

# HDR Image Reconstruction from Single LDR Images with Two-Stage Method

## Abstract

本文提出了一种基于双阶段神经网络的方法，通过单张低动态范围（LDR）图像生成不同曝光的LDR图像并融合它们来重建高动态范围（HDR）图像。在第一阶段，我们使用卷积神经网络（CNN）进行监督学习，生成多张不同曝光的LDR图像，这些图像模拟了实际拍摄中的多次曝光效果，为HDR重建提供了丰富的图像信息。在第二阶段，我们采用深度耦合反馈网络（CF-Net），结合多曝光融合（MEF）和超分辨率（SR）技术，将一对极端曝光的低分辨率LDR图像转化为高动态范围和高分辨率的图像。实验结果表明，该方法在视觉质量和细节保留方面具有一定效果，为从单一LDR图像重建HDR图像提供了可能有效的解决方案。

## 1. Introduction

高动态范围（HDR）图像能够捕捉场景中亮度范围极大的细节，提供比传统低动态范围（LDR）图像更丰富的视觉体验。然而，传统的HDR图像生成方法通常依赖于多张不同曝光的LDR图像，这在实际拍摄过程中存在一定的困难和局限性。例如，拍摄多张曝光图像需要固定的相机位置和静止的场景，否则会导致图像出现鬼影等问题。此外，多次曝光拍摄也增加了拍摄时间，不适用于快速移动的场景<sup>1</sup>。因此，如何从单一的LDR图像重建出高质量的HDR图像成为一个具有挑战性的问题。

近年来，随着深度学习技术的迅猛发展，基于神经网络的方法在图像处理领域取得了显著成果。许多研究工作已经探索了利用深度学习从单一LDR图像重建HDR图像的方法。例如，有些方法利用生成对抗网络<sup>2</sup>（GAN）或卷积神经网络（CNN）来预测HDR图像。然而，这些方法通常忽略了生成的HDR图像在高分辨率和高动态范围两个方面的需求，导致结果在细节和清晰度上存在不足。

针对上述问题，我们提出了一种基于双阶段神经网络的方法，通过单张LDR图像生成不同曝光的LDR图像<sup>3</sup>并融合它们来重建HDR图像<sup>4</sup>。在第一阶段，我们采用监督学习的方法，训练神经网络合成多张不同曝光的LDR图像。这些合成的LDR图像模拟了真实拍摄时的多次曝光效果，为后续的HDR重建提供了丰富的图像信息。在第二阶段，我们使用了一个深度耦合反馈网络（Coupled Feedback Network, CF-Net），同时实现多曝光融合（MEF）和超分辨率（SR）。具体来说，CF-Net能够输入一对极端过曝和欠曝的低分辨率LDR图像，并生成一张具有高动态范围和高分辨率的图像。通过这种方式，不仅提高了HDR图像的动态范围，还增强了图像的清晰度和细节表现。

总之，我们提出的两阶段方法有效地解决了单一LDR图像重建HDR图像的挑战，通过合成不同曝光的图像并进行多曝光融合和超分辨率处理，生成了高质量的HDR图像。实验结果表明，我们的方法在视觉质量和细节保留方面在HDR重建领域具有一定优势，为实际应用提供了更加灵活和高效的解决方案。

## 2. Related Work

在这一部分我们将回顾两个方面的相关工作，“从单个LDR图像生成一系列不同曝光的LDR图像”和“从多曝光LDR图像合成高质量的HDR图像”。

## 2.1 Deep Reverse Tone Mapping

一个很重要的工作是从单个LDR图像生成HDR图像以克服这些限制。其中值得注意的方法是由Endo Y等人(2017)提出的基于深度学习的框架Deep Reverse Tone Mapping(DrTMO)<sup>3</sup>。该方法使用卷积神经网络(CNN)从单个LDR输入推断出一系列不同曝光的LDR图像，然后将这些合成图像合并以重建HDR图像。关键的创新是使用3D反卷积网络来学习由于不同曝光而导致的像素值的相对变化，使模型能够比以前的方法更有效地再现饱和像素的自然色调和颜色。

## 2.2 Deep Coupled Feedback Network

该领域的另一个重要贡献是Deng X等人(2021)的Deep Coupled Feedback Network(CF-Net)<sup>4</sup>。该网络将多曝光融合[^5-9] (MEF) 和超分辨率[^10-15] (SR) 任务集成到一个统一的深度学习框架中。CF-Net采用耦合递归子网络，通过特征提取、超分辨率和反馈块处理过度曝光和曝光不足的LDR图像。耦合反馈机制增强了图像融合和分辨率，从而获得高质量的HDR图像，改善了细节和动态范围。

虽然这些方法使该领域取得了重大进展，但仍存在一些挑战。例如，在严重曝光不足或过度曝光的区域，准确推断缺失的信息仍然很困难。此外，在HDR重建过程中保持颜色一致性和避免伪影是一个持续的问题。未来的研究应该专注于通过提高HDR重建算法的鲁棒性和准确性来解决这些挑战，可能通过更复杂的网络架构或先进的训练技术来更好地处理各种现实世界的照明条件。

## 3. Algorithm pipeline

在本节中，我们详细介绍我们提出的方法，该方法结合了Deep Reverse Tone Mapping (DrTMO)和Deep Coupled Feedback Network (CF-Net)两篇论文的技术，以从单一LDR图像重建高质量的HDR图像。

### 3.1 多曝光LDR图像合成

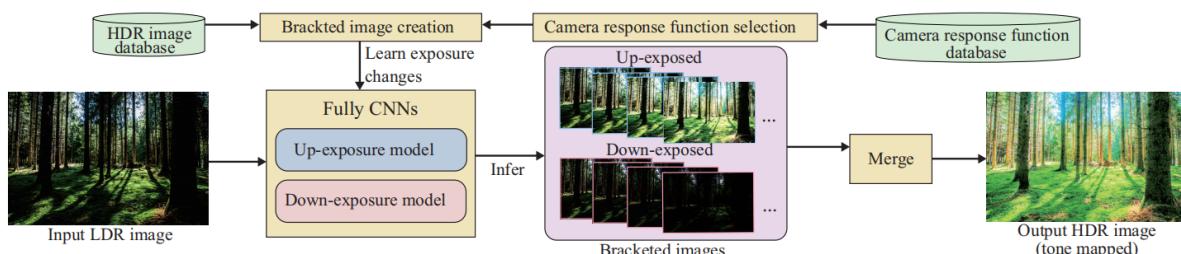


Fig.1. Overview of the proposed method. The flow is decomposed into the learning and inference phases. In the learning phase, the bracketed LDR images are first created from HDR databases by simulating cameras. Next, we let our fully CNNs learn the changes in the exposures of the bracketed images. In the inference phase, the learned CNNs compute LDR images with different exposures from a single input LDR image. The final HDR image is then generated from these bracketed LDR images.

Figure 1 说明了我们的方法在多曝光LDR图像合成阶段的整体流程，通过推断Bracketed LDR images并合并它们，间接地从单个LDR输入重建HDR图像。流程被分解为学习和推理阶段。在学习阶段，首先通过模拟相机从HDR数据库创建Bracketed LDR images (第3.1.1节)。接下来，让神经网络模型 (up-/down-exposure networks) 学习Bracketed images 曝光的变化 (第4节)。在推理阶段，学习的模型计算来自单个输入LDR图像的不同曝光的LDR图像。在up/down-exposure networks中，分别推断出Brighter/dimmer bracketed images。然后从这些Bracketed LDR images生成最终的HDR图像。

### 3.1.1 创建Bracketed Images用于训练

本阶段使用的训练数据集由真实的HDR图像和相应的Bracketed LDR images图像集组成。为了解释不同的非线性相机响应函数(CRF)引起的LDR图像的颜色变化，合成了一组具有不同CRF和HDR图像的每个曝光值的LDR图像。为此，使用以下方程来模拟相机

[Debevec and Malik 1997]:

$$Z_{i,j} = f(E_i \Delta t_j) \quad (1)$$

式中 $Z_{i,j}$ 表示每个像素*i*的像素值和曝光时间指数*j*： $f$ 、 $E_i$ 和 $\Delta t_j$ 分别表示CRF、辐照度、和曝光时间。在本文中， $Z_{i,j}$ 和 $E_i$ 分别代表LDR和HDR图像。

为了定义CRF，使用了Grossberg and Nayar的响应函数数据库(DoRF)[Grossberg and Nayar 2003]。这个数据库由作者收集的201种常见品牌的胶片、电荷耦合器件(CCDs)和数码相机的响应曲线组成。数据库中的所有CRF都是单调的，在[0, 1]范围内归一化，并以1000点采样。所有CRF都显示在Figure 2的左图。然而，使用所有的CRF是多余的，并且不必要的增加了训练时间，因此只使用K-means聚类选择的具有代表性的CRF (Figure 2, 右)。在本次实验中使用5个CRF用样条插值。

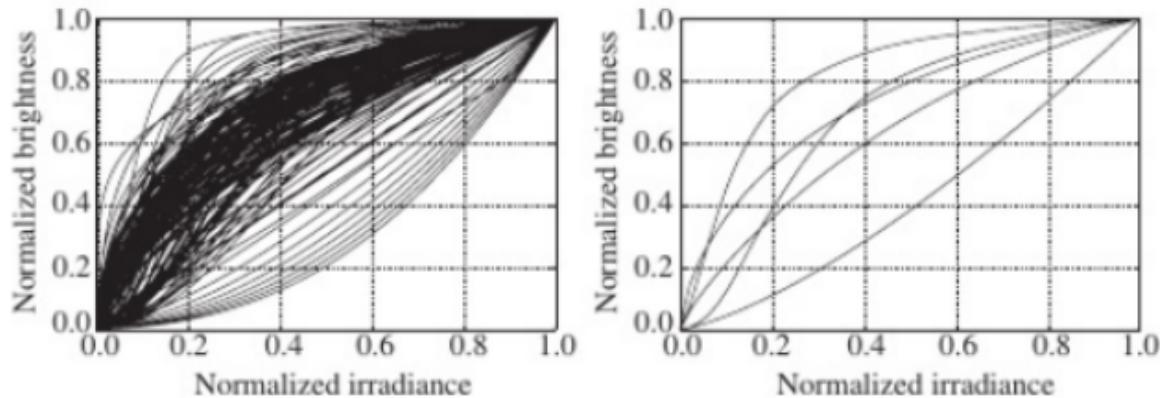


Fig.2. Camera response curves for creating training data. Using all response curves in the database [Grossberg and Nayar 2003] is redundant(left), so we choose five representative curves using k-means clustering(right).

因为DoRF中的CRF是归一化的，我们必须确定 $E_i$ 和 $\Delta t_j$ 的绝对标准。因此适当调整观测信号 $E_i \Delta t_j$ 的范围，具体的，将 $\Delta t_j$ 设置为 $\tau$

$$\Delta t_j = \frac{1}{\tau^{T/2}}, \dots, \frac{1}{\tau^2}, \frac{1}{\tau}, 1, \tau, \tau^2, \dots, \tau^{T/2} \quad (2)$$

其中T为偶数， $j = 1, 2, \dots, T + 1$ 。然后将 $E_i \Delta t_j$ 归一化，使

$E_i \Delta t_{T/2+1}$ (=  $E_i$  because  $\Delta t_{T/2+1} = 1$ )的平均像素值等于0.5，在原作者的实验中，使用 $T = 8, \tau = \sqrt{2}$ 。Figure 3为得到的不同曝光的LDR图像，虽然 $E_i$ 的归一化和 $\Delta t_j$ 的选择决定了推断的HDR图像的动态范围，但 $E_i$ 或 $\Delta t_j$ 的线性缩放可以通过推断HDR值的线性缩放来补偿。因此，如果推断的HDR图像有点太暗或太亮，用户可以通过线性缩放像素值来调整它。

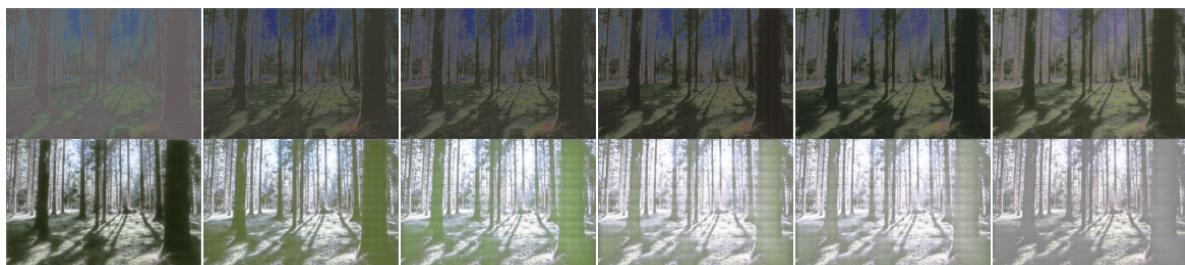


Fig.3. Examples of bracketed images created from the HDR datasets. Exposures are selected to avoid completely white or black regions.

## 3.2 HDR图像重建

在第二阶段，我们使用CF-Net将合成的多张LDR图像融合，生成最终的HDR图像。首先是将输入的多张不同曝光的LDR图像通过卷积神经网络进行特征提取处理，提取出各个图像的关键特征 $\{F_1, F_2, \dots, F_n\}$ ，然后将图像特征通过CF-Net进行处理。该子网络包括多个循环单元（第4节），通过反馈机制逐步优化特征表示，每个单元在特定的递归次数内对特征进行更新和优化，提高图像的细节和分辨率。

图像通过超分辨率模块进行分辨率提升。该模块使用卷积层和反卷积层，将低分辨率的图像特征映射到高分辨率空间。最后，通过多曝光融合模块将处理后的图像特征进行融合，生成最终的HDR图像。这个融合过程充分利用了不同曝光图像的优点，保留了丰富的细节和动态范围。

## 4. Model and Methods

在本节中，我们描述了我们使用的神经网络架构包括UP/DOWN-Exposure Models、GAN、CF-Net以及相应设计的损失函数

### 4.1 UP/DOWN-Exposure Models

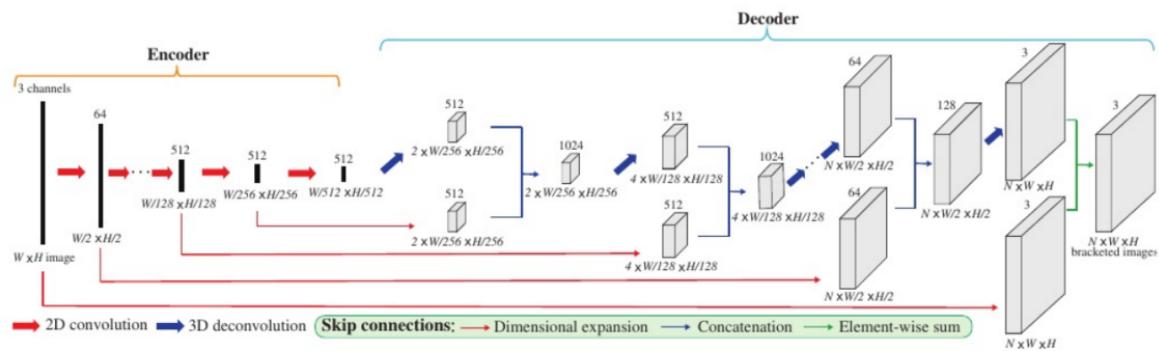


Fig.4. Original network architecture. The same architecture is used both for up- and down-exposure models.

Figure 4显示了原论文<sup>3</sup>中该部分网络的架构，网络的输入是一张 $W * H * C$ 的LDR图像，其中W、H和C代表宽度、高度和通道数。基本使用 $512 * 512 * 3$ 的RGB图像作为输入来训练网络，在推理阶段可以接受更大的图像作为输入。

由于资源限制，无法完全复现该网络架构，所以我们对该网络架构稍作调整，Figure 5是调整后的网络架构。

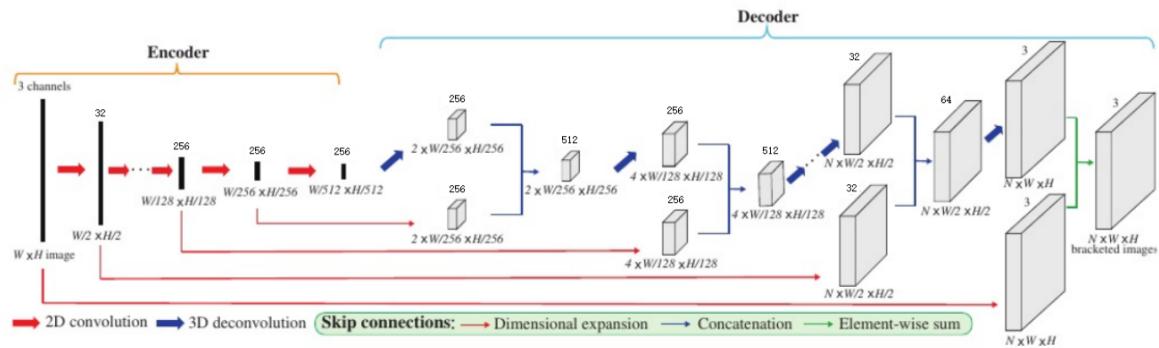


Fig.5. Our new network architecture. The same architecture is used both for up- and down-exposure models.

输入图像的像素值在[0, 1]范围内进行归一化处理。该网络首先使用二维卷积神经网络将输入LDR图像编码为潜在语义特征(latent semantic features)，结果是第1个卷积层上 $\frac{W}{2^l} * \frac{H}{2^l} * c_l$ 的三维张量。在编码器的最后一层，输入图像被缩减到一个像素，得到一个平展的256维潜在向量(latent vector)。接下来，该网络使用3D反卷积神经网络将语义特征解码为不同曝光时间下拍摄的LDR图像，输出是一个 $N * W * H * c$ 的四维张量，其中N表示曝光持续时间的数量。如3.1节所述，UP/DOWN-Exposure Models通过学习数据库中bracketed images的曝光变化，从 $\tau$ （或 $\frac{1}{\tau}$ ）依次增加/减少输入图片的曝光。

下面将详细的描述CNN架构。

#### 4.1.1 Encoder

对于编码器，采用了类似 Isola et al. [2016]的pix2pix模型的架构，除了输入图像的大小和层数。具体而言，编码器由步幅(2, 2)、 spatial padding (1, 1)的九层 $4 * 4$ 卷积组成，从第一层到最后一层，卷积核的数量（即输出通道的数量）分别为32、64、128、256、256、256、256、256、256。在第二层及后续层，对卷积输出进行批量归一化，通过对每个输入特征进行批量归一化来改善在网络中的学习。在本网络中，批归一化通过仅使用其在学习和推理阶段的统计数据来归一化一个input batch。每一层的激活函数为leaky ReLU函数。

#### 4.1.2 Decoder

在2D CNN之后，解码器使用卷积特征作为输入，生成一个 $N * W * H * c$ 的四维张量，由N张不同曝光的图像组成。为了生成具有不同曝光的一致图像，采用了3D反卷积神经网络，三维CNN是二维CNN的扩展，通过在时间和空间上进行卷积来获得temporal-coherent videos。解码器由9层组成，前3层是 $4 * 4 * 4$ 反卷积（按照曝光、宽度、高度顺序），步长为(2, 2, 2)，padding 为(1, 1, 1)，其余图层为 $3 * 4 * 4$ 反卷积，步长为(1, 2, 2)，padding为(1, 1, 1)，也就是说，前3层的曝光轴和空间轴上将图层输入加倍，其余图层仅在空间轴上加倍。从第一层到最后一层，卷积核的数量分别是256、256、256、256、256、128、64、32、3。在除最后一层以外的层中，应用批处理归一化，激活函数为ReLU函数。最后一层通过sigmoid函数输出一个像素值在[0, 1]之间的四维张量。

#### 4.1.3 Skip Connections

为解决3D CNN产生的跨时间轴的不一致噪声，实验采用了跳跃连接来扩展网络。在上面的Encoder-Decoder网络中，解码器使用完全编码的向量，代表整个图像的潜在特征。为了逐步将输入图像中的本地和低级信息合并到解码器中，在U-Net<sup>5</sup>扩展后添加跳跃连接和残差单元<sup>6</sup>。U-Net在第*i*层和第*n-i*层之间有跳跃式连接，其中*n*为总层数，它连接了两个层之间的所有channels。这种架构使解码器能够利用局部信息并加速学习。残差单元可以用一般形式表示为 $x_{l+1} = f(h(x_l) + F(x_l))$ ，其中 $x_l$ 和 $x_{l+1}$ 表示第*l*个单元的输入和输出， $f$ 表示激活函数， $F$ 是残差函数。直观地说，残差单元可以从输入中学习变化，我们假设从输入中学习曝光变化比学习如何从头开始生成新图像更容易。

U-Net和ResNet是为2D卷积和反卷积设计的，由于具有跳跃连接的层上输出维度不同，因此不能直接应用于2D卷积核3D反卷积，所以，在编码器中扩展输入图像和中间特征的维度。具体而言，我们复制并连接输入图像和编码特征，以便每个张量的维度与解码器相应层的维度相匹配。将 $W * H * c$ 输入图像张量和 $\frac{W}{2^l} * \frac{H}{2^l} * c_l$ 编码特征转换为 $N * W * H * c$ 和 $N * \frac{W}{2^l} * \frac{H}{2^l} * c_l$ 。对于U-Net，将 $N * \frac{W}{2^l} * \frac{H}{2^l} * c_l$ 编码特征和连接的反卷积层的输出连接起来。对于残差单元，将 $N * W * H * c$ 输入图像张量添加到激活函数之前的最后一层，在残差单元 $x_{l+1} = f(h(x_l) + F(x_l))$ 中 $x_l$ 是一个输入图像， $x_{l+1}$ 是一个输出，函数 $f$ 是sigmoid函数， $h$ 是恒等映射， $F$ 是没有最后一个sigmoid函数的整个网络的函数。

#### 4.1.4 损失函数

如3.1节所述，为每个场景和每个CRF合成一组 $T + 1$ 张Bracketed Images。设D为T+1张Bracketed images的集合， $I_j \in D$ 是曝光比指数为j的LDR图像。给定 $I_j$ ，UP-Exposure Model通过参考D中其余曝光较高的图像即 $I_{j+1}, I_{j+2}, \dots, I_{j+1+N}$ 作为真相，来学习曝光增加后的相对变化。UP-Exposure Model的损失函数定义为：

$$\Sigma_D \Sigma_{j=1}^T ||I_{j+1 \rightarrow j+1+N}^{up} \oplus O_j - M_j \circ G(I_j, \theta)||_1 \quad (3)$$

其中 $I_{j+1 \rightarrow j+1+N}^{up}$ 表示将 $I_{j+1}$ 的图像连接到 $I_{\min\{j+1+N, T+1\}}$ 得到的一个 $\min\{N, T+1-j\} * W * H * c$ 张量。 $O_j$ 和 $\oplus$ 是一个 $\min\{j+N-T-1, 0\} * W * H * c$ 的零张量和一个拼接算子。 $M_j$ 和 $\circ$ 表示一个 $N * W * H * c$ 的张量和一个元素积。为了掩盖数据不存在的区域，如果第一维的索引小于 $T+2-j$ 则 $M_j$ 的每个元素为1，否则为0。 $G(I_j, \theta)$ 是网络的输出 $C * W * H * c$ 张量， $\theta$ 表示网络权值。对于DOWN-Exposure Model，网络从相同的训练数据反向学习图像，其损失函数定义为：

$$\Sigma_D \Sigma_{j=1}^T ||I_{T+1-j \rightarrow T+1-j-N}^{down} \oplus O_j - M_j \circ G(I_{T+2-j}, \theta)||_1 \quad (4)$$

其中 $I_{T+1-j \rightarrow T+1-j-N}^{down}$ 表示一个通过将 $I_{T+1-j}$ 反向连接到 $I_{\max\{1, T+1-j-N\}}$ 得到的 $\min\{N, T+1-j\} * W * H * c$ 张量。

使用Adam优化器以固定学习率0.0002和momentum为0.5的batch大小为1的随机梯度下降来训练网络。用标准差为0.02的零均值高斯噪声初始化2D卷积和3D反卷积的所有权重。在批量归一化后将50%的dropout率应用于解码器的前三个反卷积层，以便使解码器对编码特征中的噪声具有鲁棒性。

## 4.2 CF-Net

本节将介绍CF-Net。首先在4.3.1节中介绍整个网络架构，然后在4.3.2节中分析在网络中使用的耦合反馈块(CFB)的新架构，最后在4.3.3节中介绍损失函数和训练策略。

### 4.2.1 Network Architecture

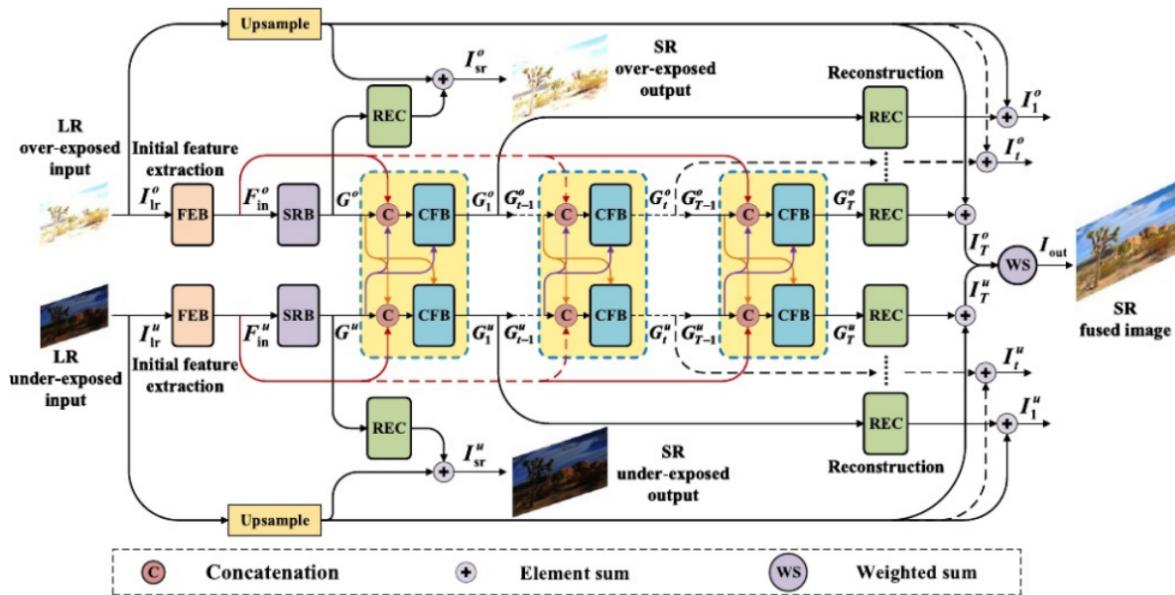


Fig.6. Network architecture of the proposed CF-Net. The overall network is composed of two sub-nets with LR over-exposed and under-exposed images as inputs, respectively. Each sub-net is composed of an initial feature extraction block (FEB), a super-resolution block (SRB) and several coupled feedback blocks (CFB). The two sub-nets interact and communicate with each other through the CFB, to boost the super-resolution and exposure fusion performance simultaneously.

Figure 6是提出的CF-Net整体网络架构，可以看到，CF-Net是由两个耦合的递归子网络组成，过度曝光或曝光不足的LR图像作为输入，每个子网络包含一个初始特征提取块(FEB)、一个超分辨率块(SRB)和几个耦合反馈块(CFB)。具体来说，FEB设计用于从LR输入中提取基本特征，以支持后续的SRB和CFB。将LR过曝光和欠曝光的图像输入分别记为 $I_{lr}^o$ 和 $I_{lr}^u$ ，则FEB提取的对应特征 $F_{in}^o$ 和 $F_{in}^u$ 可通过以下方法得到：

$$F_{in}^o = f_{FEB}(I_{lr}^o) \quad (5)$$

$$F_{in}^u = f_{FEB}(I_{lr}^u) \quad (6)$$

其中 $f_{FEB}$ 表示特征提取块的操作。之后，FEB由PReLU激活的两个卷积层组成：第一层有256个大小为3的卷积核\*3，第二层有64个大小为1的卷积核\*1。第一层用于提取基本的LR特征，第二层用于进一步整合跨通道特征，使其更加紧凑。提取的特征 $F_{in}^o$ 和 $F_{in}^u$ 作为后续SRB和CFB的基本输入。

以基本特征 $F_{in}^o$ 和 $F_{in}^u$ 为输入，SRB的作用是学习更多的高级特征，提高图像分辨率。采用<sup>7</sup>中的反馈块架构作为我们的SRB架构。具体来说，它包含几个projection groups，其中有密集的跳跃连接。每个projection group包括一个上采样和一个下采样操作，SRB学到的高级特征可以表示为：

$$G^o = f_{SRB}(F_{in}^o) \quad (7)$$

$$G^u = f_{SRB}(F_{in}^u) \quad (8)$$

其中 $f_{SRB}$ 表示SRB操作， $G^o$ 和 $G^u$ 分别是对曝光过度和曝光不足的图像提取的高级特征。为了重建超分辨率图像，使用重建块(REC)将 $G^o$ 和 $G^u$ 映射到高分辨率图像，重建块由一个反卷积层和一个卷积层组成。考虑到跳跃连接，将双线性上采样图像与重构残差图像相加，即可得到SRB后的最终超分辨率图像：

$$I_{sr}^o = f_{UP}(I_{lr}^o) + f_{REC}(G^o) \quad (9)$$

$$I_{sr}^u = f_{UP}(I_{lr}^u) + f_{REC}(G^u) \quad (10)$$

其中 $f_{UP}$ 为双线性提高(bilinear upscaling)， $f_{REC}$ 是重构操作。REC块之间不共享参数。注意， $I_{sr}^o$ 和 $I_{sr}^u$ 只是 $I_{lr}^o$ 和 $I_{lr}^u$ 的超分辨率版本，并没有任何融合信息。这两个图像的主要作用是保证高级特征 $G^o$ 和 $G^u$ 的有效性，是后续CFB的重要输入。

耦合反馈块(CFB)是CF-Net的核心组件，旨在通过复杂的网络连接同时实现超分辨率和图像融合。与前面提到的FEB和SRB各只涉及一个块不同，我们使用了一系列相互连接的CFB，如Figure 6所示。

$(t-1)-th(t > 1)$ CFB的输出是第 $t-th$ CFB的输入。除了这个输入端，第 $t$ 个CFB还有另外两个输入端。具体来说，在以 $I_{lr}^o$ 为原始输入的上层子网络中，其第 $t$ 个CFB的输出可表示如下：

$$\mathbf{G}_t^o = f_{CFB}(\mathbf{F}_{in}^o, \mathbf{G}_{t-1}^o, \mathbf{G}_{t-1}^u), \quad (11)$$

其中， $F_{in}^o$ 是FEB提取的基本特征， $\mathbf{G}_{t-1}^o$ 是其前 $(t-1)$ 个CFB提取的高级特征， $\mathbf{G}_{t-1}^u$ 是下层子网络中 $(t-1)$ 个CFB提取的高级特征。在 $t=1$ 的情况下， $\mathbf{G}_{t-1}^o$ 和 $\mathbf{G}_{t-1}^u$ 分别变为 $G^o$ 和 $G^u$ ，这意味着我们接受SRB的输出作为第一个CFB的输入。在公式(11)中的三个输入中，前两个输入对超分辨率性能的贡献更大，而最后一个输入则用于改善融合结果。这三个输入并不是简单地串联起来输入CFB，而是有一个精心设计的连接，将在第4.3.2节部分详细介绍。与公式(11)类似，对于以 $I_{lr}^u$ 为输入的下层子网络，第 $t$ 个CFB的输出 $\mathbf{G}_t^u$ 如下：

$$\mathbf{G}_t^u = f_{CFB}(\mathbf{F}_{in}^u, \mathbf{G}_{t-1}^u, \mathbf{G}_{t-1}^o). \quad (12)$$

在每个CFB之后，我们可以通过以下方式重建融合的SR图像：

$$\mathbf{I}_t^o = f_{UP}(\mathbf{I}_{lr}^o) + f_{REC}(\mathbf{G}_t^o), \quad (13)$$

$$\mathbf{I}_t^u = f_{UP}(\mathbf{I}_{lr}^u) + f_{REC}(\mathbf{G}_t^u). \quad (14)$$

请注意，这里的  $I_t^o$  和  $I_t^u$  都是高动态范围的超分辨率图像。由于  $G_t^o$  和  $G_t^u$  是通过整合曝光过度和曝光不足的特征生成的，因此  $I_t^o$  和  $I_t^u$  都具有较高的动态范围。上标 o 和 u 只是用来表示它们是由哪个子网络生成的。假设每个子网络中的 CFB 数量为 T，我们可以生成 2T 张高动态范围的超分辨率图像，即  $\{I_t^o\}_{t=1}^T$  和  $\{I_t^u\}_{t=1}^T$ 。实验结果表明，第 t 个 CFB 的性能优于之前的第 (t-1) 个 CFB；因此，我们只使用最后一个 CFB 的结果，即  $I_T^o$  和  $I_T^u$ ，来生成最终的重建图像。下面的公式显示了我们如何获得最终的重建图像  $I_{out}$ ：

$$I_{out} = w_o I_T^o + w_u I_T^u. \quad (15)$$

这里， $w_o$  和  $w_u$  是加权参数。本文将  $w_o$  和  $w_u$  都设为 0.5。虽然  $\{I_t^o\}_{t=1}^{T-1}$  和  $\{I_t^u\}_{t=1}^{T-1}$  并不直接用于获取最终图像，但它们在提高最终图像的性能方面发挥着重要作用。从公式 (13) 和 (14) 中可以看出，只有准确重建  $I_t^o$  和  $I_t^u$ ，才能得到有效的高级特征  $G_t^o$  和  $G_t^u$ 。只有利用这些有效的高级特征，才能高精度地重建最终图像。

#### 4.2.2 Coupled Feedback Block (CFB)

耦合反馈块 (CFB) 是 CF 网络的基本核心组件。许多研究都证实，反馈机制有助于图像复原。在本文中，提出了一种耦合反馈机制，并证明它能为图像超分辨率和图像融合任务带来巨大好处。Figure 7 显示了所提出的 CFB 的详细架构。虽然网络中存在一系列 CFB，但每个 CFB 的结构都是相同的。在此，我们以上层子网络中的第 t 个 CFB 为例，介绍其内部结构及其与其他块的交互。

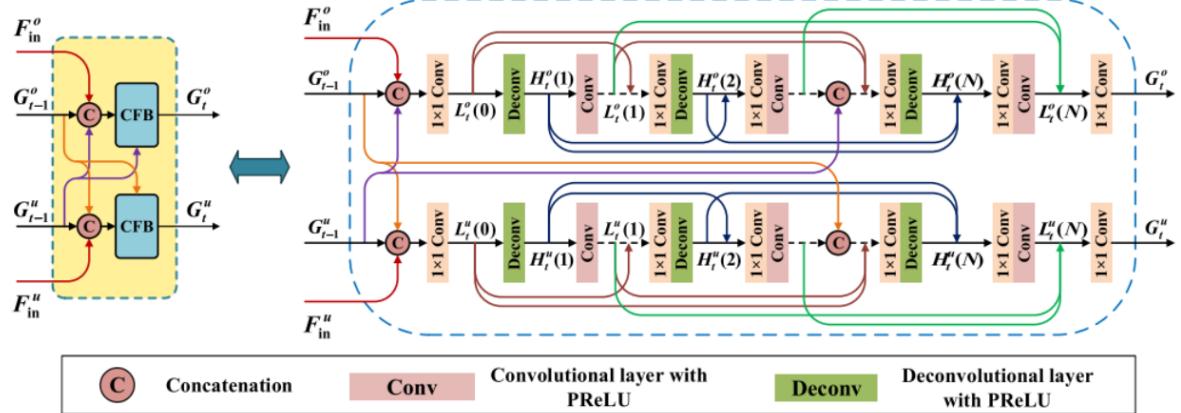


Fig.7. The symmetric architecture of the t-th CFB in the upper and lower sub-networks. The upper CFB accepts  $F_{in}^o, G_{t-1}^o$  as inputs, and output  $G_t^o$ , while the lower CFB accepts  $F_{in}^u, G_{t-1}^u$  as inputs, and output  $G_t^u$ .

#### 4.2.3 损失函数

##### 4.2.3.1 平均结构相似度(Mean Structural Similarity Index Metric, MSSIM)

本文采用MSSIM损失对网络进行端到端训练，用于测量失真图像与参考图像斑块之间的差异，与均方误差(MSE)相比，它能更好地描述图像的感知质量。给定一个畸变图像patch  $x$ 和一个参考图像patch  $y$ ，SSIM定义如下公式：

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (16)$$

其中  $\mu_x$  和  $\mu_y$  为 patch 的平均值， $\sigma_x$  和  $\sigma_y$  表示 patch 的标准差， $\sigma_{xy}$  是协方差值， $C_1$  和  $C_2$  是用于避免数值不稳定的常量正值。

为测量图像级SSIM值，通过平均所有patch上的SSIM值来计算MSSIM：

$$MSSIM(X, Y) = \frac{1}{J} \sum_{j=1}^J SSIM(x_j, y_j) \quad (17)$$

$X$ 和 $Y$ 分别表示畸变图像和参考图像,  $x_j$ 和 $y_j$ 分别表示它们在第j个位置的patch。MSSIM取值范围为0~1, 取值越大表示图像质量越好, 在此基础上, 将图像 $X$ 和 $Y$ 之间的MSSIM损耗定义为:

$$L_{MS}(X, Y) = 1 - MSSIM(X, Y) \quad (18)$$

#### 4.2.3.2 损失函数

CF-Net总损失函数定义:

$$\begin{aligned} L_{total} = & \lambda_o L_{MS}(I_{st}^o, I_{gt}^o) + \lambda_u L_{MS}(I_{sr}^u, I_{gt}^u) \\ & + \sum_{t=1}^T \lambda_t (L_{MS}(I_t^o, I_{gt}) + L_{MS}(I_t^u, T_{gt})) \end{aligned} \quad (19)$$

其中 $I_{gt}^o$ 和 $I_{gt}^u$ 分别为过曝光和欠曝光的HR真相,  $I_{gt}$ 是高动态范围HR真相, 这是我们的最终目标。 $\lambda_o$ 、 $\lambda_u$ 和 $\{\lambda_t\}_{t=1}^T$ 是每个损失的权重。(19)中的损失函数可以分为两部分, 前两个损失用于保证SRB的有效性, 而最后一个损失用于确保每个CFB都能正常工作。也就是说, 形成前两个损失是为了保证超分辨率性能, 构造最后一个损失是为了同时保证超分辨率和曝光融合性能。前两个损失也作为最后一个损失的重要基础。通过最小化(19)中定义的损失, 以端到端的方式训练整个网络, 对于训练好的模型, 在测试阶段用(15)得到最终的输出图像。

## 5. Experiments

在本节中, 我们测试了所提出的两阶段方法的性能。第5.1节介绍了使用到的数据集以及实验设置, 第5.2节和5.3节分别展示了与其他先进方法的定量和定性比较结果。

### 5.1 Dataset

#### 5.1.1 第一阶段训练集

CNN:在论文中, 原作者收集了网上的在线数据库 (包含1043张HDR图像) 用于生成五种色调曲线和九种不同曝光时间的LDR图像, 所有的训练图像的大小都调整为512x512, 我们在训练模型的过程中使用了代码中原有的数据集进行训练, 最后的整体图片明显发红, 经分析判断, 发现是训练集中存在偏色的图片导致红色通道分布异常, 使模型学习到错误的颜色分布。在网络上收集了部分数据集并对图片进行调整之后, 我们构建了一个新的训练集, 在经过相同的迭代次数之后, 明显新训练集的成果更好, Figure 8为生成的曝光图片的前后对比。

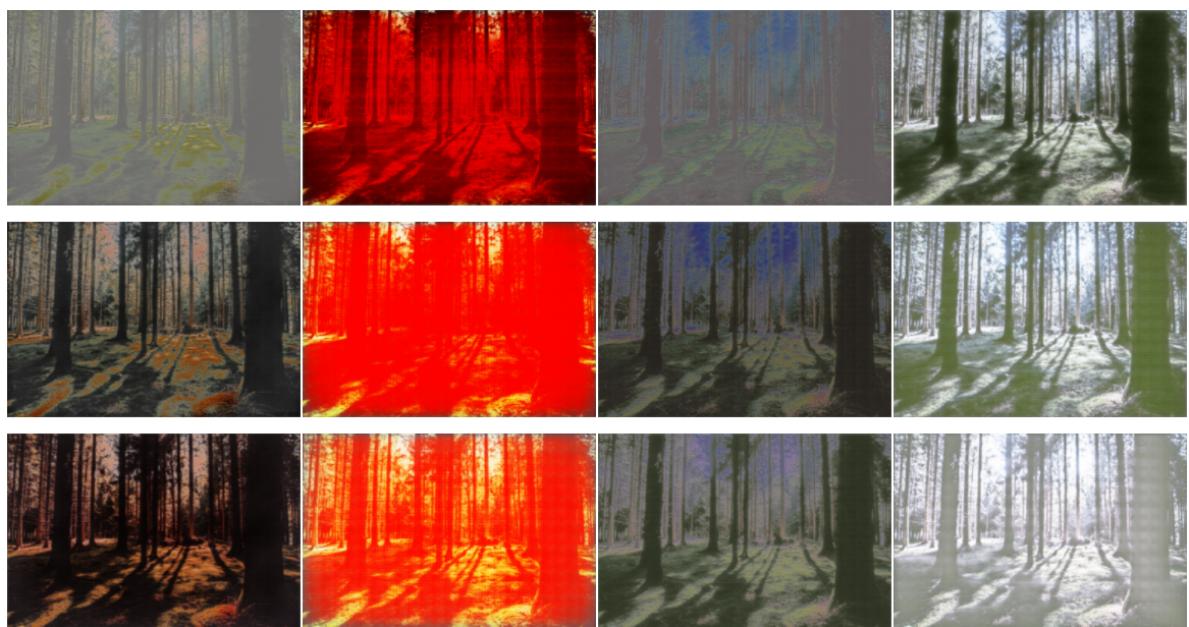


Fig.8 Construct a new training set for comparison. The six graphs on the left are the effects of the original training set, and the six graphs on the right are the effects of the reconstructed training set

## 5.1.2 第二阶段训练集

CF-Net：训练数据来自SICE数据集<sup>8</sup>，该数据集提供了7种不同曝光水平的多曝光图像序列。由于工作重点是解决极端曝光融合问题，因此我们只选择了一对极度过度曝光和曝光不足的图像进行训练。这些图像的一些例子如图i所示，其中涵盖了广泛的场景，包括人类、自然景观和人造建筑等。可以看到，曝光不足的图片极度黑暗，而过度曝光的图片极度明亮，这意味着很多细节都隐藏在两者之中。我们的方法能够还原这些细节，并进一步提高图像的分辨率。注意，ground-truth融合图像是由SICE数据集<sup>8</sup>提供的，这对训练过程有很大的好处。总共有450对曝光过度和曝光不足的图像，但限于设备，我们只能随机选择15对用于训练进行测试。

## 5.2 Quantitative Comparison Results

我们使用三个指标来衡量我们方法的性能，包括峰值信噪比(PSNR)、SSIM和MEF-SSIM。其中，PSNR和SSIM用于评估SR精度，MEF-SSIM用于描述融合性能。

Methods Combination	EDSR	RCAN	SRFBN	SAN	CFN	Ours
PSNR	18.20dB	18.20dB	18.16dB	18.18dB	22.52dB	9.45dB
SSIM	0.8136	0.8141	0.8100	0.8128	0.8698	0.5552
MEF-SSIM	0.8490	0.8489	0.8464	0.8481	0.9218	0.5923

Tab.1. Quantitative comparison with other advanced methods

这三个指标的值越高，表明图像的质量和融合性能越好，由于资源限制，我们的方法和模型是经过简化的版本，所以在性能的表现上并不如其他先进的方法，但我们有理由相信在相同的训练资源的情况下，我们的模型能够取得一个较好的表现。

## 5.3 Qualitative Comparison Results

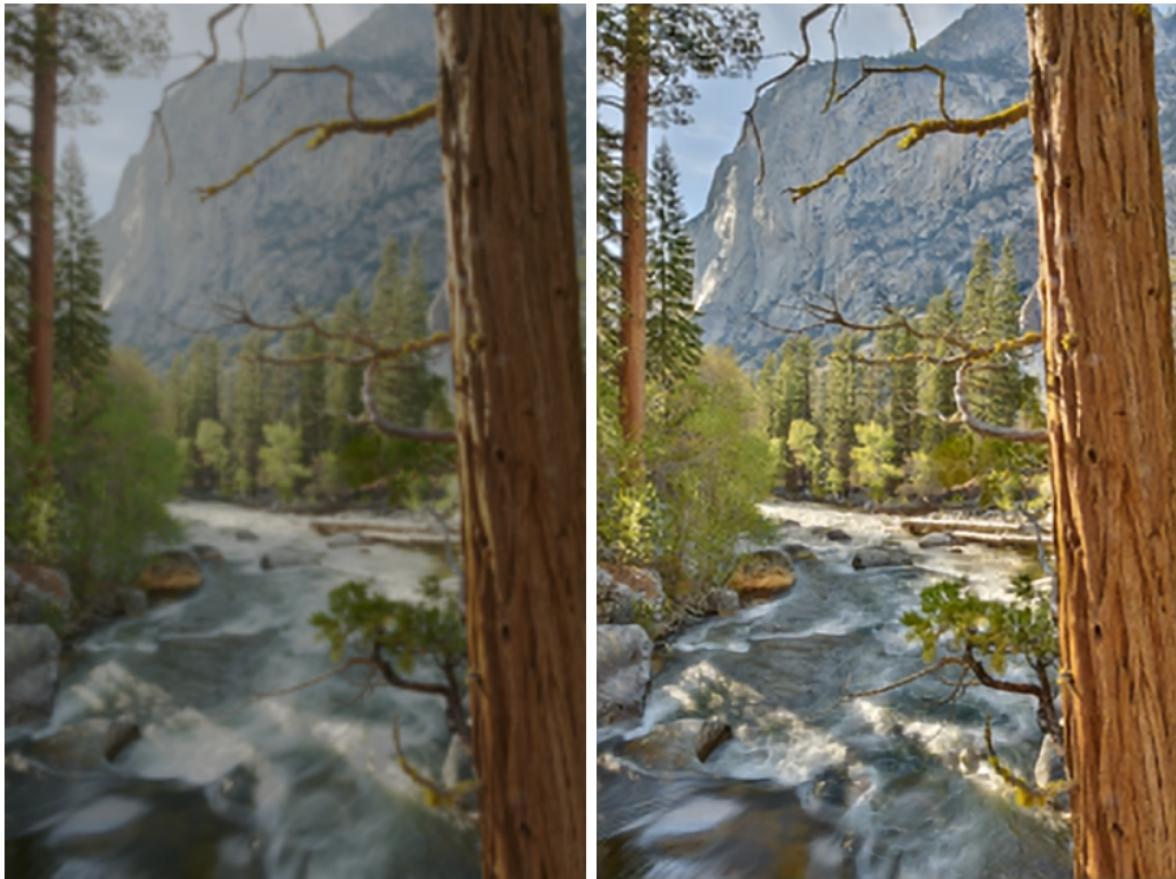


Fig.9. Qualitative comparison between our method and advanced methods. The left image is the image synthesized by our method, and the right image is the image synthesized by CFN

根据Figure 9可以看出，在排除对资源限制的影响看考虑，直观上可以看出，我们的模型在一定程度上对合成HDR图像存在其可行之处，在细节上若是有更好的处理，那么生成的图片质量将会得到大幅提高。

## 6. Discussion and Conclusions

在第一阶段我们采用的方法是CNN，但是实际上效果并不理想，希望能够尝试使用gan网络来进行生成不同曝光的图片，也是后续可以再作尝试的方向，参考论文<sup>9</sup>使用的GAN方法，针对使用深度学习的逆色调映射方法存在的问题，即从给定的图像生成具有不同曝光值的图像时需要额外的子网络，每一个子网络代表输入图像与不同曝光值图像之间的关系，造成了额外网络的数量线性增加，需要不同的数据集和优化过程来训练额外的网络。为了解决无法对图像中缺失的图案凭空恢复的问题，考虑曝光值的变换方向（正负）定义了两个神经网络，这些网络被约束为使用条件gan生成考虑相邻像素的图像，然后使用这些网络为给定的图像推断出具有相对曝光-T和+T的图像。这是我们目前想到可以对第一阶段实验的一种改进方法，但尚未进行实验以得到验证。

需要特别注意的是，对于实验的第二个阶段，即利用第一阶段生成的不同曝光的LDR图像合成HDR，硬件条件为NVIDIA 3050RTX GPU，显存为4GB。由于显存过小，经过不断尝试，发现训练能够使用的数据集至多为15张图片，而原数据集为478张，这将直接导致FEB和SRB模块对于低级特征和高级特征的学习不够充分也不够准确，从而对最后的SR图像重建和不同曝光图像的融合效果造成了较大的负面影响。另一方面，论文中采用了4次SR和不同曝光图像的融合，在实验过程中发现，由于硬件条件的限制，会出现无法分配足够空间的报错，所以在本实验中只进行了2次SR和曝光融合。另外，在试验过程中发现，将epoch设置为较大的数时，会在运行过程中出现电脑卡死的情况，原因不明，猜测是因为GPU显存不足，所以将epoch设置为了100，同时基本学习率也从1e-5相应地修改为1e-4.最后需要说明的是，CNN生成的不同曝光的LDR共有16张，但CF-Net的输入只需要一张极度高曝和一张极度低曝的图像，所以在融合之前，需要从16张不同的LDR中挑选两张曝光符合要求的图像，经过格式转换后放入CF-Net的

输入中。但是融合的过程中，再次出现了GPU空间不足的报错，经排查，发现是因为CNN生成的LDR图像过大，经过裁剪后问题解决。

## References

---

1. Kalantari N K, Ramamoorthi R. Deep high dynamic range imaging of dynamic scenes[J]. ACM Trans. Graph., 2017, 36(4): 144:1-144:12. 
2. Niu Y, Wu J, Liu W, et al. HDR-GAN: HDR image reconstruction from multi-exposed LDR images with large motions[J]. IEEE Transactions on Image Processing, 2021, 30: 3885-3896. 
3. Endo Y, Kanamori Y, Mitani J. Deep reverse tone mapping[J]. ACM Trans. Graph., 2017, 36(6): 177:1-177:10.   
4. Deng X, Zhang Y, Xu M, et al. Deep coupled feedback network for joint exposure fusion and image super-resolution[J]. IEEE Transactions on Image Processing, 2021, 30: 3098-3112.  
5. Siddique N, Paheding S, Elkin C P, et al. U-net and its variants for medical image segmentation: A review of theory and applications[J]. Ieee Access, 2021, 9: 82031-82057. 
6. Targ S, Almeida D, Lyman K. Resnet in resnet: Generalizing residual architectures[J]. arXiv preprint arXiv:1603.08029, 2016. 
7. M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1664–1673. 
8. J. Cai, S. Gu, and L. Zhang, "Learning a deep single image contrast enhancer from multi-exposure images," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 2049–2062, Apr. 2018.    
9. 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). 2018: 596-611. 