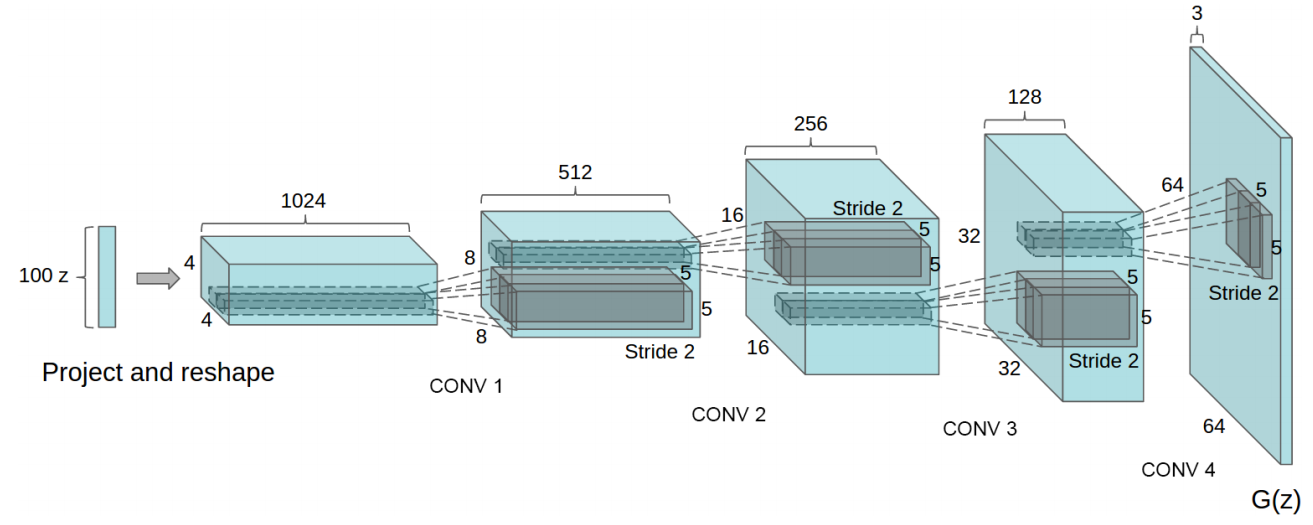
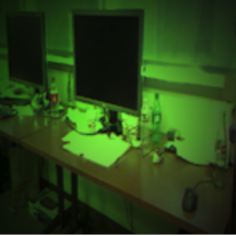
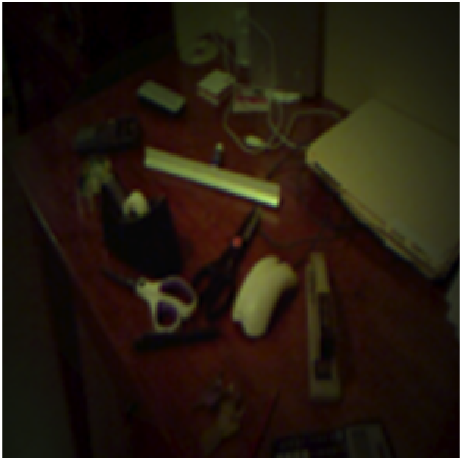


图4：判别器网络结构。

3 实验结果

在本节中，我们展示了算法的生成结果。将训练好的网络模型应用在两个不同的RGB-D数据集上进行测试，图5展示了在Microsoft 7-Scenes数据集下的生成结果图，具有良好的主观视觉效果，符合水下图像基本特征。







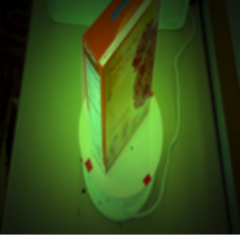
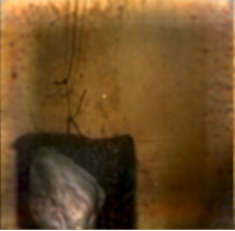
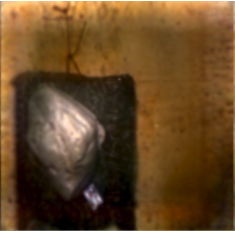
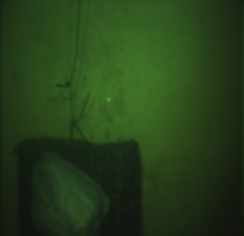


图5：在Microsoft 7-Scenes数据集下的测试结果图。

第一行和第三行为原始图像，第二行和第四行为与之对应的结果图。

为了进一步测试网络的性能，我们在另一个RGB-D数据集上进行测试，测试结果在图6中展示，所使用的数据集来自于密歇根大学海洋流体动力学实验室（University of Michigan Marine Fluid Dynamics Laboratory，MHL）。依然具有良好的主观效果。







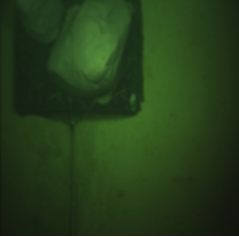
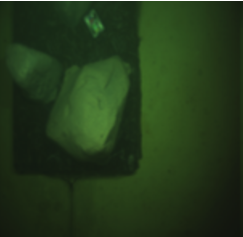
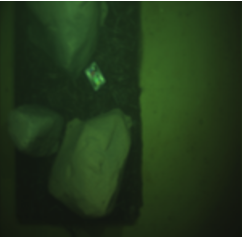
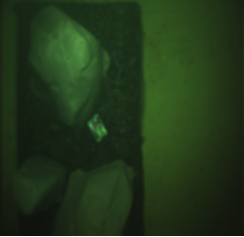


图6：MHL数据集测试结果。

第一行和第三行为原始图像，第二行和第四行为与之对应的结果图。