
《人工神经网络》大作业最终报告

王纪霆

计算机科学与技术系
2015011251

wangjt15@mails.tsinghua.edu.cn

张卡尔

计算机科学与技术系
2015011025

zke_1996@126.com

路橙

计算机科学与技术系
2015010137

c-lu15@mails.tsinghua.edu.cn

Abstract

单一图片的去雾问题是一个十分具有挑战性的任务，因为实际像素的深度和传播图无法从图像中直接得到。基于物理模型的去雾算法多依赖于大量的先验信息，然而在许多极端的例子下可能会失效，普适性并不高。另一种基于深度神经网络的监督学习方法可以很好的估计图像的传播图，然而对训练的数据依赖性过强，不具有灵活性。本文提出了Dehazing Conditional Generative Adversarial Network(DHCGAN)网络，一种弱监督学习网络，利用CGAN训练，通过多尺度CNN提取图像特征，并使用感知损失函数训练，同时解决了物理模型本身的缺陷和纯监督模型过于依赖训练数据的问题。本文使用真实的带雾图片对网络进行测试，并计算PSNR、SSIM、UQI等多种量化评价方式评判去雾效果，验证了本网络具有良好的去雾效果。

1 引言

复杂光照环境下的图像信息提取一直是计算机视觉领域的一个很重要的问题。视野范围近、对比度差、颜色偏差、缺少锐度等问题破坏了原有图像的美感，同时也增大了计算机提取原有信息的难度。雾霾是一种常见的天气现象。雾霾中的微小颗粒使其中的光线反射、折射和散射，导致视野范围变近、对比度下降、颜色偏暗、锐度降低。因此，图像去雾问题引起了计算机视觉领域的广泛关注[6][8]。

雾霾图像的数学表达式是由Koschmieder[15]首先提出，并本文中所有的相关工作都用到了该表达式。该表达式如下：

$$I(x) = J(x)t(x) + A(1 - t(x)) \quad (1)$$

其中， $I(x)$ 为观测到的带雾的图像， $J(x)$ 为原始无雾图像， A 为环境光， $t(x)$ 透射率分布图。透射率分布图上每个像素点的值与实际对应点到观测点的深度 $d(x)$ 呈指数正比：

$$t(x) = e^{-\beta d(x)} \quad (2)$$

从公式(1)中我们可以观察到，雾图是原图与全局光照的叠加，而叠加的比例就是透射率分布图 $t(x)$ ，因此只要得到合适的透射率分布图，就可以将带雾的图片转化为无雾的图。然而对于给定的带雾图片，我们很难直接得知每个像素的深度。因此，估计深度和透射率分布图是本文的主要工作。

有许多成功的去雾算法都是通过估计透射率分布图来去雾的，例如，著名的暗通道算法[8]和近年提出的Non-Local Image Dehazing[10]算法，都是基于物理模型的算法。然而这类

算法在光照环境复杂或者包含非常规光时效果却不理想，例如当物体的光和环境光接近暗通道算法效果不好，而当环境光明显亮与物体本身时，Non—Local算法效果不好。

近年来图片去雾算法的研究多为通过卷积神经网络估计透射率分布图或者直接估计生成去雾图。尽管这类监督学习模型在雾图-原图一一对应的训练集上表现的非常出色，但实际上图片去雾的目标并非是要去雾图和原图完全一致，而是要去雾图和原图等效，即通过人眼观察，去雾图成功去雾而且十分真实地表现原有信息。因此，我们基于条件生成对抗网络[1](CGAN)提出了Dehazing Conditional Generative Adversarial Network(DHCGAN)网络，一种弱监督学习网络，同时解决了物理模型本身的缺陷和纯监督模型过于依赖训练数据的问题，并通过多尺度CNN提取图像特征，使用感知损失函数训练。相比于其他纯粹为了从雾图到原图变换的模型，我们的模型能够生成让人看上去感觉真实的去雾图，而非要和原图完全一样去雾图。

在本文中，我们主要做了一下几方面的工作：

- 我们提出了一种基于条件生成对抗网络、多尺度CNN网络和雾霾图像物理表达式的端到端的Dehazing Conditional Generative Adversarial Network(DHCGAN)网络，传播图和去雾结果可以直接通过损失函数得到。训练时使用对抗学习的方法和感知损失函数进行优化。
- 我们针对去雾领域缺乏训练数据的问题，收集了大量的带有深度信息的室内图像，并对每张图使用造雾算法生成两个雾图，共4496张雾图。
- 生成对抗网络训练难度比一般的网络要大，因此我们使用了多种损失函数并调整比例。
- 我们对模型生成的图像以及原图，采用Peak Signal to Noise Ratio(PSNR), Structural Similarity Index(SSIM), Universal Quality Index(UQI)等多种量化评价方式评判去雾效果。
- 我们选取了多张有特点的真实图片，将我们的网络和现有的其他网络的结果进行比较。

2 相关工作

2.1 单一图像去雾

在以往的图像去雾算法中，基于强先验假设的物理模型占据了大多数。[9]基于统计信息的规律，假定图像中深度信息变化是平滑的，且无雾图的对比度相较于有雾图而言要高得多，因此将雾图做白平衡处理后，通过最大化局部对比度求解图像的深度信息与局部光照。[11]将大气散射模型[12]中的未知图像 $J(x)$ 分解成表面反射率 R 与浓度系数 t 的乘积，并假设 R 与 t 局部不相关，从而求解透射率分布图。[8]基于暗通道先验理论提出了一种高效、普适性高的去雾算法，且之后的去雾算法中很多基于暗通道理论进行扩展。[13]将大气散射模型中的景深信息分解为反射率与景深深度，并假定二者相互独立，通过最大化后验概率求得景深。[14]提出了基于颜色衰减先验的线性模型来估计景深，即假定深度和亮度与饱和度之差正相关，从而通过极大似然估计求解景深。与此相对的是，[10]提出了基于全局的颜色先验，即通过统计发现雾图的RGB线性空间中的线性相关度较高，因此雾图可被几百个单个颜色代表，通过雾线的迭代估计透射率分布图。

这些去雾算法都基于通过统计信息规律得到的先验假设，在一些极端的例子下可能会失效，普适性并不很高。并且，有些算法需要的运算量较大，实时性也相对较差。此外，这些算法都基于一定的物理模型，通过先验假设和优化算法计算出透射率分布图。[17]首次提出了用来估计透射率分布图的端到端的CNN网络，摒弃了所有的先验假设。[18]提出了多尺度的CNN网络，首先训练一个粗尺度的模型对透射率分布图进行粗略估计，接着用该模型的输出作为另一模型的中间输入，可以更精确地估计透射率分布图。然而，这些模型对网络的训练方向只有一个，即让网络估计出的透射率分布图尽可能地接近原图的透射率分布图。然而，在生成去雾图时，需要通过透射率分布图和雾图估计大气光照，目前常用的方法为通过透射率分布图中最小的0.1%的像素点在原图的RGB值估计大气光照[8]，但这样的估计假定了全图中最亮的点深度为无穷大，因此在近景去雾的情形下并不适用。在本文中，我们通过条件生成对抗网络[1]中的判别器对网络进行判别，使得网络生成的去雾图并不一定严格与原图相同，而只需要混淆判别器即可，这样我们的网络在一定程度上降低了先验假设的限制。

2.2 生成式对抗网络(GAN)

近年来,生成式对抗网络[16](以下简称GAN)在图像生成中取得了很大的成效。在GAN中,判别器与生成器同时进行训练,判别器的任务是尽可能地辨别真实图片与生成图片,而生成器的任务则是尽可能地生成出让判别器无法辨别的图片。特别地,条件生成式对抗网络[1](以下简称CGAN)通过对生成器添加条件变量,进一步提高了生成器的性能。[2]使用CGAN进行图像去雨,将雨图作为生成器的条件变量,生成器通过该条件生成去雨图,再由判别器进行判别。在本文中,我们的CGAN网络与此类似,将雾图作为条件变量,生成器生成对应的去雾图。但由于图像中雾的信息与深度信息相关,因而生成器不再适合使用残差网络,我们的生成器使用了multi-scale CNN进行训练。

2.3 损失函数

损失函数作为神经网络训练过程中最主要的部分,在神经网络的拟合中起着重要的作用。通常而言,神经网络的损失函数使用L-1正则或L-2正则(Euclidean loss),通过逐像素计算结果图与原图的差距而求得损失。但正因为计算过程为逐像素计算,会损失图像的连续信息和语义信息。为了克服这一缺陷,我们新增加了来自判别器的损失函数Adversarial loss[16]和感知损失函数Perceptual loss[19],来对生成器的估计进行修正。

3 方法

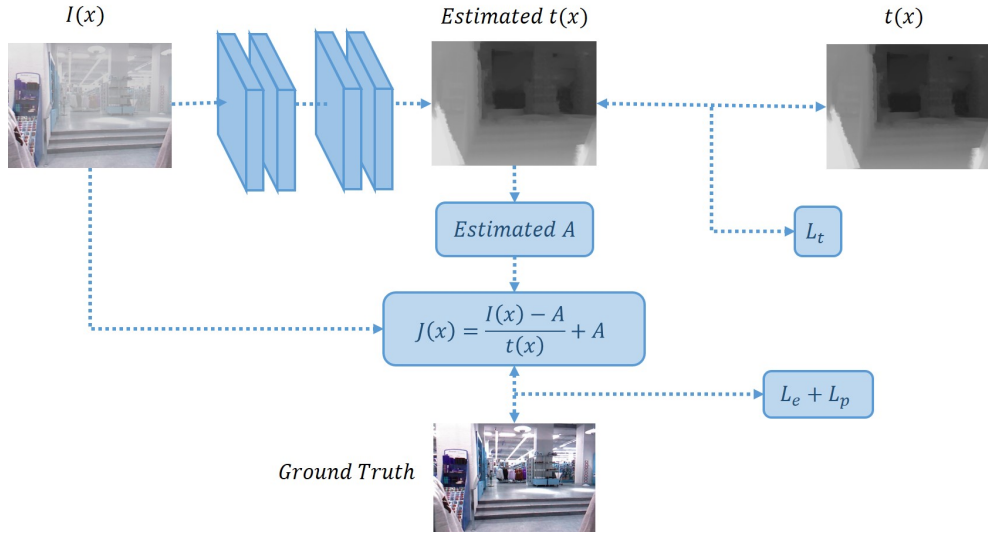


Figure 1: 生成模型G网络结构示意图

我们使用CGAN框架进行单图片去雾模型的构建与训练。框架由生成模型G与判别模型D两个子网络构成,其中G网络根据输入带雾图 x ,能够输出相应的去雾图片 $G(x)$,而D网络则希望通过训练区分G生成雾图与实际图片 y 。具体而言,我们需要完成以下优化问题:

$$\min_G \max_D \lambda_d (\log(1 - D(x, G(x))) + \log D(x, y)) + L \quad (3)$$

其中,为了增加模型的稳定性,我们使用自定义的损失函数 L 以协助GAN的训练过程, λ_d 用于平衡 L 与判别模型 D 之间的权重。各部分的结构将于以下详述。

3.1 生成模型G

生成模型的结构与功能如图1所示。其包含三个部分:透射率分布图估计,环境光估计,以及去雾图片生成。

为了从雾图中恢复原图，我们从雾生成的表达式出发，可以得到

$$J(x) = \frac{I(x) - A}{t(x)} + A \quad (4)$$

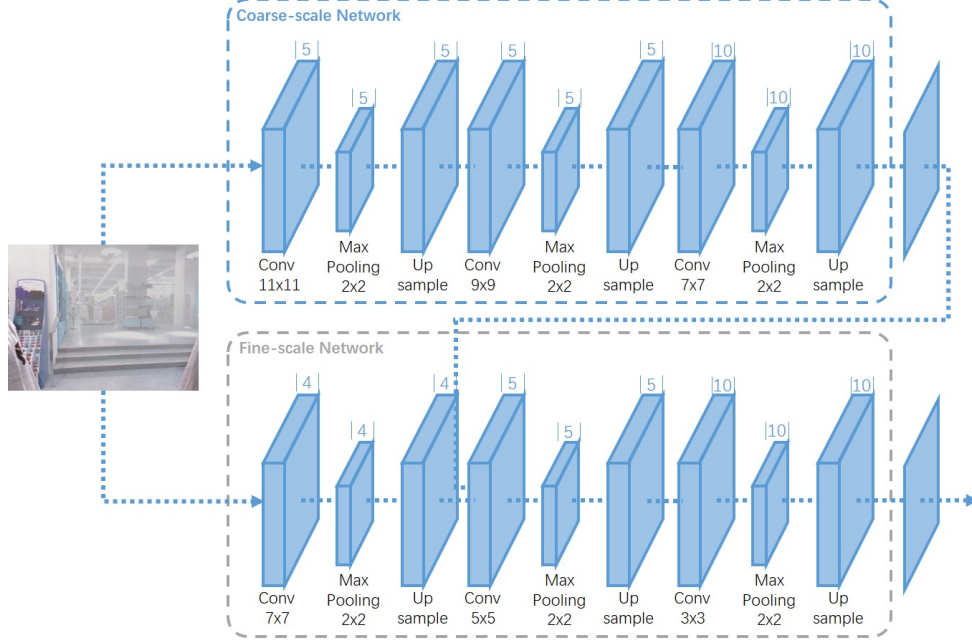


Figure 2: 用于透射率分布图估计的多尺度卷积神经网络

为使用此式，我们使用卷积神经网络，根据雾图产生相对应的透射率分布图 $t(x)$ 。网络结构如图2所示，由于传输图的估计同时需要全局尺度上与细节尺度上的信息，我们使用两个不同尺度的网络分别实现这两种所需特征的提取。粗尺度网络通过较大的卷积核，获取图像较大范围内粗糙的深度图信息，再与细尺度网络结合得出透射率分布图的估计。

对于环境光 A ，考虑到式1中，当 $t(x)$ 接近于0时， $J(x)$ 也将接近于 A 。因此，我们假设图片的深度足够，使用全图中 $t(x)$ 最小的0.1%的像素点作为对 A 的估计。

以上估计完成之后，即可根据4式完成图像去雾，为了避免 $t(x)$ 接近于0时产生的数值计算问题，最终图片恢复公式为

$$J(x) = \frac{I(x) - A}{\max\{0.1, t(x)\}} + A \quad (5)$$

3.2 判别模型D

判别模型用于判断生成图片是否为真实图片，以与生成模型对抗，生成更加真实的图片。我们使用带Batch Normalization与PReLU的多层卷积实现判别模型，其具体结构如图3所示。

3.3 辅助损失函数L

仅使用判别模型与生成模型，其训练结果将很难稳定。传统的逐像素比较易于产生噪声，并不能保证生成图像与真实图像在视觉上的一致性。对此，一种解决方案是使用感知损失 L_p ，将其与逐像素的欧几里得损失 L_e 加权，作为辅助损失函数 L 。而由于我们的生成模型中产生了中间结果 $t(x)$ ，我们可以将其与实际的透射率分布图进行比较，也加入损失函数中：

$$L = \lambda_e L_e + \lambda_p L_p + \lambda_t L_t \quad (6)$$

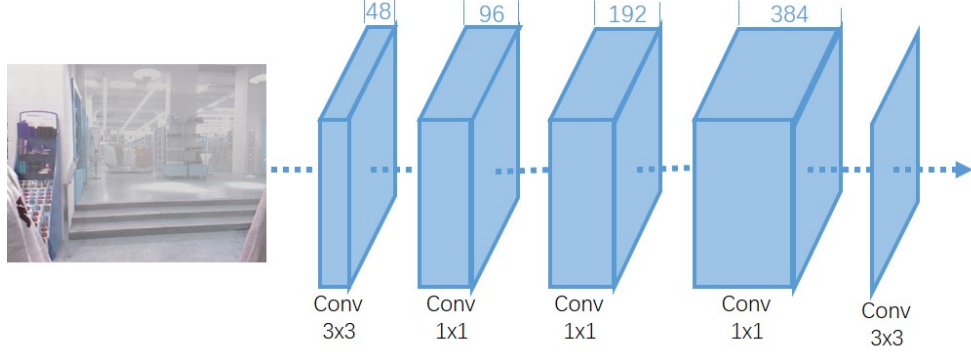


Figure 3: 判别模型D网络结构示意图

其中欧几里得损失用于衡量生成图像 y' 与原图 y 的逐像素差异，设图像的宽、高、通道数分别为 W, H, C ，其定义为：

$$L_e = \frac{1}{WHC} \sum_{i=1}^W \sum_{j=1}^H \sum_{c=1}^C (y'_{i,j,c} - y_{i,j,c})^2 \quad (7)$$

感知损失函数需要使用一高阶非线性层，用于提取图像特征，并进而衡量生成图像、原图的高阶特征之间的差异。在这里，我们使用VGG-16模型中relu2_2层的输出结果作为高阶函数 V ，并设该层的输出结果的宽、高、通道数分别为 W_1, H_1, C_1 ，感知损失函数的完整定义如下：

$$L_p = \frac{1}{W_1 H_1 C_1} \sum_{i=1}^{W_1} \sum_{j=1}^{H_1} \sum_{c=1}^{C_1} (V(y'_{i,j,c}) - V(y_{i,j,c}))^2 \quad (8)$$

最后， L_t 用于衡量估计透射率分布图 t' 与实际分布图 t 之间的差异。对于宽、高分别为 W_2, H_2 的图片，其定义为

$$L_t = \frac{1}{W_2 H_2} \sum_{i=1}^{W_2} \sum_{j=1}^{H_2} (t'_{i,j} - t_{i,j})^2 \quad (9)$$

对于Batch size为 N 的网络，在生成器的损失函数中，来自判别器的损失得到的损失函数为

$$L_a = -\frac{1}{N} \sum_{i=1}^N \log(D(y'_i)) \quad (10)$$

因此，判别器的损失函数为

$$L_g = \lambda_a L_a + L = \lambda_a L_a + \lambda_e L_e + \lambda_p L_p + \lambda_t L_t \quad (11)$$

4 实验

4.1 数据集

我们根据NYU Depth dataset V1中的2248幅原始图像以及Kinect所提供的深度图，在式1中，取 $A = (k, k, k), k \in [0.6, 0.8]$ ，在式2中，取 $\beta \in [-0.25, -0.13]$ ，为每个原始图像生成2幅带雾图，共4496幅作为训练集。此外，我们使用此文[6]中提供的数据集，由NYU Depth dataset V2生成的共1449组数据中，取300幅作为验证集。为了使测试数据与训练数据有分布上的区别，我们使用Middlebury stereo database[7]所生成的20张图片用于测试最终成果。

4.2 模型与训练参数

整个网络使用Tensorflow在一块Tesla M60上进行训练。其中 $L_e = 1$, $L_p = 0.1$, $L_t = 40$, $L_a = 0.005$ 。Learning rate为0.001, 使用随机梯度下降法进行训练, Momentum为0.9, Learning rate每经过400个iteration乘以0.96。训练过程中batch size为4, 共训练了80个epoch。

4.3 消融分析

为了更好地证明模型的合理性, 我们采取了如下对比实验:

- 模型的所有参数均调整为默认值 (即4.2中的参数取值)
- 不使用 L_e , 其余参数均为默认值
- 不使用 L_p , 其余参数均为默认值
- 不使用 L_a , 其余参数均为默认值
- 不使用 L_t , 其余参数均为默认值

下面七张图为不同模型在训练过程中, 不同数值的变化对比图:

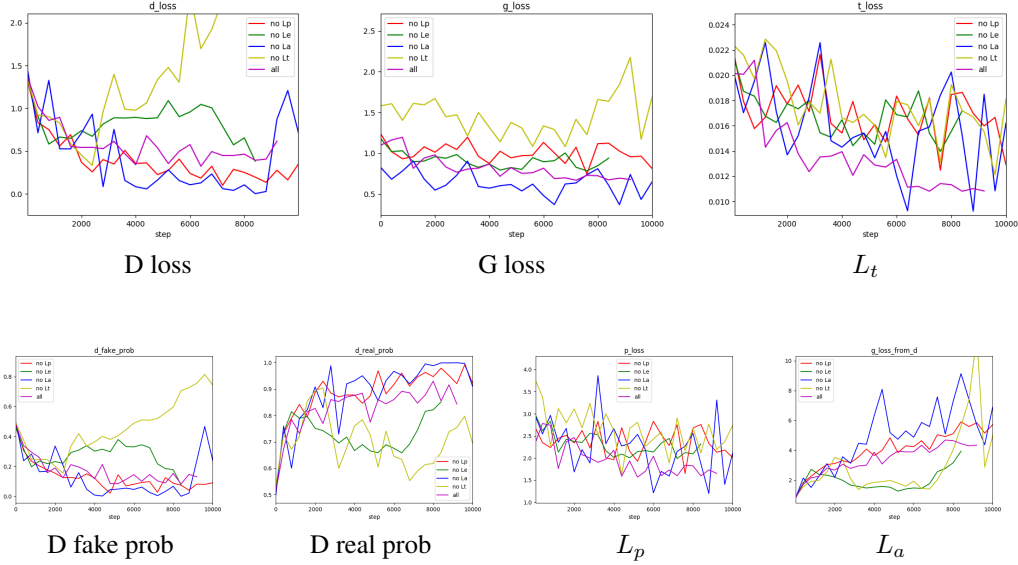


Figure 4: 模型训练过程对比图

根据不同数值的变化图以及在测试集上的效果图, 我们可以得出以下结论:

1. 分析 L_t 对结果的影响:
由图可知, 当缺少深度信息的损失影响时, 模型让判别器无法分辨, 但与此同时, 生成器与判别器的loss都在飞快上涨, 且最终效果几乎没有去雾, 因此 L_t 会修正CGAN的不稳定性, 从而保证CGAN更容易训练。
2. 分析 L_p 对结果的影响:
由图可知, 当缺少感知损失函数的修正时, 生成图的色彩鲜艳度有突变, 即不能保证图像的色彩均匀变化。此外, 由于训练过程中 L_p 占比例较小, 因此在大型场景中的体现并不明显。
3. 分析 L_a 对结果的影响:
由图可知, 当缺少GAN对生成器的影响时, 尽管生成器和判别器的损失函数的值都很小, 但生成的图效果却很差。由此可以猜测, 我们的模型在训练集上训练得到的只是判断近景的深度信息, 但这种深度信息在远景的观察中效果较差, 必须通过判别器的修正来让模型更具有普适性, 即不需要让模型十分接近去雾的原图, 只需要让模型朝着判别器无法分辨的方向训练即可。

4. 分析 L_e 对结果的影响:

由图可知, 当缺少原图的修正时, 模型整体效果偏暗, 且感知损失也较差。然而, 由于此时模型并不朝着更像原图的方向进行训练, 不需要朝着一个方向训练, 因而更容易受到判别器的影响, 从而让判别器的loss下降较慢, 即生成的图在一定程度上是无雾, 但却整体偏暗。

因此, 我们的模型中提出的各个损失函数都对模型去雾产生了一定的影响, 通过这些损失函数, 我们的模型在一定程度上达到了去雾的效果。



Figure 5: 不同损失函数去雾效果对比图

4.4 与baseline的比较

	He[8]	Ours
SSIM	0.7641	0.6943
PSNR	17.36	13.48

Table 1: 测试集上的量化指标

我们对测试集中图片进行定量衡量, 使用Dark channel算法作为对比, 其结果如表1所示。我们的训练集与测试集来源于不同的两组数据集, 其雾的产生与背景分布都有一定差异, 而Dark channel算法则不受数据分布差异影响。因此, 可以看出, 在定量分析上我们模型的效果并没有超过Dark channel的表现。

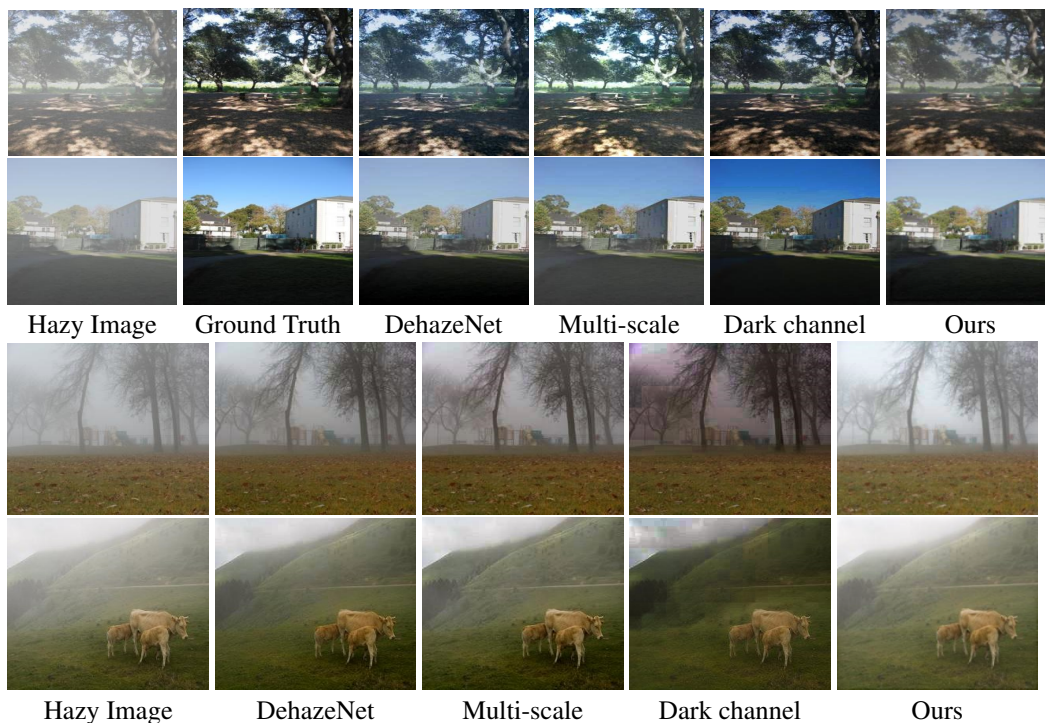


Figure 6: 模型去雾效果对比

此外，我们在真实场景中，将模型与其他算法得出的结果进行比较，结果如图6所示。由于真实场景没有可以对照的量化指标，因此只能通过视觉直观分析。可见，我们的模型在去雾效果方面，由于训练集均为室内场景，对于远景的去雾效果并不十分理想；但在较近场景有着较好的表现。

而如Dark channel等依赖数学的算法，在这里改变了图片整体的色调；这使得其结果偏暗或是偏红，与真实场景产生了较大差距，效果并不理想。

5 结论

本次大作业中，我们学习到了关于图像到图像的multi-CNN网络、生成对抗模型（GAN）、感知损失函数、图像处理等方面的诸多知识，自己用tensorflow + tensorlayer搭建了CGAN的框架，神经网络工具的熟练度更高了，收获很大。这次大作业持续了近半学期的时间，从一开始的选题讨论、文献调研，到中期编写代码，不断尝试和借鉴不同的网络结构，等待训练结果，到最后总结分析，我们体验到了一个近乎完整的科研项目研究过程。从中收获的不仅仅是文献中的先进神经网络理论知识，更是一些科研方法和态度上的经验。积极学习他人的经验，站在巨人的肩膀上，并从中受到启发，思考改进空间，付诸行动并有所收获，这是我们感受到的最真实的科研经历。我们的大作业模型仍有很多可以进步的空间，比如尽管我们投入了不少资金在租用运算设备上，我们今后仍可以尝试使用运算能力更强的服务器。我们在模型搭建时没有借鉴其他成熟GAN的网络（如pix2pix），因此在调参时经历了十分痛苦的阶段。但是正是调参的过程，让我们对GAN和Perceptual loss的理解更深，我们也更换了几种不同的生成器的框架，发现了残差网络的适用特点。通过调参和对模型的调整，我们对神经网络结构的理解更深入了。此外，由于大作业是小组合作，我们总体的分工也很明确，感受到了小组共同讨论、参与科研的快乐。

我们的github链接是：<https://github.com/maoyuchaxue/ANN-dehazing>

参考文献

- [1] Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. *Computer Science*, 2672-2680.
- [2] Zhang, H., Sindagi, V., & Patel, V. M. (2017). Image de-raining using a conditional generative adversarial network.
- [3] X.-J. Mao, C. Shen, and Y.-B. Yang, "Image denoising using very deep fully convolutional encoder-decoder networks with symmetric skip connections," *arXiv preprint arXiv:1603.09056*, 2016.
- [4] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *European Conference on Computer Vision*. Springer, 2016, pp. 391-407.
- [5] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.
- [6] Ancuti, C., Ancuti, C. O., & Vleeschouwer, C. D. (2016). D-HAZY: A dataset to evaluate quantitatively dehazing algorithms. *IEEE International Conference on Image Processing* (pp.2226-2230). IEEE.
- [7] Scharstein, D., Hirschmüller, H., Kitajima, Y., Krathwohl, G., Nešić, N., Wang, X., et al. (2014). High-Resolution Stereo Datasets with Subpixel-Accurate Ground Truth. *German Conference on Pattern Recognition* (Vol.8753, pp.31-42). Springer International Publishing.
- [8] He, K., Sun, J., & Tang, X. (2009). Single image haze removal using dark channel prior. *Computer Vision and Pattern Recognition, 2009. CVPR 2009*. IEEE Conference on (Vol.33, pp.1956-1963). IEEE.
- [9] Tan, R. T. (2008). Visibility in bad weather from a single image. *Computer Vision and Pattern Recognition, 2008. CVPR 2008*. IEEE Conference on (pp.1-8). IEEE.
- [10] Berman, D., Treibitz, T., Avidan, S. (2016). Non-local Image Dehazing. *IEEE Conference on Computer Vision and Pattern Recognition* (pp.1674-1682). IEEE Computer Society.
- [11] Fattal, R. (2008). Single image dehazing. *Acm Siggraph* (Vol.27, pp.1-9). ACM.
- [12] Land, E. H. (1964). The retinex. *American Scientist*, 52(2), 247-264.
- [13] Kratz, L., Nishino, K. (2009). Factorizing Scene Albedo and Depth from a Single Foggy Image. *IEEE, International Conference on Computer Vision* (Vol.30, pp.1701-1708). IEEE.

- [14] Zhu, Q., Mai, J., Shao, L. (2015). A fast single image haze removal algorithm using color attenuation prior. *IEEE Transactions on Image Processing*, 24(11), 3522-3533.
- [15] Koschmeider, H. (1924). Theorie der horizontalen sichtweite. *Beitr.phys.freien Atm*, 12.
- [16] Goodfellow, I. J., Pougetabadie, J., Mirza, M., Xu, B., Wardefarley, D., Ozair, S., et al. (2014). Generative adversarial networks. *Advances in Neural Information Processing Systems*, 3, 2672-2680.
- [17] Cai, B., Xu, X., Jia, K., Qing, C., Tao, D. (2016). Dehazenet: an end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11), 5187-5198.
- [18] Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., Yang, M. H. (2016). Single Image Dehazing via Multi-scale Convolutional Neural Networks. *Computer Vision – ECCV 2016*. Springer International Publishing.
- [19] Johnson, J., Alahi, A., Li, F. F. (2016). Perceptual losses for real-time style transfer and super-resolution. 694-711.