

基于条件生成对抗网络的 HDR 图像生成方法

贝悦¹, 王琦¹, 程志鹏¹, 潘兴浩¹, 杨默涵², 丁丹丹^{2,*}

(1. 咪咕视讯科技有限公司, 上海 201201; 2. 北京观止创想科技有限公司, 北京 100036)

摘 要: 高动态范围(HDR)图像相比低动态范围(LDR)图像有更宽的色域和更高的亮度范围, 更符合人眼视觉效果, 但由于目前的图像采集设备大都是 LDR 设备, 导致 HDR 图像资源匮乏, 解决问题的一种有效途径是通过逆色调映射将 LDR 图像映射为 HDR 图像。提出了一种基于条件生成对抗网络(CGAN)的逆色调映射算法, 以重建 HDR 图像。为此, 设计了基于多分支的生成对抗网络与基于鉴别块的鉴别网络, 并利用 CGAN 的数据生成能力和特征提取能力, 将单张 LDR 图像从 BT.709 色域映射到对应的 BT.2020 色域。实验结果表明: 与现有方法相比, 所提出的网络能够获得更高的客观与主观质量, 特别是针对低色域中的模糊区域, 所提方法能够重建出更清晰的纹理与细节。

关 键 词: 条件生成对抗网络; 卷积神经网络; 逆色调映射; 色域转换; 特征提取

中图分类号: TP391

文献标识码: A

随着显示设备技术产业的发展, 普通低动态范围(Low Dynamic Range, LDR)的显示器逐渐不能满足大众影音需求, 具有更高亮度、更广色域的高动态范围(High Dynamic Range, HDR)显示设备逐渐出现。与普通的 LDR 图像相比, HDR 图像的编码位数由 8 bit 提升至 10 bit 或以上, 图像色域范围由 BT.709 标准提升至 BT.2020 标准, 能提供更多的动态范围和图像细节。然而, 目前与 HDR 显示设备相适配的视频图像资源十分匮乏。因此, 通过逆色调映射, 将原本的 LDR 图像生成对应的高质量 HDR 图像, 也就是逆色调映射, 是有效提升原有图像视频质量的有效途径之一^[1]。

从 LDR 图像到 HDR 图像的映射是一个非线性的、不适定问题, 存在多种解。与 LDR 图像相比, HDR 图像色域更广, 广义上的 HDR 图像色域应能覆盖到 90% 以上的 DCI-P3 色域范围; 位深由 LDR 图像的 8 bit 增加到 10 bit。因此, 在不

同场景下, 由 LDR 图像生成对应的 HDR 图像需要良好的对应点像素生成能力。由此启发可以使用神经网络来完成上述逆映射。

卷积神经网络 (Convolutional Neural Networks, CNN) 首先被用于上述逆色调映射, 以完成 HDR 重建。实验结果表明, 一些基于 CNN 的方法能够获得比传统方法更高的性能^[1-7]。但是, 这些基于 CNN 的方法通常涉及大量参数, 因而在训练中会花费很长时间。此外, 在网络训练中往往使用 L2 损失函数, 虽然获得了较高的峰值信噪比(Peak Signal-to-Noise Ratio, PSNR), 但图像主观质量偏于平滑, 降低了 HDR 的观看体验。近年来, 为克服上述问题, 研究人员提出使用基于对抗生成网络 (Generative Adversarial Network, GAN) 的 HDR 重建方法。以往工作^[1-11]一般都使用 U-net 做为 GAN 的生成器, 使用卷积配合降采样的结构作为鉴别器, 提高了重建 HDR 的主观性能。这些研究表明了 GAN 网络在 HDR

收稿日期: 2020-09-14; 录用日期: 2021-04-23; 网络出版时间: 2021-05-17 17:00

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210517.1523.003.html

基金项目: 浙江省自然科学基金 (LY20F010013)

*通讯作者: E-mail: DandanDing@hznu.edu.cn

引用格式:

重建方面具有巨大的潜力。GAN 网络无要求一个假设的数据分布,而是使用一种分布直接进行采样,从而达到理论上可以完全逼近真实数据,这也是 GAN 网络最大的优势。但是,GAN 网络无需预先建模的方法对于较大图像而言,生成目标的可控性不强,无法生成具有特定性质的图像。

受上述工作的启发,本文提出一种基于条件生成对抗网络(Conditional Generative Adversarial Network, CGAN)的从单张 LDR 图像生成对应 HDR 图像的方法。利用 CGAN 网络的思想,通过在 GAN 网络的生成网络和鉴别网络中添加条件性的约束,来解决训练过于自由的问题,实现了由单张 LDR 图像到对应 HDR 图像的端到端的非线性映射。实验结果表明,本文方法不仅能够将 8 bit 的 LDR 图像自然地映射到 10 bit 的 HDR 图像,还能挖掘 LDR 图像中的一些隐藏特征,与以往方法相比,所提方法获得了较高的主观与客观质量。

1 相关工作

基于神经网络方法以端到端的方式解决逆色调映射问题具有巨大潜力。

1.1 多曝光的方法

多曝光的方法是处理逆色调映射的最直接方法,其合并不同曝光拍摄的多个 LDR 图像,最终生成单个 HDR 图像。不过,一般多曝光的图像获取需要使用专业相机,这对于实际应用并不友好,因为实际应用中所涉及的图像几乎都是非专业用户拍摄的单曝光图像。为此,普遍采用的方式是:首先,使用软件算法从单个输入 LDR 图像生成若干个不同曝光的 LDR 图像;然后,使用这些生成的 LDR 图像重建出对应的 HDR 图像。如此,多曝光方法一般包括 2 个步骤,即生成和重建,且每个步骤都可以由神经网络来实现。

基于上述多曝光的思想,Endo 等^[2]开发了一对基于“编码器-解码器”结构的 CNN 网络 DrTMO,包括向上曝光模型和向下曝光模型,分别推断 2 组曝光的 LDR 图像,使用已有的合并方法将这些 LDR 图像合并为一个 HDR 图像。Lee 等^[7]提出了一个 CNN 链式结构,采用 2 个并行的 CNN,每个 CNN 推断 3 个 LDR 图像。后来, Lee 等^[10]基于 GAN 的结构进一步改进了设计,使用 2 个神经网络 G_{plus} 和 G_{minus} 来生成曝光图像。Xu 等^[8]将 GAN 用于 LDR 图像的合并,先利用相机响应曲线来更改输入 LDR 图像的曝光时

间,并获得 2 个曝光过度和 2 个曝光不足的图像,训练了一个 GAN 网络来合并这些 LDR 图像以进行 HDR 图像重建。

1.2 单曝光的方法

相对于多曝光的方法,单曝光的方法将输入的 LDR 图像直接转换为输出的 HDR 图像,灵活性更高。

1.2.1 基于 CNN 的单曝光

Takeuchi 等^[12]提出了一种基于 CNN 的、从 BT.709 到 BT.2020 的色彩空间转换方法。Eilertsen 等^[3]基于自动编码器网络重构输入 LDR 图像中的过饱和区域,所提出的 HDRCNN 方法在图像中的过饱和像素比例低于 5% 时,可以达到较好的效果。Marnierides 等^[5]设计了一个 ExpandNet, LDR 图像通过局部、中等和全局 3 种 CNN 分支后进行融合,得到对应的 HDR 图像。ExpandNet 能够提升一些图片的主观质量,但在某些图像上,会出现颜色偏移或者主观质量不高等问题。Kinoshita 和 Kiya^[6]在现有 U-net 的基础上,添加了一个全局编码器,以和现有的 U-net 编码器并行工作,来自 2 个编码器的结果被一起并送入 U-net 解码器进行 HDR 重建。

1.2.2 基于 GAN 的单曝光

GAN 网络已经在很多方面取得了巨大的进展,如图像超分辨率^[1-14]、图像风格迁移^[1-16]、人体姿态估计^[17]、人体运动传递^[18]等。在逆色调映射方面, GAN 网络刚刚起步。Ning 等^[9]基于 U-Net 设计了生成器,并采用了基于 CNN 的鉴别器。Lee 等^[10]基于 GAN 进行超分辨率与 HDR 的联合重建。以上方法都是基于现成的网络展开,针对逆色调映射,有必要设计新的网络结构配合新的损失函数,以获得更好的重建效果。

1.3 GAN 网络和 CGAN 网络

GAN 网络由 Goodfellow 等^[19]提出,是一种基于对抗生成的博弈思想创造出的模型,其组成有 2 个网络,分别为生成网络 G 和鉴别网络 D。生成网络和鉴别网络相互对抗,所采用的思想是零和博弈思想,即双方在平等的情况下开局,各自使用各自的策略去与对方对抗,在对抗中双方再利用对方的策略及时对自己的策略进行调整。双方充分遵循“损人利己”去实现获得胜利的目的。这个思想拓展到深度学习领域,即生成网络和鉴别网络为博弈的双方,生成网络不断发挥自己拟合数

据的能力去欺骗鉴别网络，而鉴别网络则尽力去辨别生成网络传送来的拟合数据，两者的最终目标是达到纳什平衡^[20]，最终让生成网络学习到真实的数据分布。

如图1所示，生成网络 G 根据给出的标签样本的数据分布，将服从随机分布的噪声 z 包装成真实的样本数据，该过程也被称为生成样本数据的过程。生成网络用该样本数据去欺骗鉴别网络。鉴别网络 D 则用来鉴别生成网络输出的样本数据的真假性。

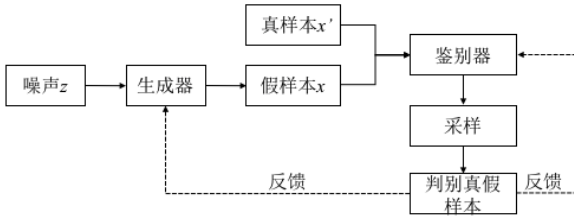


图1 GAN网络基本结构

Fig.1 Basic structure of GAN network

生成器为了能够学习真实的样本分布 $P_{data}(x)$ ，先通过随机分布的噪声 z （一般设置为高斯分布）创建映射 $G(z; \theta_g)$ ，再通过鉴别器的映射 $D(x; \theta_d)$ 判断所生成的数据是否符合真实样本的分布，其中 G 、 D 是由参数为 θ_g 、 θ_d 的神经网络表示的微分函数。在GAN网络中，通常采用的代价函数为

$$\min_G \max_D V(D, G) = E_{x \sim P_{data}(x)} [\log D(x)] + E_{z \sim P_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

式中： x 为真实数据分布；最大化 $D(x)$ 表示鉴别器将尽可能地学习真实样本数据分布； $P_z(z)$ 为输入的噪声分布； $G(z)$ 表示生成器生成的样本数据分布； $D(G(z))$ 表示鉴别器将尽力去鉴别生成器生成的样本数据是否为真。在训练过程中，先将两者之一参数固定，优化另一方，一段时间后交换固定和优化角色。如此反复博弈，生成器和鉴别器最终将达到纳什平衡。

理论上，GAN网络有能够拟合任何复杂样本数据的可能性，但在实际应用中，GAN在拟合复杂样本数据时表现不够稳定。因此，GAN网络被提出。在GAN网络基础上，CGAN网络通过加入条件的方式加强对数据生成的引导，加入的条件是任意的，如标签信息、数据属性信息等。在CGAN网络中，生成网络接收的就是随机噪声 z 和加入的条件 c 。因此，CGAN的代价函数一般

为

$$\min_G \max_D V(D, G) = E_{x \sim P_{data}(x)} [\log D(x|c)] + E_{z \sim P_z(z)} [\log(1 - D(G(z|c)))] \quad (2)$$

通过这种方式，原本的GAN网络从无监督或者半监督的学习转化成有监督的学习。CGAN网络和GAN网络一样都是相互交替训练。CGAN网络从训练鉴别网络开始，再训练生成网络。可见，CGAN网络可以从添加的条件中得到需要的生成信息，从而有利于图像生成任务的进行。

2 所提出的逆色调映射算法

2.1 生成网络

受Marnerides等^[5]采用多分支网络分别提取低频、中频和高频特征等工作的启发，本文所设计的生成网络的结构如图2所示。该模型使用 256×256 大小的LDR图像作为输入，通过3个分支后，对得到的3个特征图进行融合并输出。在layer层中，卷积核尺寸为 3×3 ，边界使用1作为填充，并使用RELU激活函数。由于使用目的不同，步长在不同的分支中略有不同。

1) 分支一。使用了6个layer层和一个 4×4 的卷积层，layer层使用的卷积核大小为 3×3 ，步长为2，这是为了将输入的 256×256 大小的图片汇聚为1个像素点，以获得整张图片的感受野，从而获得色调映射前图像的整体色彩变化，本分支最终输出 $1 \times 1 \times 64$ 尺寸的特征图。

2) 分支二。使用3个layer层和一个 3×3 的卷积层，layer层步长是1，以利于生成网络学习到LDR图像的中频信息的分布。在该网络分支中，输入图像的通道数是3，输出图像的通道数是64。

3) 分支三。仅由2个layer层组成，每个layer层都使用 3×3 的卷积核，步长为1。这使得该分支能够获得 5×5 的感受野，从而能够获得图像局部较为剧烈的变化特征，该分支最终得到 $256 \times 256 \times 128$ 尺寸的特征图。

所输入的LDR图像经过3个分支后，先将分支一获得的能够反映全局色调变化的一个像素点扩充为 256×256 大小，这样是为了在下面的融合中，使图像能够反映映射前全局色调变化，并与另外2个分支生成的特征图按照维度串联成 $256 \times 256 \times 256$ 大小，最终通过一个layer层将特征图数目减少到3，得到 $256 \times 256 \times 3$ 尺寸的生成

图像, 该 layer 层中使用的激活函数是 SELU。

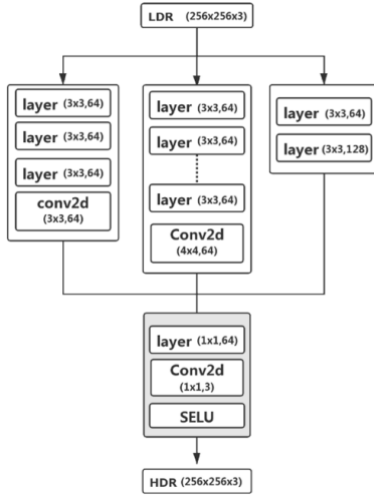


图2 所提出的生成网络结构

Fig.2 Structure of the proposed generative network

2.2 鉴别网络

鉴别网络用于判别生成图像是否符合真实的数据分布。在本文方法中, 鉴别网络的结构如图3所示。

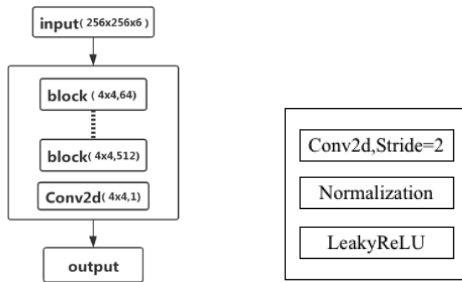


图3 所提出的鉴别网络及其内部鉴别块

Fig.3 Structure of the proposed authentication network and its internal authentication block

将生成网络生成的图片和原 HDR 图像按照通道数串联成 $256 \times 256 \times 6$ 大小, 使用图3所示的鉴别块 (block) 进行下采样。其中, 每个鉴别块使用卷积核为 4×4 , 步长为 2, 边界填充为 1 的卷积层, 并跟随归一化处理, 使用 0.2 的 LeakyRELU 激活函数。图像依次经过 4 个鉴别块, 每经过一个鉴别块后特征图都会缩小 2 倍, 在得到 $32 \times 32 \times 512$ 尺寸的图片之后, 使用一层与鉴别块相同的卷积层。鉴别器将最终的结果进行是或非的判别。

2.3 代价函数

鉴别网络对真实样本的分布通过鉴别后判断为真 (即输出 1), 对生成网络生成的样本进行

鉴别后输出为假 (即输出 0)。生成网络则尽力使自己生成的样本分布获得真的评价 (即获得 1)。因此, 网络的一部分损失函数可以表示为式 (2)。一方面, 鉴别网络要对真实的样本分布经过鉴别网络后生成的分布与相同大小的全 1 矩阵做交叉熵, 用于识别真实的样本分布; 另一方面, 生成网络生成的样本分布经过鉴别网络后, 与相同大小的全 0 矩阵做交叉熵, 用于使得鉴别器尽可能地识别虚假的数据分布。为了使生成网络生成的数据分布能够尽力欺骗鉴别网络, 将生成网络生成的数据分布与全 1 矩阵做交叉熵后再送入鉴别网络。此外, 在生成网络的损失函数中, 使用均方误差 (Mean Square Error, MSE) 进行逐像素的差别计算, 以保证生成图像和真实样本图像之间的相似性。其中, MSE 的表达如下:

$$\text{MSE} = \frac{1}{2n} \sum_{i=1}^n \|y(x_i) - a(x_i)\|^2 \quad (3)$$

式中: y 为真实的数据分布; a 为生成网络生成的数据分布; x_i 为对应的像素分布。

最终, 整个网络的损失函数为

$$\text{Loss} = \min_G \max_D V(G, D) + \text{MSE} \quad (4)$$

3 实验结果与分析

3.1 实验设置

通过从海量视频中收集了 8262 张 HDR 高清图像, 在逐一剔除不良数据之后, 选取了 6736 张大图像, 再将这些大图像逐一剪裁成 30312 张尺寸为 960×480 的小图像。在剔除相似场景后, 最终得到 7632 张 960×480 大小的图像。本实验选取其中 4000 张图像, 并随机剪裁成 256×256 大小, 用来作为标签数据。对于这些图像, 使用 OpenCV 中的色调映射算法随机进行映射, 得到的图像被作为对应的 LDR 图像样本。除以上 4000 张训练图像外, 从获取的 7632 张图像中去掉训练集后, 随机选取不重复的 20 张图像作为测试集, 测试集的缩略图如图 4 所示。

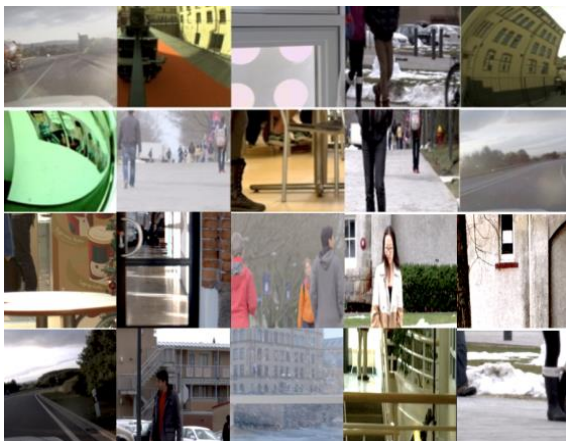


图4 实验所使用的 20 张 LDR 测试图片

Fig.4 20 LDR test pictures used in this experiment

本实验使用 Adam 优化器，生成网络的学习率设为 10^{-3} ，鉴别网络的学习率设为 5×10^{-4} 。在训练的过程中，即使已经将生成网络的学习率设置成鉴别网络的 2 倍，判别网络也总能轻而易举地在对抗中取得上风，这样就会导致生成网络的损失函数几乎没有下降，产生 GAN 网络常见的梯度消失问题，损失函数容易陷入局部最优解。因此，训练中，在更新 k 次生成网络参数的同时只更新一次鉴别网络的参数，以保证控制鉴别网络的鉴别能力与生成网络的生成能力尽量持平。如果 k 值过大，就会让生成网络的损失函数的值来回震荡； k 值过小，鉴别网络会达到局部最优解，产生梯度消失问题。在实验中，根据训练集大小，让生成网络更新 5 次，鉴别网络更新 1 次。如此，模型能够较快收敛且重建图像的质量较为稳定。

实验基于 Pytorch 平台展开，使用 CPU i9-9900K CPU@3.60GHz 处理器，32GB 运行内存，NVIDIA GeForce RTX 2080Ti GPU 进行训练与测试。网络训练时间约为 40h。

3.2 实验结果

使用所得到的模型对图 4 所示的 20 张图像进行测试，并给出了客观与主观测试结果。

使用 PSNR、MPSNR、SSIM、MS-SSIM 及 HDR-VDP 共 5 个客观指标对 HDR 重建图像的质量进行评价。其中，MS-SSIM 客观反映了图像在多尺度上的结构相似性，HDR-VDP 用算法来模拟人眼观看图像的过程。MS-SSIM 与 HDR-VDP 是用来评价 HDR 重建图像质量的重要指标。使用不同的方法对图 4 所示的测试集进行测试，所得到的平均值如表 1 所示。可以看出，与以往方

法 DrTMO^[2]、ExpandNet^[5]与 HDRCNN^[3]相比，本文方法获得了最高 SSIM、MS-SSIM 与 HDR-VDP 值，反映了所提方法能够较好地重建出 HDR 图像。

此外，还对比了不同方法所得到的 HDR 图像的主观质量。由于普通显示器无法显示 HDR 图像，将所有方法得到的 HDR 图像都通过同样的方法再映射为 LDR 图像，以使它们能够在普通显示器上可视化。如图 5 所示，DrTMO^[2]方法生成的图像偏白，与原 HDR 图像差距较大；ExpandNet^[5]在有些图像上有严重的色度失真，影响了主观质量；HDRCNN^[3]对于曝光区域低于 5% 的图像有较好效果，但是也会产生色度偏移，如第二行右上角的天空有明显的色度偏移。与上述方法相比，本文方法可以得到质量较高的 HDR 图像，图像看上去自然，与原 HDR 图像较为接近。ExpandNet^[5]方法中出现的色彩偏移等现象也得到了解决。此外，所训练得到的模型不仅仅针对某一类场景，在各种场景，各种亮度条件下都能得到较好质量的重建图像。如图 6 所示，在 LDR 图像的 BT.709 色域难以显示的模糊特征也能够再所得到的 HDR 图像中有更加清晰的表达；如图 7 所示，在明亮场景下重建的图像色彩自然，模糊区域也有一定程度的亮度提升，并且没有明显的色彩偏移现象。

4 消融实验

为了验证多分支网络的有效性，对生成网络进行了消融研究。

首先，去掉了提取中频信息的分支，所得实验结果如表 2 所示。可以看出，去掉中频分支后的生成网络所获得的客观指标数值均低于所提出的三支网络，获得的主观图像质量的色彩饱和度不足，亮度不够高，如图 8 (a) 所示。

然后，为进一步验证多分支网络的有效性，分别对 3 个分支进行了单独训练，实验结果如表 2 所示。图 8 展示了保留不同分支后的主观结果，可以看出，单独分支所生成的图像存在明显的偏色，且纹理细节不够丰富。值得注意的是，由于低频分支会将输入的 256×256 大小的图像汇聚到 1 个像素点，以获得整张图片的感受野，低频分支最终的输出尺寸是 1×1 。为直观地展示低频分支的输出图像，将其放大为 256×256 大小的图像，如图 8 (d) 所示。4 种消融实验证明，本文所提

出的生成网络能够显著提高图像的质量,使得生成图像色彩饱和度更高,主观质量更好,更加接近参考图像。

此外,将3个分支输出的特征图进行了可视化,结果如图9所示。

5 结论

基于GAN网络架构,本文提出了一种LDR图像到HDR图像的逆色调映射方法。(1)为此,我们设计了一种新的生成网络,使用多分支结构,

提取图像低频、中频、高频特征,并将不同尺度的特征进行融合得到最终的HDR图像;(2)设计了一种新的鉴别器网络,使用组合的鉴别块完成鉴别任务。(3)实验结果表明,所提出的方法在客观指标上超过了以往的方法,尤其是在直接反映HDR重建图像质量的MS-SSIM与HDR-VDP指标上取得了较高性能。在主观质量上,所重建的HDR图像主观质量自然并且无显著色彩失真。

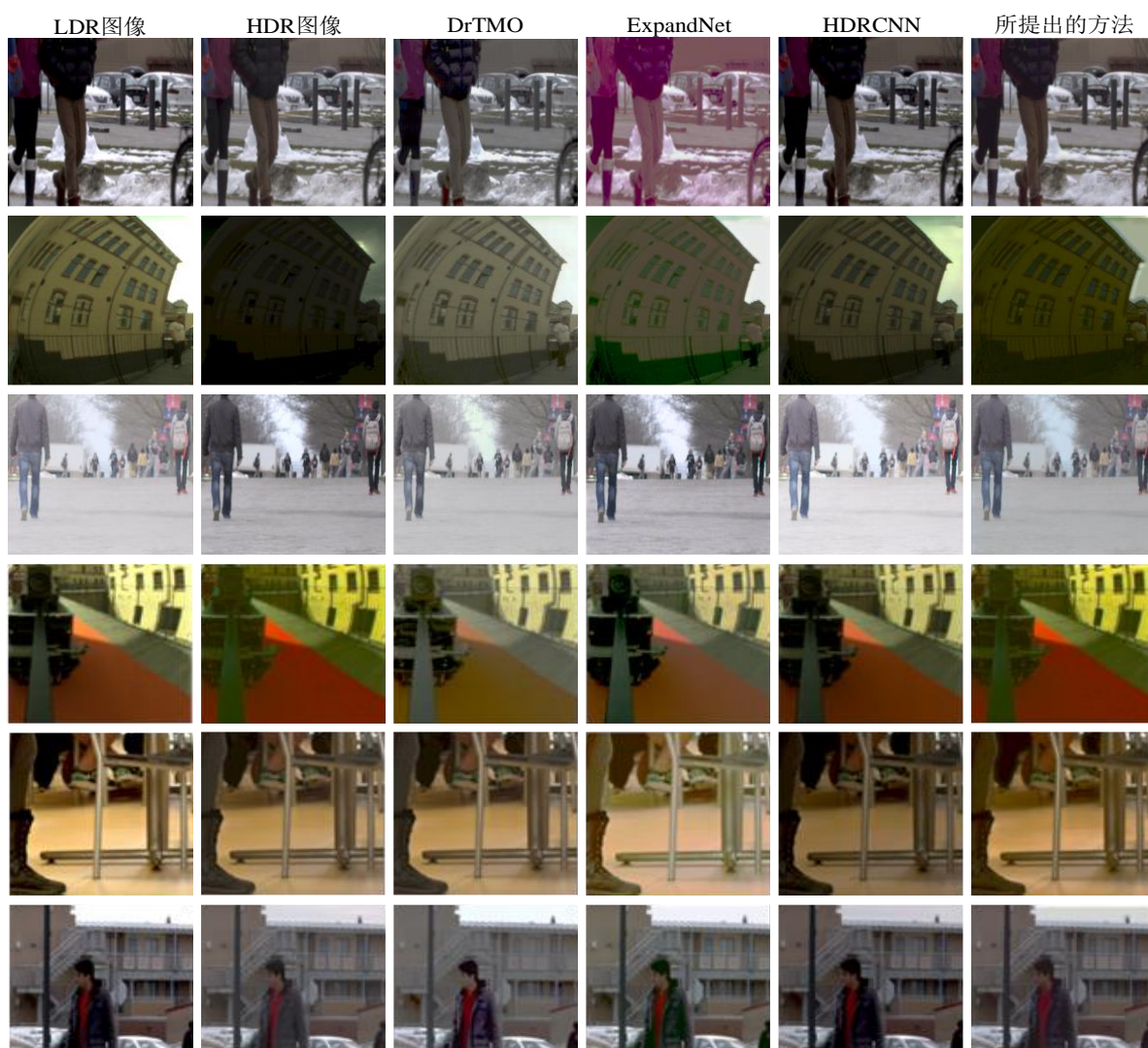


图5 不同方法得到的HDR图像的主观效果对比

Fig.5 Comparison of subjective effects of HDR images obtained by different methods

表1 不同方法的客观性能比较

Table 1 Comparison of objective performance among different methods

方法	PSNR	MPSNR	SSIM	MS-SSIM	HDR-VDP-2
DrTMO ^[2]	22.31	22.44	0.58	0.59	63.79
ExpandNet ^[5]	23.61	23.79	0.70	0.71	78.02
HDRCNN ^[3]	25.70	25.95	0.60	0.63	70.93
所提出的方法	24.99	25.26	0.71	0.77	78.67

表 2 消融实验：网络中不同分支的客观性能比较

Table 2 Comparison of objective performance of different branches of network in ablation experiment

方法	PSNR	MPSNR	SSIM	MS-SSIM	HDR-VDP-2
所提出的方法	24.99	25.26	0.71	0.77	78.67
去掉中频分支	22.06	22.46	0.61	0.65	77.73
仅中频分支	22.36	30.46	0.42	0.45	77.54
仅高频分支	22.84	42.44	0.45	0.47	78.72


 图 6 低色域模糊场景下的 HDR 图像重建
(左: LDR, 右: 重建 HDR)

 图 7 一般场景下的 HDR 图像重建
(左: LDR, 右: 重建 HDR)

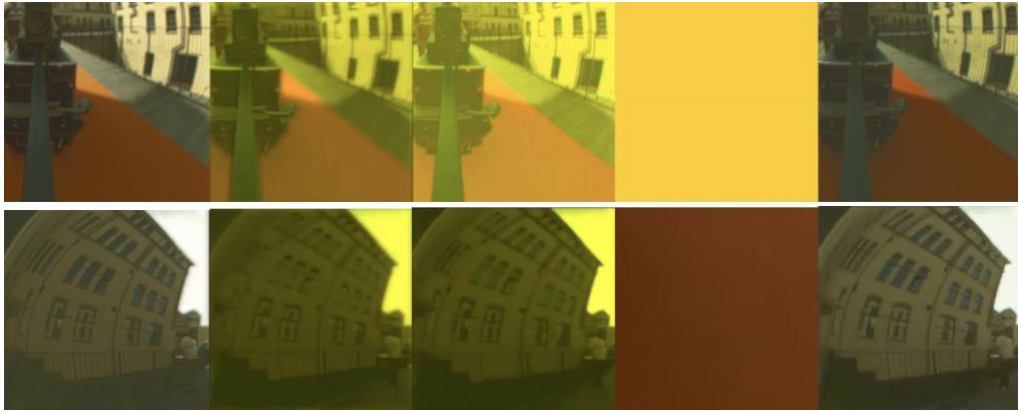
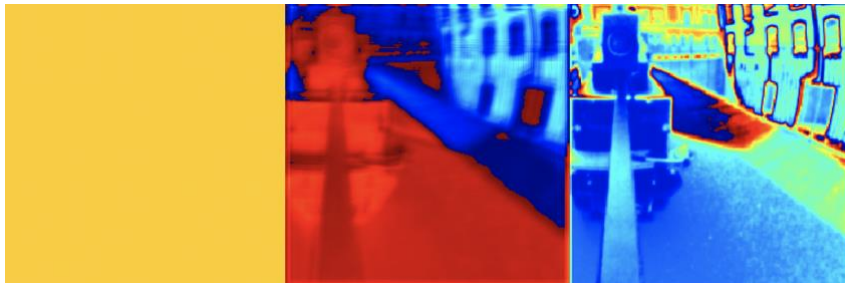
 Fig.6 HDR image reconstruction in low-color-gamut blurred scene。 Fig.7 HDR image reconstruction in general scenes
(Left: LDR, Right: Reconstructed HDR)

 图 8 消融实验：保留网络中不同分支所得到的主观图像质量
(左 1: 去掉中频分支, 左 2: 仅中频分支, 左 3: 仅高频分支, 左 4: 仅低频分支, 左 5: 所提出方法:)
 Fig.8 Subjective image quality by retaining different branches of network in ablation experiment
(Left 1: Remove the middle-frequency branch, Left 2: Only middle-frequency branch, Left 3: Only high-frequency branch, Left 4: Only low-frequency branch, Left 5: Proposed method)

图 9 不同分支输出的特征图 (左: 低频分支, 中: 中频分支, 右: 高频分支)

 Fig.9 Feature maps output from different branches
(left: low-frequency branch, middle: middle-frequency branch, right: high-frequency branch)

参考文献 (References)

- [1] 马正先. HDR 技术及其在 4K 超高清电视上的应用[J]. 电视技术, 2019, 43(1):33-39.
MA Z X. HDR technology and application on 4K ultra-high-definition TV[J]. Television Technology, 2019, 43(1):33-39(in Chinese).
- [2] ENDO Y, KANAMORI Y, MITANI J. Deep reverse tone mapping[J]. ACM Transactions on Graphics, 2017, 36(6):177:1-177:10.
- [3] EILERTSEN G, KRONANDER J, DENES G, et al. HDR image reconstruction from a single exposure using deep CNNs[J]. ACM Transactions on Graphics, 2017, 36(6):1-15.
- [4] XU Y C, SONG L, XIE R, et al. Deep video inverse tone mapping[C]//2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM). Piscataway: IEEE Press, 2019:142-147.
- [5] MARNERIDES D, BASHFORD - ROGERS T, HATCHETT J, et al. ExpandNet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content[C]//Computer Graphics Forum, 2018, 37(2):37-49.
- [6] KINOSHITA Y, KIYA H. iTM-Net: Deep inverse tone mapping using novel loss function considering tone mapping operator[J]. IEEE Access, 2019, 7:73555-73563.
- [7] LEE S, AN G H, KANG S J. Deep chain HDRI: Reconstructing a high dynamic range image from a single low dynamic range image[J]. IEEE Access, 2018, 6:49913-49924.
- [8] XU Y C, NING S Y, XIE R, et al. Gan based multi-exposure inverse tone mapping[C]//2019 IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE Press, 2019:1-5.
- [9] NING S Y, XU H T, SONG L, et al. Learning an inverse tone mapping network with a generative adversarial regularizer[C]//2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2018:1383-1387.
- [10] LEE S, AN G H, KANG S J. Deep recursive HDRI: Inverse tone mapping using generative adversarial networks[C]//Proceedings of the European Conference on Computer Vision (ECCV). Berlin: Springer, 2018:596-611.
- [11] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Berlin: Springer, 2015:234-241.
- [12] TAKEUCHI M, SAKAMOTO Y, YOKOYAMA R, et al. A Gamut-extension method considering color information restoration using convolutional neural networks[C]//2019 IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE Press, 2019:774-778.
- [13] LEDIG C, THEIS L, HUSZÁR F, et al. Photo-realistic single image super-resolution using a generative adversarial network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017:4681-4690.
- [14] WANG X T, KE Y, WU S X, et al. EsrGAN: Enhanced super-resolution generative adversarial networks[C]//Proceedings of the European Conference on Computer Vision (ECCV). Berlin: Springer, 2018:0-0.
- [15] ISOLA P, ZHU J Y, ZHOU T, et al. Image-to-image translation with conditional adversarial networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017:1125-1134.
- [16] 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. Piscataway: IEEE Press, 2017:2223-2232.
- [17] SIAROHIN A, SANGINETO E, LATHUILIÈRE S, et al. Deformable GANs for pose-based human image generation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018:3408-3416.
- [18] CHAN C, GINOSAR S, ZHOU T H, et al. Everybody dance now[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2019:5933-5942.
- [19] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Advances in Neural Information Processing Systems, 2014:2672-2680.
- [20] RATLIFF L J, BURDEN S A, SASTRY S S. Characterization and computation of local Nash equilibria in continuous games[C]//2013 51st Annual Allerton Conference on Communication, Control, and Computing. Piscataway: IEEE Press, 2013:917-924.

HDR image generation method based on conditional generative adversarial

network

BEI Yue¹, Wang Qi¹, CHENG Zhipeng¹, PANG Xinghao¹, YANG Mohan², DING Dandan^{2*}

(1. MIGU Video Co., Ltd., Shanghai 201201, China;

2. Beijing Bravo Video Technologies Incorporation, Beijing 100036, China)

Abstract: Compared with Low Dynamic Range (LDR) images, High Dynamic Range (HDR) images have a wider color gamut and higher brightness range, which is more in line with human visual effects. However, since most of the current image acquisition devices are LDR devices, HDR image resources are scarce. An effective way to solve this problem is to map LDR images to HDR images through inverse tone mapping. This paper proposes an inverse tone mapping algorithm based on Conditional Generative Adversarial Network (CGAN) to reconstruct HDR images. To this end, a multi-branch-based generation network and a discrimination network based on discrimination blocks are designed, and the data generation and feature extraction capabilities of CGAN are used to map a single LDR image from the BT.709 color gamut to the corresponding BT.2020 color area. The experimental results show that the proposed network can obtain higher objective and subjective quality compared with the existing methods. Especially for fuzzy areas in the low color gamut, the proposed method can reconstruct clearer textures and details.

Keywords: conditional generative adversarial network; convolutional neural network; inverse tone mapping; gamut mapping; feature extraction

Received:2020-09-14; **Accepted:**2021-04-23; **Published online:**2021-05-17

URL: <https://kns.cnki.net/kcms/detail/11.2625.v.20210517.1523.003.html>

Foundation item: Zhejiang Provincial Natural Science Foundation of China (LY20F010013)

***Corresponding author.** E-mail: DandanDing@hznu.edu.cn