

基于循环注意力学习机制的欠曝光图像增强的深度对称网络（中文翻译版）

Lin Zhao^{1*} Shao-Ping Lu^{1*} Tao Chen² Zhenglu Yang¹ Ariel Shamir³

¹TKLNDST, CS, Nankai University, Tianjin, China

²Elephant Technologies, China

³The Interdisciplinary Center, Herzliya, Israel

lin-zhao@mail.nankai.edu.cn; {slu,yangzl}@nankai.edu.cn; tao.chen1@vcg.com; arik@idc.ac.il

摘要

欠曝光图像的增强在很多研究领域都有重要意义。在本文中，我们把此问题视为欠曝光图像与其成对增强版本之间的图像特征变换，并针对该问题提出了一种深度对称网络。我们利用可逆神经网络（*Invertible Neural Networks*, INN）作为对称网络进行图像间的双向特征学习，为确保相互传播的可逆性，我们专门构造了两对具有相同预训练参数的编码器-解码器。利用这种双向特征变换的可逆机制，我们可以在避免颜色偏差的同时有效地恢复图像增强的内容。此外，我们提出了一种新的循环残差注意力模型（*Recurrent Residual-Attention Module*, RRAM），其中循环学习网络旨在逐渐执行所需的颜色调整。执行消融实验可以显示我们的新结构中各组件的作用。我们在两个数据集上进行了大量实验，结果表明，我们的方法在欠曝光图像增强方面达到最优。代码可通过 <https://www.shaopinglu.net/proj-iccv21/ImageEnhancement.html> 获得。

1. 引言

由于数码相机在日常生活中的广泛使用，数码摄影越来越受欢迎。然而，糟糕的拍摄环境、不合适的相机参数或缺乏摄影技巧都可能导致图像质量不理想。很多时候，后期处理需要对照片颜色和局部细节等曝光敏感方面进行调整。

图像增强仍然是一个具有挑战性的问题，尤其是对欠曝光图像的增强处理。为了提高图像质量，我们需要对其颜色进行调整，同时保留图像的内容特征。传统的全局调整算法，比如直方图均衡化 [31, 40, 41]，对比度调整 [18, 44] 和伽马校正 [22, 42] 方法，无法编辑和改变图像中的局部细节。最近的基于神经网络的方法 [7, 11, 21, 34, 46] 在复杂的欠曝光条件下仍然会出现颜色偏差或伪影。确切地说，当照片在弱光环境下拍摄时，视觉特征被隐藏在黑暗区域。为修正这种情况，我们不仅需要调整颜色，还需要恢复内容（见图1中的示例）。有一些方法试图通过分别使用多个不同的模块来处理这些需求。然而，这种方案可能会引入累积的训练错误，并导致视觉伪影。

在本文中，我们将图像增强问题设想为欠曝光图及其增强版本（这可能是训练过程中的标准参考图）的图像对之间进行可逆特征变换的统一架构。因此，受最新可逆神经网络（INN, *Invertible Neural Networks*） [1, 8, 27, 48] 的启发，我们提出了一种基于可逆特征变换器（IFT, *Invertible Feature Transformer*）的深度对称网络。为高度实现前向和后向的传播操作，我们专门设计了两对完全共享相同参数的预训练编码器-解码器用于应对图像对（即，曝光不足和增强图像）与其相应特征之间的相互变换。我们的对称网络同步进行前向、后向学习，成功解决了由缺少大量成对图像训练数据和颜色特征学习困难导致的图像颜色偏差问题。为了准确地恢复图像的期望特征，我们进一步提出了一个带有循环残差注意力模型（RRAM, *Recurrent Residual-Attention Module*）的循环学习计划。

*共同一作。

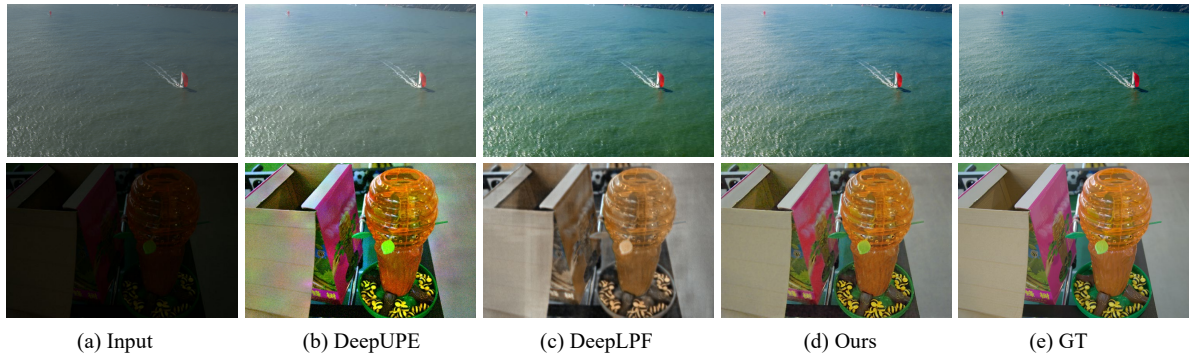


图 1. 不同方法对挑战性图像的处理结果。DeepUPE [46]和DeepLPPF [34]是目前最先进的办法。我们的方法在保证图像结构和纹理特征的同时，可以有效地调整图像颜色。

由于为特征图不同通道分配了不同的权值，我们的网络可以聚焦于恢复结构特征信息。在不增加网络参数数量的情况下，循环学习使我们的网络能逐步学习复杂的特征变换，从而实现对图像颜色的调整。在公共数据集上的大量实验证实了我们方法的优势。

综上所述，我们的贡献如下：

- 据我们所知，我们最先将INN引入到欠曝光图像增强问题中，该对称结构可以同步进行双向特征学习，与其他欠曝光图像增强的解决方案相比，我们取得了最优结果。
- 我们提出了一种具有循环残差注意力模型（RRAM）的特征变换循环学习方案，允许在不增加网络参数的情况下逐步进行颜色调整。

2. 相关工作

在过去的几十年中，人们引入了许多图像增强的研究方法。在这里，我们简要讨论一些重要工作，尤其是欠曝光图像增强方面的工作。此外，本节将介绍一些与我们工作最相关的方法。

现有图像增强方案。 有很多通过调整像素值实现图像增强的算法。一些传统的算法提出了增强图像对比度和亮度的方法。例如，Ying *et al.* [51]使用曝光融合框架增强图像对比度。Aubry *et al.* [2]采用快速局部拉普拉斯（Laplacian）滤波器来增强细节。近年来，深度学习已成功应用于图像增强。Gharbi *et al.* [11]提出了一种用于颜色变换的双边网络处理网络。为了估计全局先验并获得满意的性能，He *et al.* [16]提出了一个基础网络之外的条件网络。此外，基于深度强化学习的方法也被用来美化图像 [21, 37]。最近的一些工作

[7, 23, 24]也建立在生成性对抗网络（GANs, Generative Adversarial Networks）的基础上来解决这个问题。

具体来说，许多方法聚焦于增强极度欠曝光图像。由于原始图像容易丢失内容信息，人们提出了多种方法来恢复这些内容，其中一些方法还将恢复模块引入到图像增强管道中。但是，这一类的处理机制会导致累积的训练错误。Chen *et al.* [5]提出了一种关注原始传感器数据而非低光RGB图像的端到端管道来避免训练错误。Xu *et al.* [49]依次恢复低频层和高频层，在低频层恢复目标。此外，许多研究者根据同态滤波（Retinex）理论 [28]，将弱光图像增强作为亮度估计任务 [14, 29, 30, 36, 47]。最近，一些特定的数据集也被引入，以适应针对各种图像增强任务的新的训练策略。例如，Jiang *et al.* [25]对来自不同数据集域的微光图像采用基于U-Net的生成性对抗网络（GAN, Generative Adversarial Network）。Guo *et al.* [13]提出了一种可在没有标准参考图像的情况下进行训练的网络，以避免过度拟合的风险。Yan *et al.* [50]提出了从高质量图像中获得感知指导并使用半监督学习来增强图像的DRBN方法。

现有的一些方法试图在保证原始图像特征分布的同时增强图像。例如，为了解复杂的图像调整过程，Wang *et al.* [46]设计了一个图像光照映射模型。但是，该模型没有对图像进行全局调整。Moran *et al.* [34]提供了一种通过学习空间局部滤波器来增强图像的新方法，但由于没有考虑全局滤波器，他们的结果可能会引入一些颜色偏差。相反，我们的对称网络在进行图像增强时，既保持了图像颜色的一致性，又恢复了图像内容。最后，还有一些其他的任务，如人脸增强 [10, 39]和阴影增强 [52]。与这些特定的主题相比，我

们的工作更加一般化，可用于增强各种类型的欠曝光图像。

循环注意力模型。 与其他前馈神经网络不同，递归神经网络（RNN）通常以序列数据作为输入，并在序列 [17] 的渐进方向上以递归方式执行。RNN最初用于解决自然语言处理问题 [32, 45]，最近被引入到计算机视觉任务中 [9, 12, 35]。此外，Mnih *et al.* [33] 提出了用于图像分类的循环注意力机制。随后，具有相似思想的不同模型被广泛应用于其他任务中。例如，Chen *et al.* [6] 提出了一种基于RNN的视觉注意模型，用于学习视图序列进行三维形状分类；Haque *et al.* [15] 将循环注意模型引入身份识别；在 [26] 中，使用了一种软注意力机制来跟踪目标，Bendre *et al.* [3] 将一组帧序列放入RNN中进行人体行为识别。据我们所知，我们最先将循环注意力机制引入到欠曝光图像增强问题。

3. 对称网络

我们对称网络的总体架构如图2所示，在训练过程中进行双向传播操作。当网络执行前向操作（见绿色虚线箭头）时，预增强的输入图像进行前向变换处理。当处理反向操作（见红色虚线箭头）时，标准参考图像被可逆地变换为其欠曝光的版本。这个对称框架使我们的任务在双向传播方式下高度可解。

我们的对称网络包含一个基于最新的INNs（可逆神经网络） [1, 8, 27, 48] 的可逆特征变换器（IFT）。需要注意的是，由于缺乏深度特征，直接将现有INN（可逆神经网络）架构应用于图像块无法获得期望的图像增强效果（更多详细信息将在消融部分给出）。这就需要我们对图像特征进行更复杂的重组，以满足双向传播操作应高度可逆的硬约束。因此，我们专门设计了两对共享相同参数的预训练编码器-解码器网络。在我们系统的传播两端，编码器用于执行从图像到其相应特征的转换，而解码器用于将特征转换为相应的图像。

对于前向传播，我们使用第一对编码器-解码器在图像与其相应特征之间进行转换，我们的IFT（可逆特征变换器）执行前向特征学习。形式上，这种传播工作如下：

$$\begin{aligned} x_{f_{L_1}} &= E_1(x_{LQ}), \\ x_{f_{H_1}} &= IFT(x_{f_{L_1}}), \\ x_{HQ_f} &= D_1(x_{f_{H_1}}), \end{aligned} \quad (1)$$

其中 x_{LQ} 表示输入图像。 $x_{f_{L_1}} \Delta x_{HQ_f}$ 、 $x_{f_{H_1}}$ 分别表示编码器、解码器和IFT的输出结果。 $[E_1(\cdot)$ 、 $D_1(\cdot)]$ 和 $IFT(\cdot)$ 分别是前向编码器-解码器和前向特征变换。类似地，第二对编码器-解码器和IFT（可逆特征变换器）涉及反向传播，可公式化为：

$$\begin{aligned} x_{f_{H_2}} &= E_2(x_{HQ}), \\ x_{f_{L_2}} &= IFT_R(x_{f_{H_2}}), \\ x_{LQ_f} &= D_2(x_{f_{L_2}}), \end{aligned} \quad (2)$$

其中 x_{HQ} 表示标准参考图像。 $x_{f_{H_2}} \Delta x_{LQ_f}$ 、 $x_{f_{L_2}}$ 分别是相对应的结果。 $[E_2(\cdot)$ 、 $D_2(\cdot)]$ 和 $IFT_R(\cdot)$ 分别表示后向编码器-解码器和后向特征变换。因此，我们的网络在两个传播方向上保持了特征的一致性，双向约束解决了欠曝光图像的颜色偏差问题。

3.1. 预训练的编码器和解码器

如图2所示，我们采用对称编译器-解码器结构进行图像与特征之间的转换，确保了图像涉及的特征的完整性。我们的目的是保证 $(x_{f_{L_1}}, x_{f_{L_2}})$ 和 $(x_{f_{H_1}}, x_{f_{H_2}})$ 两对高度一致并保持图像的全局特征，使图像的结构信息在欠曝光图像增强中得到很好的保存。

为了确保所提出的框架是可逆的，我们进一步限制编码器与译码器的参数完全相同。在我们的解决方案中，这些参数是从ImageNet [43] 上预训练VGG-16模型的前两个卷积层中提取的。需要注意的是，更多的CNN（卷积神经网络）层数会使全局特征被破坏，重构结果也会因此受到影响。

3.2. 可逆特征变换器（IFT）

欠曝光图像通常有颜色和内容的损失问题。为了正确地恢复图像而不产生伪影，我们的IFT（可逆特征变换器）不仅学习从低质量图像到高质量图像的正向映射，还学习从高质量图像到低质量图像的反向映射。我们的IFT（可逆特征变换器）由几个可逆块组成（默认为8个）。对于第 i 个块，输入特征 x_f^i 根据通道号平均划分为 $x_{f_1}^i$ 和 $x_{f_2}^i$ ，然后通过转换模块：

$$\begin{aligned} x_{f_1}^{i+1} &= \frac{x_{f_1}^i - T_{i,3}(x_{f_2}^i)}{ES(T_{i,1}(x_{f_2}^i))}, \\ x_{f_2}^{i+1} &= x_{f_2}^i - T_{i,2}(x_{f_1}^{i+1}), \end{aligned} \quad (3)$$

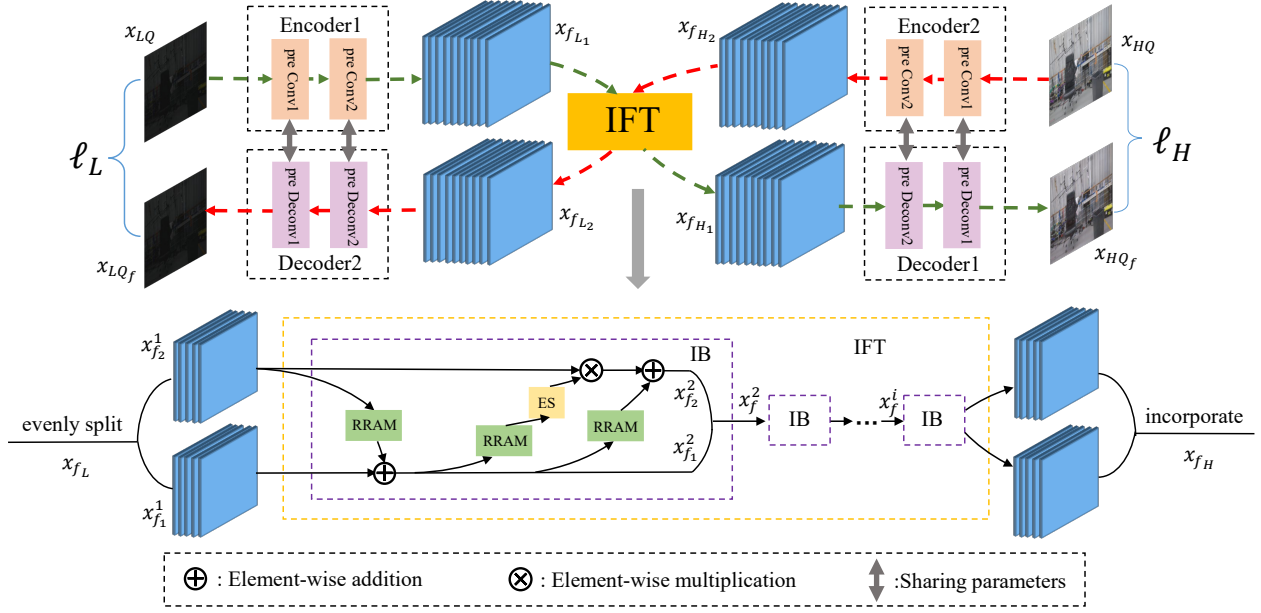


图 2. 我们方法的总体网络架构。绿色和红色虚线箭头分别表示网络的双向传播操作。preConvi 和 preDeconvi 分别为使用预训练 VGG-16 [43] 的卷积层参数 i ($i=[1, 2]$) 进行卷积和反卷积。在特征 x_f^i 进入第 i 个可逆块 (IB) 之前, x_f^i 根据通道数量等分为 $x_{f_1}^i$ 和 $x_{f_2}^i$ 。

其中 $T_{i,j}(\cdot)$ 表示第 i 个块中第 j ($j=1, 2, \text{ or } 3$) 个转换模块。ES(\cdot) 表示带指数的 S 形 (Sigmoid) 函数。我们使用 ES(\cdot) 作为乘数来增强转换能力。同样, 当反向传播发生时, 很容易得出:

$$\begin{aligned} x_{f_2}^i &= x_{f_2}^{i+1} + T_{i,2}(x_{f_1}^{i+1}), \\ x_{f_1}^i &= x_{f_1}^{i+1} * ES(T_{i,1}(x_{f_2}^i)) + T_{i,3}(x_{f_2}^i). \end{aligned} \quad (4)$$

3.3. 循环残差注意力模型 (RRAM)

许多曝光不足和标准参考的图像对之间存在明显的色差, 使用现有方法很难消除。因此, 我们引入 RRAM (循环残差注意力模型) 来解决这个问题。考虑到图像间颜色特征的变换学习仍然是一个非常具有挑战性的问题, 这里我们将 RRAM 作为一个多轮递归模块, 在不增加网络参数的情况下, 利用注意力结构对目标进行 t 轮的逐步递归学习。

多轮循环学习。 我们循环学习的核心思想是将需要解决的任务划分为几个连续步骤, 并在欠曝光图像的增强任务中逐步学习颜色调整。如图 3 所示, 我们使用 $[h^1, h^2, \dots, h^t]$ 将一个隐藏状态序列表示为 RNN (递归神经网络), t 是循环轮数。此外, 隐藏状态使用多轮内存可存储前几轮中获得的特征信息。我们网络的

特点是使用对应元素加法来减轻每个模块的学习任务。正式来说, 第 t 轮:

$$h^t = \begin{cases} 0, & \text{if } t = 1 \\ h^{t-1} + x_{fo}^{t-1}, & \text{otherwise} \end{cases}, \quad (5)$$

其中 x_{fo}^{t-1} 表示第 $(t-1)$ 轮残差软通道注意力机制的输出。因此, 第 t 轮递归式表示为:

$$x_{fo}^t = f_{RSCA}(W_h h^t + W_x x_{fi}), \quad (6)$$

其中, RSCA 表示残差软通道注意力机制, W_h 和 W_x 表示平衡权值。我们选择 $W_h = W_x = 1$ 和 $t = 3$ 。通过这种设计, 隐藏状态有效地包含了之前学习到的特征信息, 并且它可以集中在每一轮中遗漏的剩余信息中。因此, RRAM (循环残差注意力模型) 可以有效学习欠曝光图像与其标准参考图像之间的明显色差。

残差软通道注意力机制。 采用残差软通道注意力机制使网络更加关注通道结构信息。如图 4 所示, 该机制首先在转换中进行特征学习, 如下所示:

$$x_h = f_C(\delta(f_C(x_{fi}))), \quad (7)$$

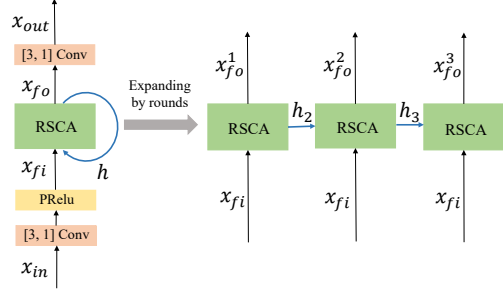


图 3. 我们的循环残差注意力模型RRAM的结构图。蓝线表示隐藏状态的转移。RSCA表示残差软通道注意力机制 (Residual Soft Channel Attention Mechanism)。[i, j]Conv 表示一个需要 $i \times i$ 内核大小和 $j \times j$ 步幅大小的卷积运算。

其中 $f_C(\cdot)$ 和 $\delta(\cdot)$ 表示卷积函数和参数校正线性单元 (Prelu), x_{fi} 和 x_h 表示该机制的输入和获得的特征。接下来, 动态学习了 x_h 通道之间的相互依赖关系。为了实现这一目标, 全局平均池将 $C \times H \times W$ 大小的特征 x_h 转换为 $C \times 1 \times 1$ 大小的未学习图 M_i , 作为SENet [19], 该图在一定程度上具有全局感受野。通过学习注意力图, 我们自动获得了每个特征通道的重要性:

$$M_o = f_{CS}(f_{CR}(M_i)), \quad (8)$$

其中 $f_{CS}(\cdot)$ 和 $f_{CR}(\cdot)$ 分别表示后跟Sigmoid层的卷积和后跟Prelu层的卷积。 M_o 是获得的用来与 x_h 相乘的软注意力图。此外, 我们使用残差学习优化网络, 如下所示:

$$x_{fo} = x_h \times M_o + x_{fi}, \quad (9)$$

其中 x_{fo} 表示获得的机制特征。

如先前所述, 转换到每个可逆块 (即 图2中的紫色虚线框) 时, 特征平均分为两部分。为充分提取通道的特征信息, 了解通道间关系, 我们在RRAM的前端和末端都加倍并恢复了特征通道的数量,

$$x_{fi} = \delta(f_C(x_{in})), \quad (10)$$

$$x_{out} = f_C(x_{fo}^t), \quad (11)$$

其中 x_{in} 和 x_{out} 表示RRAM的输入和输出特性。 x_{fi} 和 x_{fo}^t 分别为我们注意力机制的输入和第 t 个输出。

3.4. 损失函数

在两种传播操作中, 我们都使用 L_2 距离作为训练

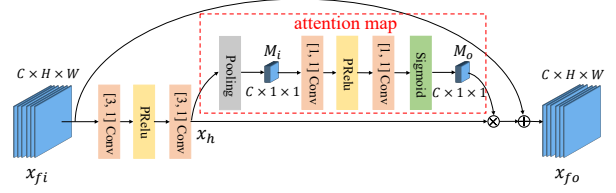


图 4. 我们的残差软通道注意力机制, 采用2D自适应平均池化。 C 、 H 、 W 分别表示通道数、特征长度、特征宽度。

损失函数。前向传播操作的损失函数定义为:

$$\ell_H = \frac{1}{N} \sum_{i=1}^N \|x_{HQ_f} - x_{HQ}\|^2, \quad (12)$$

其中 N 表示训练图像的个数。同理, 向后传播损失函数为:

$$\ell_L = \frac{1}{N} \sum_{i=1}^N \|x_{LQ_f} - x_{LQ}\|^2. \quad (13)$$

因此, 我们将最终的训练损失定义为以上两种损失的加权和:

$$\ell_{tr} = \lambda_H \ell_H + \lambda_L \ell_L, \quad (14)$$

其中, λ_H 和 λ_L 为平衡权重。在我们的训练中, 我们根据经验设定 $\lambda_H = \lambda_L = 1$ 。

4. 实验

数据集. 我们以MIT-Adobe FiveK [4]和LOL [47]数据集作为基准数据集对我们的模型进行测试。在MIT-Adobe FiveK [4]数据集中, 有 5000 张低质量图像, 每一张图像都由 5 位专家处理。我们遵循 [7, 34, 37, 46]准则, 选取Expert C调整后的图像作为标准参考图像。原始的MIT-Adobe FiveK [4]数据集只代表了欠曝光图像的某些特定分布, 当面对输入值的分布变化较大时, 可能会导致较差的泛化结果。在 [37]中, 标准参考图像经过多个随机畸变算子的预处理, 进一步合成各种欠曝光结果作为输入。这里我们完全遵循这个过程。我们分别使用了前 4500 张图像和后 500 张图像进行训练和测试。我们使用LOL [47]数据集来测试我们的模型对于极端欠曝光图像的性能。LOL [47]数据集包含 500 个低光/正常光真实图像对, 我们使用400对图像进行训练, 100对图像进行测试。

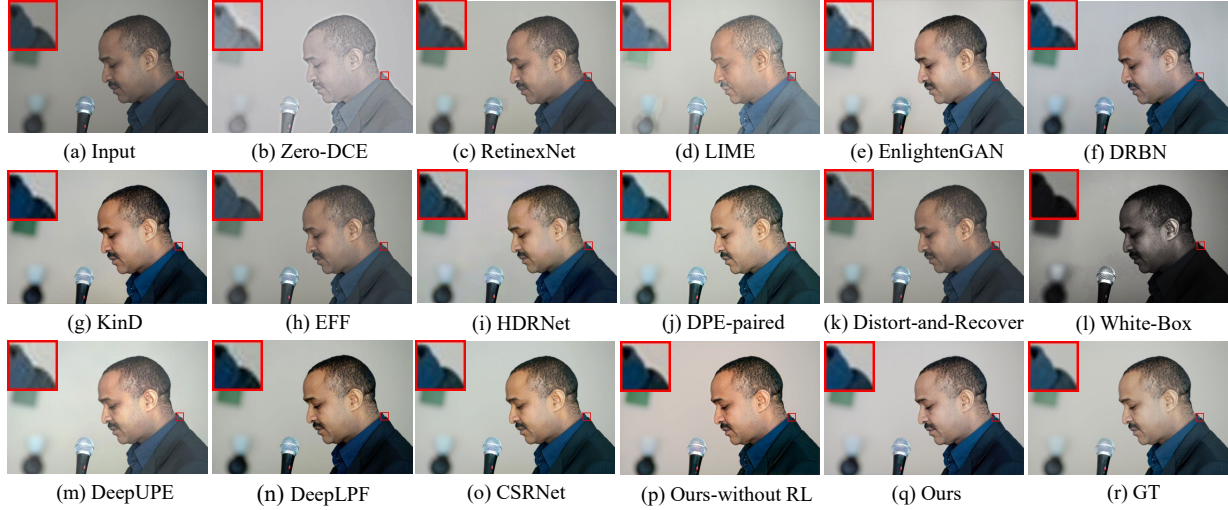


图 5. 在MIT-Adobe FiveK [4]数据集上, 对欠曝光图像以不同方法处理后的视觉结果。在这里, 我们也展示了我们的方法在未使用循环学习RL的情况下的结果。

实现细节。 我们的网络在Pytorch [38] 和 Jittor [20]两个平台上实现, 使用Nvidia 2080Ti显卡。对于每个数据集上的训练, 我们将原始图像随机剪裁为 180×180 大小的图像块, 并使用初始学习率为 $2e-4$ 的ADAM优化器。我们的网络对MIT-Adobe FiveK [4]数据集共训练200期, 且每训练50期学习率降低一半。对于LOL [47]数据集, 相应的训练期数值为250和100。我们使用PSNR (峰值信噪比) 和SSIM (结构相似性) 作为测试的图像质量评估标准。更高的PSNR (峰值信噪比) 和SSIM (结构相似性) 值意味着更好的处理效果。

4.1. 与最先进技术的对比实验

为了验证我们的方法的有效性, 我们将其与另外14个现有方法进行了比较: HDRNet [11], DPE (paired) [7], Distort-and-Recover [37], White-Box [21], 对于欠曝光图像的CSRNet [16], Zero-DCE [13], RetinexNet [47], LIME [14], EFF [51], EnlightenGAN [25], DRBN [50], 对于极度欠曝光图像的KinD [53], DeepUPE [46], 适用于两种情况的DeepLPF [34]。为了确保对比的公平性, 我们对所有方法进行了重训练。

定量比较。 表1显示了这些方法在两个数据集上的结果。从上到下的三个部分是聚焦于弱光增强、图像修饰以及两者皆有的方法。我们的方法在LOL数据集上获得了除SSIM之外的最大值。

定性比较。 评估可应用于两个数据集。我们从MITAdobe FiveK数据集中选择一张欠曝光图像,

Method	LOL	FiveK
	PSNR/SSIM	PSNR/SSIM
Zero-DCE [13]	13.08/0.470	12.30/0.673
RetinexNet [47]	17.03/0.707	20.20/0.781
LIME [14]	16.92/0.540	14.30/0.731
EnlightenGAN [25]	17.79/0.769	21.28/0.818
EFF [51]	16.94/0.592	18.15/0.784
DRBN [50]	19.24/0.847	21.71/0.855
KinD [53]	20.08/0.822	21.72/0.833
HDRNet [11]	19.62/0.716	23.29/0.842
DPE (paired) [7]	18.08/0.659	21.67/0.846
Distort [37]	20.46/0.666	21.29/0.812
White-Box [21]	17.59/0.633	17.30/0.755
CSRNet [16]	19.57/0.681	24.13/0.878
DeepUPE [46]	16.78/0.468	20.83/0.795
DeepLPF [34]	16.58/0.678	23.63/0.875
w/o RL	20.63/0.826	23.32/0.888
ours	21.71/0.834	24.27/0.900

表 1. 不同方法对MIT-Adobe FiveK [4]和LOL [47]数据集的定量结果。“w/o RL”指的是未进行循环学习时我们的方法。“Distort”表示 Distort-and-Recover 方法。

图5显示所有方法的结果对比。可以看出, Zero-DCE、LIME、White-Box的结果存在颜色偏差, RetinexNet、EFF、DRBN、Distort-and-Recover无法调整图像亮度, DeepLPF无法恢复局部信息。DPE配对和CSRNet没有正确调整图像颜色。EnlightenGAN、KinD、HDRNet、Distort-and-Recover和DeepUPE会在图像中产生伪影, 导致一些纹理细节消失。

如图6所示, 我们与多个图像修饰的方法进行对比, 可以看出我们的方法在这方面占据优势。当输入

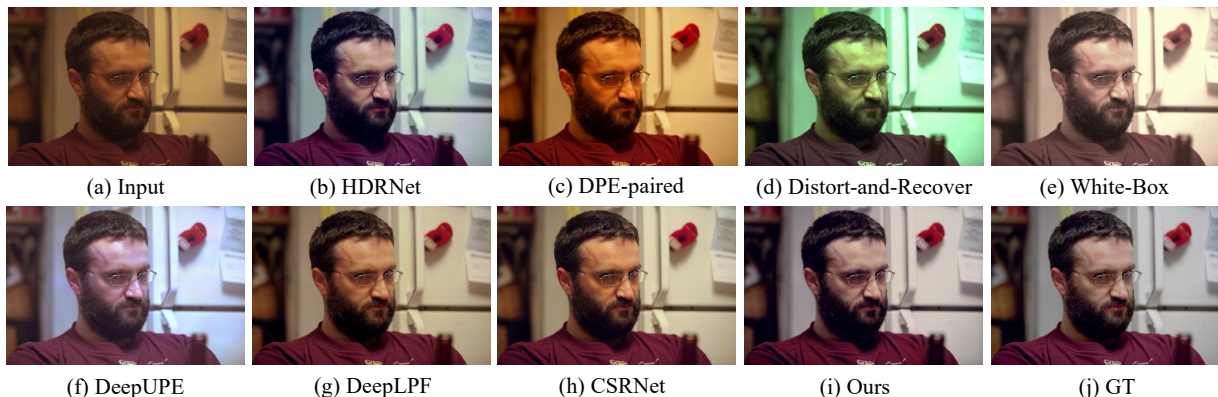


图 6. 在MIT-Adobe FiveK [4]数据集上图像修饰算法的对比结果。

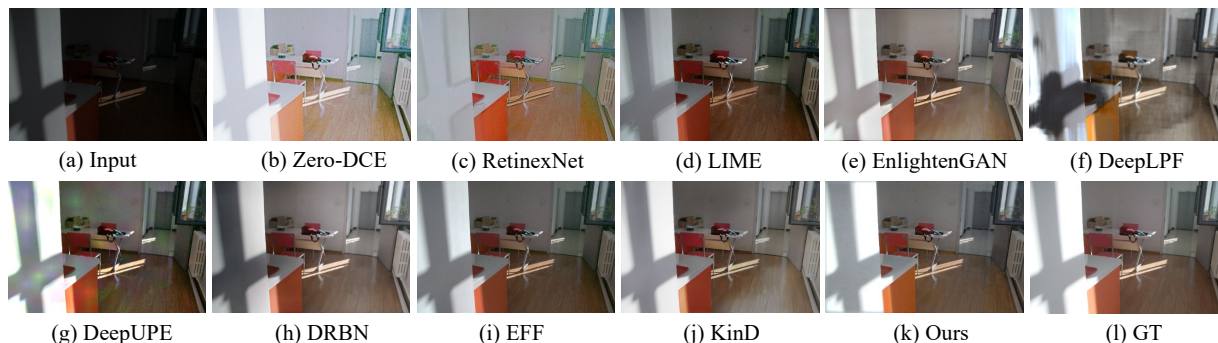


图 7. 各方法对LOL [47]数据集中的图像进行低光图像增强的结果。

图像与标准参考图像存在色差时，我们的结果图像比其他方法更贴近标准参考图像。

此外，我们在LOL数据集上和其他低光图像增强方法进行比较，图7显示了一些极度欠曝光场景的结果。一般来说，RetinexNet、DeepLPF、KinD和DRBN方法获取的结果并未很好地恢复图像颜色。此外，Zero-DCE、LIME、DeepUPE、EFF和EnlightenGAN等方法更可能产生视觉伪影。总的来说，我们的方法在极度欠曝光图像增强方面取得了更好的效果。

4.2. 消融实验

如图7和表1所示，大多数方法在LOL数据集上都不能取得令人满意的结果，因此我们使用数据集进行消融实验，详细说明我们每个模块的作用。

在表2中，我们从三个不同的方面进行了实验，验证了双向转换和损耗、RRAM（循环残差注意力模型）和对称结构的效果。对于双向转换，我们取消了向后损失，以验证IFT（可逆特征变换器）的意义。我们还

将损耗函数改为 l_1 ，使PSNR [7]（峰值信噪比）最大，以进行比较。对于RRAM，我们将其改为其他网络。此外，我们还去除了评估重复学习有效性的注意力机制，并改变循环学习的次数来证明循环学习的合理性。为了验证对称结构的作用，我们改变了可逆块的数量，专门用Haar小波替换了我们预先训练好的编码器-解码器，这些小波通常用于之前基于INN（可逆神经网络）的网络 [27, 48]。从表2中各变化点的PSNR和SSIM可以看出，我们各模块对性能提高都有贡献。

为了进一步说明注意力机制和具有双向损失的对称结构的具体作用，我们在图8中展示了一些具有代表性的图像的对比结果。从图8 (b)可以看出，在RRAM（循环残差注意力机制）中除去注意力机制后，有一部分内容是我们无法恢复的。当网络不进行向后学习时（见图8 (c)），图像曝光部分会出现误差。此外，基于Haar小波的结果存在明显的伪影。

最后，我们也验证了我们的循环学习架构的有效性。如图5 (p)-(q)和图9 (b)-(c)所示，当我们比较我们的网络在有无循环学习机制这两种情况下的结果时，

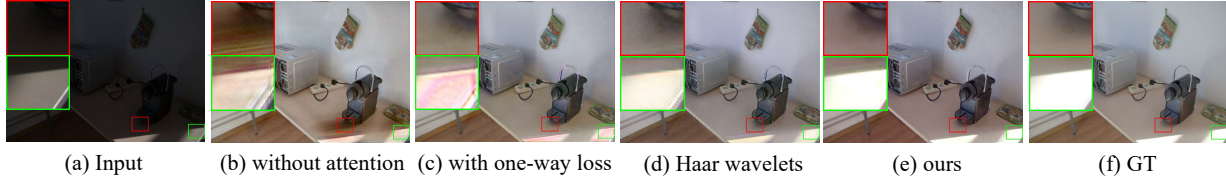


图 8. 用我们的方法对LOL [47]数据集中的图像进行不同消融实验的视觉结果。

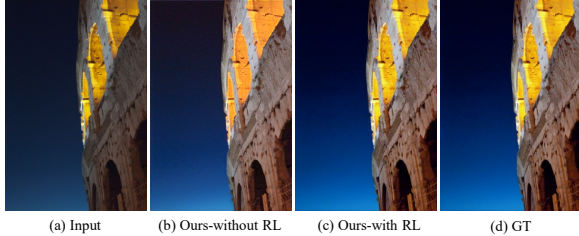


图 9. 有无对图像循环学习的视觉结果表现了很明显的颜色调整效果。

Condition	PSNR	SSIM
$L_2 \rightarrow L_1$ loss	-0.98	-0.024
$L_2 \rightarrow$ PSNR loss	-0.20	-0.007
w/o backward loss	-1.28	-0.014
w/o attention	-3.10	-0.038
RESBLK	-3.50	-0.067
SE-RESBLK	-1.32	-0.034
DenseNet	-2.74	-0.039
w/o RL	-1.08	-0.008
two-round RL	-0.42	-0.009
4 IB	-1.42	-0.026
Haar wavelets	-1.74	-0.027

表 2. 在LOL [47]数据集中，对不同消融实验与我们方法的默认设置进行定量比较。这里“RL”表示循环学习，“IB”表示可逆块，“SE-RESBLK”表示嵌入SENet [19]的残块。

很明显，只有通过循环学习，网络才能正确地学习图像颜色调整。此外，没有循环学习机制时，两个数据集上的PSNR/SSIM值将下降（表1）。

4.3. 用户测试

个人偏见和专业背景强烈影响图像修饰的性能和美学。为了更好地评估我们的方法，我们邀请了20名参与者进行用户研究，其中5名专家是专业摄影师/编辑，有多年的工作经验，而其他参与者是随机搜索的图像增强领域的业余爱好者。我们从他们日常的5类修饰任务中选择了50张原始照片，分别是“人”、“夜”、“建筑”、“自然”和“极端天气”。对于每个类别，我们随机选择10张照片。我们认为这个测试集很好地代表了真实世界的图片修饰任

Me \ Pe	CSRNet	DeepUPE	DeepLPF	ours
experts	23.20%	7.60%	8.80%	60.40%
amateurs	24.13%	8.40%	22.13%	45.33%

表 3. 在用户研究者首选方法的百分比。‘Me’和‘Pe’分别表示方法和百分比。

务。对于每个图像，我们分别运用CSRNet、DeepLPF、DeepUPE和我们的方法生成4个修饰结果。选择了现有的这三种方法是因为它们在定量评估方面表现较好，工作量也易于管理。用于测试网络稳健性的模型全部是来自MIT-Adobe FiveK数据集的训练模型。我们要求每位测试者综合考虑每张照片的颜色、亮度、对比度和伪影（照片和结果在补充材料中），并选择最佳的修饰效果。最后，我们分别统计每种方法在专家和业余人员评选结果中所占的百分比。从表3可以看出，无论测试者是专家还是业余爱好者，我们的方法在评选中所占比例最高。

5. 总结

在本文中，我们提出了一种对称深度网络，它包括一个可逆特征变换器（IFT）和两对预先训练的编码器-解码器。对称结构允许训练过程向两个方向传播，从而保持欠曝光和增强图像对的特征一致。另外，我们的循环注意残差模型（RRAM）帮助系统更好地实现所需的色彩调整与复杂的特征变换。此外，RRAM中的残差软通道注意力机制通过关注特征通道之间的相互依赖关系使我们的网络能够更好地恢复结构特征。我们进行了大量的定量和定性对比实验，证明了这种方法的优越性。

致谢。 感谢各位审稿人的宝贵意见。感谢国家自然科学基金(No. 61972216 and No. 62111530097)、天津市自然科学基金(No. 18JCYBJC41300 and No. 18ZXZNGX00110)、北京信息技术国家研究中心(No. BNR2020KF01001)和以色列科学基金1390/19对本工作的支持。本文的通讯作者是卢少平。

参考文献

- [1] Lynton Ardizzone, Carsten Lüth, Jakob Kruse, Carsten Rother, and Ullrich Köthe. Guided image generation with conditional invertible neural networks. *arXiv preprint arXiv:1907.02392*, 2019. 1, 3
- [2] Mathieu Aubry, Sylvain Paris, Samuel W Hasinoff, Jan Kautz, and Frédo Durand. Fast local laplacian filters: Theory and applications. *ACM TOG*, 33(5):1–14, 2014. 2
- [3] Nihar Bendre, Nima Ebadi, John J Prevost, and Peyman Najafirad. Human action performance using deep neuro-fuzzy recurrent attention model. *IEEE Access*, 8:57749–57761, 2020. 3
- [4] Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. Learning photographic global tonal adjustment with a database of input/output image pairs. In *CVPR*, pages 97–104, 2011. 5, 6, 7
- [5] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *CVPR*, pages 3291–3300, 2018. 2
- [6] Songle Chen, Lintao Zheng, Yan Zhang, Zhixin Sun, and Kai Xu. Veram: View-enhanced recurrent attention model for 3d shape classification. *IEEE TVCG*, 25(12):3244–3257, 2018. 3
- [7] Yu-Sheng Chen, Yu-Ching Wang, Man-Hsin Kao, and Yung-Yu Chuang. Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans. In *CVPR*, pages 6306–6314, June 2018. 1, 2, 5, 6, 7
- [8] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016. 1, 3
- [9] Jeffrey Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, and Trevor Darrell. Long-term recurrent convolutional networks for visual recognition and description. In *CVPR*, pages 2625–2634, 2015. 3
- [10] Shan Du and Rabab K Ward. Adaptive region-based image enhancement method for robust face recognition under variable illumination conditions. *IEEE TCSVT*, 20(9):1165–1175, 2010. 2
- [11] Michaël Gharbi, Jiawen Chen, Jonathan T Barron, Samuel W Hasinoff, and Frédo Durand. Deep bilateral learning for real-time image enhancement. *ACM TOG*, 36(4):1–12, 2017. 1, 2, 6
- [12] Karol Gregor, Ivo Danihelka, Alex Graves, Danilo Jimenez Rezende, and Daan Wierstra. Draw: A recurrent neural network for image generation. *arXiv preprint arXiv:1502.04623*, 2015. 3
- [13] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *CVPR*, pages 1780–1789, 2020. 2, 6
- [14] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE TIP*, 26(2):982–993, 2016. 2, 6
- [15] Albert Haque, Alexandre Alahi, and Li Fei-Fei. Recurrent attention models for depth-based person identification. In *CVPR*, pages 1229–1238, 2016. 3
- [16] Jingwen He, Yihao Liu, Yu Qiao, and Chao Dong. Conditional sequential modulation for efficient global image retouching. In *ECCV*, 2020. 2, 6
- [17] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982. 3
- [18] A Howie. Image contrast and localized signal selection techniques. *Journal of Microscopy*, 117(1):11–23, 1979. 1
- [19] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *CVPR*, pages 7132–7141, 2018. 5, 8
- [20] Shi-Min Hu, Dun Liang, Guo-Ye Yang, Guo-Wei Yang, and Wen-Yang Zhou. Jittor: a novel deep learning framework with meta-operators and unified graph execution. *Science China Information Sciences*, 63(12):1–21, 2020. 6
- [21] Yuanming Hu, Hao He, Chenxi Xu, Baoyuan Wang, and Stephen Lin. Exposure: A white-box photo post-processing framework. *ACM TOG*, 37(2):1–17, 2018. 1, 2, 6
- [22] Shih-Chia Huang, Fan-Chieh Cheng, and Yi-Sheng Chiu. Efficient contrast enhancement using adaptive gamma correction with weighting distribution. *IEEE TIP*, 22(3):1032–1041, 2012. 1
- [23] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. Dslr-quality photos on mobile devices with deep convolutional networks. In *ICCV*, pages 3277–3285, 2017. 2
- [24] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. Wespe: Weakly supervised photo enhancer for digital cameras. In *CVPRW*, pages 691–700, 2018. 2
- [25] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang

- Wang. Enlightengan: Deep light enhancement without paired supervision. *arXiv preprint arXiv:1906.06972*, 2019. [2](#), [6](#)
- [26] Samira Ebrahimi Kahou, Vincent Michalski, Roland Memisevic, Christopher Pal, and Pascal Vincent. Ratm: recurrent attentive tracking model. In *CVPRW*, pages 1613–1622, 2017. [3](#)
- [27] Durk P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. In *NeurIPS*, pages 10215–10224, 2018. [1](#), [3](#), [7](#)
- [28] Edwin H Land. The retinex theory of color vision. *Scientific american*, 237(6):108–129, 1977. [2](#)
- [29] Mading Li, Jiaying Liu, Wenhan Yang, Xiaoyan Sun, and Zongming Guo. Structure-revealing low-light image enhancement via robust retinex model. *IEEE TIP*, 27(6):2828–2841, 2018. [2](#)
- [30] Xujie Li, Hanli Zhao, Guizhi Nie, and Hui Huang. Image recoloring using geodesic distance based color harmonization. *Computational Visual Media*, 1(2):143–155, 2015. [2](#)
- [31] Shao-Ping Lu, Sen-Mao Li, Rong Wang, Gauthier Lafruit, Ming-Ming Cheng, and Adrian Munteanu. Low-rank constrained super-resolution for mixed-resolution multiview video. *IEEE TIP*, 30:1072–1085, 2021. [1](#)
- [32] Tomáš Mikolov, Stefan Kombrink, Lukáš Burget, Jan Černocký, and Sanjeev Khudanpur. Extensions of recurrent neural network language model. In *ICASSP*, pages 5528–5531, 2011. [3](#)
- [33] Volodymyr Mnih, Nicolas Heess, Alex Graves, and Koray Kavukcuoglu. Recurrent models of visual attention. *arXiv preprint arXiv:1406.6247*, 2014. [3](#)
- [34] Sean Moran, Pierre Marza, Steven McDonagh, Sarah Parisot, and Gregory Slabaugh. Deeplpf: Deep local parametric filters for image enhancement. In *CVPR*, pages 12826–12835, 2020. [1](#), [2](#), [5](#), [6](#)
- [35] Lichao Mou, Pedram Ghamisi, and Xiao Xiang Zhu. Deep recurrent neural networks for hyperspectral image classification. *IEEE TGRS*, 55(7):3639–3655, 2017. [3](#)
- [36] Qi Mu, Xinyue Wang, Yanyan Wei, and Zhanli Li. Low and non-uniform illumination color image enhancement using weighted guided image filtering. *Computational Visual Media*, 2021. [2](#)
- [37] Jongchan Park, Joon-Young Lee, Donggeun Yoo, and In So Kweon. Distort-and-recover: Color enhancement using deep reinforcement learning. In *CVPR*, pages 5928–5936, 2018. [2](#), [5](#), [6](#)
- [38] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017. [6](#)
- [39] Eli Peli, Estella Lee, Clement L Trempe, and Sheldon Buzney. Image enhancement for the visually impaired: the effects of enhancement on face recognition. *JOSA A*, 11(7):1929–1939, 1994. [2](#)
- [40] Stephen M Pizer. Contrast-limited adaptive histogram equalization: Speed and effectiveness stephen m. pizer, r. eugene johnston, james p. ericksen, bonnie c. yankaskas, keith e. muller medical image display research group. In *Proceedings of the First Conference on Visualization in Biomedical Computing, Atlanta, Georgia*, volume 337, 1990. [1](#)
- [41] Stephen M Pizer, E Philip Amburn, John D Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart ter Haar Romeny, John B Zimmerman, and Karel Zuiderveld. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, 39(3):355–368, 1987. [1](#)
- [42] Shanto Rahman, Md Mostafijur Rahman, Mohammad Abdullah-Al-Wadud, Golam Dastagir Al-Quaderi, and Mohammad Shoyaib. An adaptive gamma correction for image enhancement. *EURASIP Journal on Image and Video Processing*, 2016(1):1–13, 2016. [1](#)
- [43] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *Computer Science*, 2014. [3](#), [4](#)
- [44] J Alex Stark. Adaptive image contrast enhancement using generalizations of histogram equalization. *IEEE TIP*, 9(5):889–896, 2000. [1](#)
- [45] Sainbayar Sukhbaatar, Jason Weston, Rob Fergus, et al. End-to-end memory networks. In *NeurIPS*, pages 2440–2448, 2015. [3](#)
- [46] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In *CVPR*, pages 6849–6857, 2019. [1](#), [2](#), [5](#), [6](#)
- [47] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *BMVC*, 2018. [2](#), [5](#), [6](#), [7](#), [8](#)
- [48] Mingqing Xiao, Shuxin Zheng, Chang Liu, Yaolong Wang, Di He, Guolin Ke, Jiang Bian, Zhouchen Lin, and Tie-Yan Liu. Invertible image rescaling. *arXiv preprint arXiv:2005.05650*, 2020. [1](#), [3](#), [7](#)

- [49] Ke Xu, Xin Yang, Baocai Yin, and Rynson WH Lau. Learning to restore low-light images via decomposition-and-enhancement. In *CVPR*, pages 2281–2290, 2020. [2](#)
- [50] Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *CVPR*, pages 3063–3072, 2020. [2](#), [6](#)
- [51] Zhenqiang Ying, Ge Li, Yurui Ren, Ronggang Wang, and Wenmin Wang. A new image contrast enhancement algorithm using exposure fusion framework. In *CAIP*, pages 36–46, 2017. [2](#), [6](#)
- [52] Xuaner Cecilia Zhang, Yun-Ta Tsai, Rohit Pandey, Xiuming Zhang, Ren Ng, David E Jacobs, et al. Portrait shadow manipulation. *arXiv preprint arXiv:2005.08925*, 2020. [2](#)
- [53] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *ACM MM*, pages 1632–1640, 2019. [6](#)