河南科技学院2023-2024学年第二学期

《数字图像处理应用》

文献翻译报告

一种基于隐式神经表示（INR）的低光图像增强方法

学 院 名 称**： 人** 工 智 能 学 院

专 业 名 称**：** 人工智能

年 级 班 别： 223班

小 组 成 员 （学号，姓名）：

20222214303 党硕硕

20222214301 程景晗

20222214302 程文豪

20222214305 惠子卿

任 课 教 师： 郑建锋

2024年06月

**摘 要**

现有的低光照图像增强方法在应用中受到以下三个因素的限制：不可预测的亮度退化和噪声、度量友好版本与视觉友好版本之间的固有差距、以及有限的配对训练数据。为了解决这些限制，我们提出了一种用于协作低光照图像增强的隐式神经表示方法，称为NeRCo。该方法能够以无监督的方式稳健地恢复感知友好的结果。具体来说，NeRCo通过一个可控的拟合函数统一了真实场景中的多种退化因素，从而提高了鲁棒性。此外，对于输出结果，我们引入了基于预训练视觉语言模型的语义导向监督。它不仅仅是遵循参考图像，而是鼓励结果满足主观期望，找到更视觉友好的解决方案。进一步地，为了减少对配对数据的依赖并缩小解决方案空间，我们开发了一个双闭环约束增强模块。该模块与其他附属模块协同训练，以自监督的方式进行。

**关键词：**低光图像增强；隐式神经表示；多任务学习；协作优化；对比度增强

**目录**

[1绪论 1](#_Toc170381132)

[2相关工作 2](#_Toc170381133)

[2.1 低光照图像增强 2](#_Toc170381134)

[2.2 图像的神经表示 3](#_Toc170381135)

[2.3 多模态学习 3](#_Toc170381136)

[3我们的方法 4](#_Toc170381137)

[3.1 框架架构 4](#_Toc170381138)

[3.2 神经表示归一化 4](#_Toc170381139)

[3.3文本驱动外观判别器 6](#_Toc170381140)

[3.2.1 动机 6](#_Toc170381141)

[3.4文本驱动判别 6](#_Toc170381142)

[3.4.1基于外观的判别 7](#_Toc170381143)

[3.5双循环生成过程 8](#_Toc170381144)

[3.5.1双循环 8](#_Toc170381145)

[3.5.2协作损失 8](#_Toc170381146)

[4实验 9](#_Toc170381147)

[4.1 实现细节 10](#_Toc170381148)

[4.2 与最先进技术的比较 11](#_Toc170381149)

[4.3 消融研究 12](#_Toc170381150)

[4.4结论 12](#_Toc170381151)

[5. 结论 13](#_Toc170381152)

[参考文献 14](#_Toc170381153)

[隐式神经表示用于协作低光图像增强 - 补充文档 19](#_Toc170381154)

[摘要 19](#_Toc170381155)

[A. 归一化结果 19](#_Toc170381156)

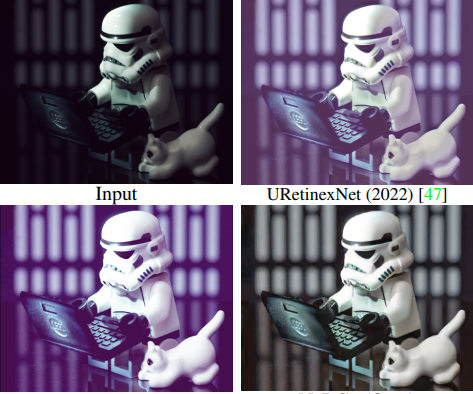
[B. 使用替代文本提示的实验 19](#_Toc170381157)

[C. TAD消融 20](#_Toc170381158)

[D. 语义评估 21](#_Toc170381159)

[E. 定性分析 21](#_Toc170381160)

## 1绪论

由于低光照图像中的亮度退化覆盖了物体并降低了对比度，低光照图像严重影响了后续的高级计算机视觉任务（例如语义分割和目标检测等）。因此，解决亮度退化问题对于探索复杂的黑暗环境具有实际意义。低光照图像增强旨在恢复退化区域中的期望内容，近年来引起了广泛关注。

图表 1与 LIME [12] 数据集上的两种最先进方法进行比较。可以看出，我们恢复了更真实的颜色和视觉友好的内容。

在过去的几年中，许多算法被提出以解决这一经典的病态问题，这些算法大致可以分为两类：传统的基于模型的方法（例如伽马校正、Retinex模型和直方图均衡化）和最近的基于深度学习的方法。前者将退化公式化为物理模型，并将增强视为估计模型参数的问题，但在表征多样的低光照因素方面存在局限性，并且需要大量手工设计的先验知识。后者则通过调整色调和对比度的各种模型，能够自动从大量数据中学习。本质上，它们被训练来学习从输入域到输出域的映射。然而，在现实场景中，许多样本远离输入域的特征空间，导致训练模型缺乏稳定效果。我们提出在增强之前对退化进行归一化，以使这些样本更接近输入域。此外，现有的监督方法高度依赖于配对训练数据，并主要尝试生成度量友好的结果，即与真实值相似。但有限的监督数据集和度量导向与视觉友好版本之间的固有差距不可避免地影响了它们的有效性。我们开发了一种协作训练策略来解决这一问题。

具体来说，我们的关键见解是：使用可控拟合函数对输入进行归一化，以减少现实场景中不可预测的退化特征。我们采用神经表示在增强操作之前再现退化场景。通过操纵位置编码，我们选择性地避免再生极端退化，从而实现归一化并降低增强难度。

使用不同模态监督输出，以实现度量友好和感知导向的增强。我们采用多模态学习，从文本和图像角度进行监督。与包含不同样本亮度的图像监督相比，设计的提示的特征空间在描述亮度方面更稳定和准确。在训练过程中，我们的结果不仅被鼓励与参考相似，还被强制匹配其相关提示。通过这种方式，我们弥合了度量友好和感知友好版本之间的差距。

开发了一种无监督训练策略，以减少对配对数据的依赖。我们提出使用双闭环协作对抗约束程序来训练增强模块，该程序以无监督方式学习。还提出了更多相关的损失函数，以进一步减少解决方案空间。受益于这些，我们恢复了更真实的色调和更好的对比度。

总体而言，我们的贡献如下：

我们首次利用神经表示的可控拟合能力进行低光照图像增强。它在不进行任何额外操作的情况下归一化亮度退化并去除自然噪声，为未来工作提供了新思路。我们首次将多模态学习引入低光照图像增强。受益于其高效的视觉语言先验，我们的方法学习了多样的特征，产生了感知上更好的结果。

我们开发了一种无监督协作对抗学习策略，以减少对配对训练数据的依赖。在这种策略中，基于外观的判别确保了从颜色和细节层面上的真实性，提高了恢复结果的质量。在代表性基准上进行了大量实验，证明了我们NeRCo相对于一系列最先进算法的优越性。特别是，它甚至超过了一些监督方法。

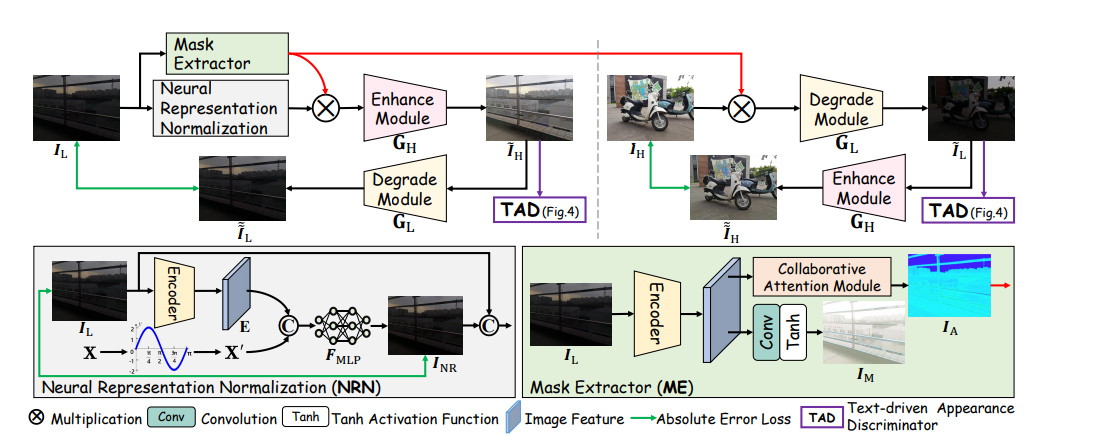
## 2相关工作

## 2.1 低光照图像增强

为了提高低光照图像的可见性，首先广泛采用了基于模型的方法。Retinex理论将观察分解为照明和反射（即清晰预测），但往往会过度曝光外观。各种手工设计的先验进一步被引入模型作为正则化项。Fu等人开发了一种加权变分模型，以同时估计反射和照明层。Cai等人提出了一种边缘保留平滑算法来建模亮度。Guo等人通过采用相对全变分来预测照明。然而，这些定义的先验是劳动密集型的，并且在现实场景中表现出较差的泛化能力。

由于这些限制，研究人员利用深度学习以数据驱动的方式进行恢复，自动从大量数据中利用先验。例如，Guo等人将光增强公式化为一种图像特定曲线估计任务，使用轻量级深度模型。Jiang等人引入了对抗训练以从未配对的监督中学习。Wei等人设计了一种端到端可训练的RetinexNet，但仍然受到严重噪声的困扰。为了缓解这一问题，Zhang等人提出了一种分解型架构来对反射施加约束。Liu等人采用架构搜索并构建了一个展开网络。尽管这些精心设计的模型实现了令人印象深刻的效果，但在现实应用中并不稳定。为了提高鲁棒性，我们在增强程序之前通过神经表示预调制退化到统一水平。

## 2.2 图像的神经表示

最近，神经表示被广泛用于描述图像。Chen等人首先利用隐式图像表示进行连续图像超分辨率。然而，多层感知器（MLP）倾向于扭曲高频成分。为了解决这个问题，Lee等人开发了一种主频估计器来预测自然图像的局部纹理。Lee等人进一步利用隐式神经表示将图像扭曲成连续形状。Dupont等人通过操纵隐藏层中的潜在代码尝试用一个MLP生成不同的对象。Saragadam等人采用多个MLP以多尺度方式表示单个图像。Sun等人基于捕获的断层特征预测连续信息。Tancik等人引入元学习来初始化MLP的参数以加速训练。Reed等人采用神经表示和参数运动场来预测器官的形状和位置。此外，一些研究人员采用神经表示来压缩视频。

图表 2我们的 NeRCo 的工作流程。它提出了一个包含双闭环分支的协作对抗增强过程，每个分支包含一个增强操作和一个退化操作。我们嵌入了一个 Mask Extractor (ME) 来描绘退化分布，以及一个神经表征规范化 (NRN) 模块来规范化输入低光图像的退化状况。所有这些都一起训练以相互约束，锁定到更准确的目标域。红色表示注意力图的转移。

然而，现有的神经表示主要应用于图像压缩、去噪和描述连续信息等。我们是第一个将其可控拟合能力应用于低光照图像增强的人。

## 2.3 多模态学习

近年来，跨模态学习引起了广泛关注。各种视觉语言模型被开发出来。Radford等人提出从语言监督中学习视觉模型，称为CLIP。在训练了4亿对图像-文本对后，它可以用自然语言描述任何视觉概念，并在没有任何特定训练的情况下转移到其他任务。此外，Zhou等人开发了软提示来替代手工设计的提示，使用可学习的向量来建模上下文词并获得任务相关的上下文。为了进一步将提示细化到实例级别，Rao等人设计了上下文感知提示，将提示与视觉特征结合。Cho等人通过更新统一框架以实现七个多模态任务的共同目标来共享不同任务的先验。Ju等人采用预训练的CLIP模型进行视频理解。

## 3我们的方法

## 3.1 框架架构

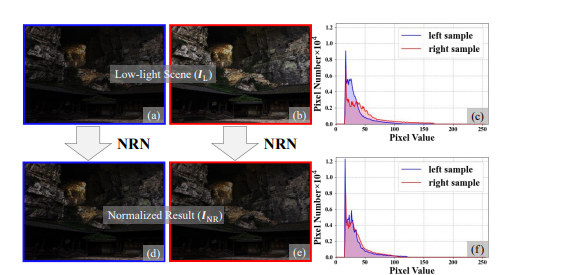
如图2所示，给定一张低光图像IL，我们首先通过神经表示规范化（NRN，见第3.2节）对其进行规范化，以提高模型对不同退化条件的鲁棒性。然后，掩码提取器（ME，见第3.4节）模块从图像中提取注意力掩码，以引导不同区域的增强。之后，增强模块GH（以ResNet表示）生成高光图像eIH。为了确保其质量，我们设计了一个文本驱动的外观鉴别器（TAD，见第3.3节）来监督图像生成，其中文本驱动的监督保证了语义可靠性，外观监督保证了视觉可靠性。eIH然后通过退化模块GL转换回低光域eeIL，并与原始低光图像IL计算一致性损失（见第3.4节）。图2的右上分支输入高光图像IH，实现方式类似。

该网络以双循环方式实现（见第3.4节），基于无配对数据实现稳定约束。它进行双向映射：增强-退化（IL到eIH再到eeIL）和退化-增强（IH到eIL再到eeIH）。这种双循环约束充分利用了低光域和高光域之间的潜在一般区分。此外，协作损失（见第3.4节）鼓励框架中的所有组件相互协作监督，进一步减少了解空间。

在训练过程中，我们运行图2中的整个过程。我们输入两张图像（即低光图像IL和高光图像IH）。IL被增强为eIH，然后转换回低光eeIL。同样，对于IH也是如此。需要注意的是，IH仅用于训练目的，即以无监督的方式训练模型以获得更好的增强而不是退化。因此，我们只使用NRN来增强IL，而不退化IH。在推理时，IL直接增强为eIH，如图2左上部分所示，没有其他操作。

## 3.2 神经表示归一化

具体来说，在NR中，图像IL被转换为特征图E ∈ R^H×W×C，其中H和W是图像分辨率。每个像素的位置记录在坐标集X ∈ R^H×W×2中，其中2表示水平和垂直坐标。IL可以通过其特征和一组坐标来表示。如图2中的神经表示归一化（NRN）模块所示，我们融合了X和E，并使用解码函数FMLP输出图像INR，该函数参数化为多层感知器（MLP）。图像的神经表示表达为：

其中[i, j]是像素的位置，INR[i, j]是生成的RGB值。通过预测每个像素的RGB值，重现图像INR。我们通过l1范数鼓励INR与IL相似。这个与NR相关的损失表达为：

图表 3捕获的低光场景（上行）与 NRN 的结果（下行）之间的比较。低光样本来自 SICE [3] 数据集。它包含许多图像集，每个图像集都有共同的内容和不同的照明条件。右侧给出了 Y 通道上图像的像素值分布。可以看出，NRN 将亮度标准化为相似。

神经表示的工作原理在于训练后的FMLP，每个特征图E可以形成一个函数FMLP(E, ·)：X → INR，将坐标映射到其预测的RGB值。没有E，FMLP无法描绘具有相同坐标X的各种RGB值。没有X，我们无法通过调整拟合能力来归一化退化，这将在下文中解释。

根据[36]，神经网络倾向于描绘较低频的信息。尽管我们的解码函数可以近似RGB值，但在渲染过程中可能会丢弃一些高频成分。例如，对于边缘附近的相邻像素，它们的RGB值变化很大，但坐标变化很小。这意味着FMLP应该基于相似的输入输出不同的结果，这是困难的。

受[30]启发，为了更好地适应高频变化，我们在将输入坐标传递给FMLP之前，将其映射到更高维空间，这称为位置编码。如图2的灰色区域所示，在将坐标与图像特征融合之前，我们使用一个高频函数γ(·)将原始坐标x从R映射到更高维空间R^2L，表达为：

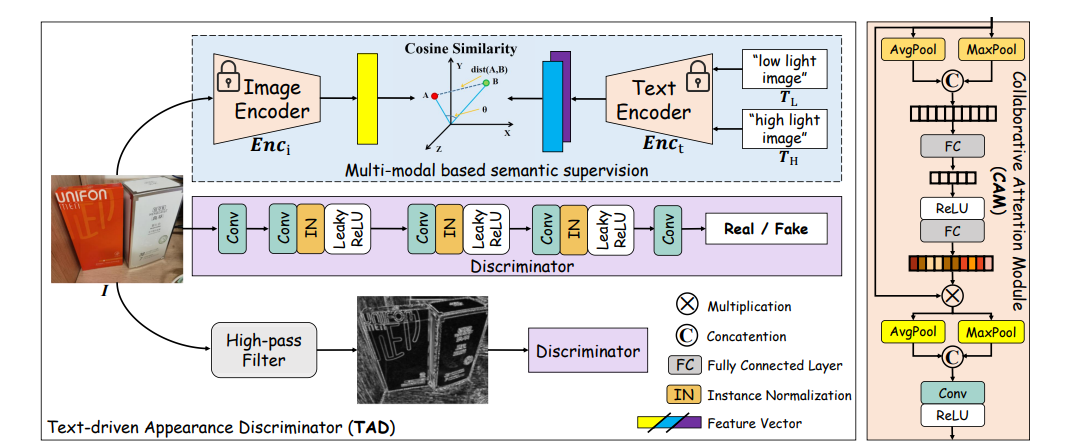
**γ(x) = (· · ·, sin(2iπx), cos(2iπx), · · ·)**

其中i的取值范围从0到L-1，L是决定维度值的超参数。最终的坐标组成为：x' = γ(x)。通过调整L的值，我们可以改变NRN模块的拟合能力，即更大的L会导致更精确的拟合。

然而，更强的拟合能力并不总是更好的。由于NRN旨在归一化各种退化，其输出不应该完全与输入相同，特别是退化成分。我们希望选择一个不会过度拟合的L，以保留所有信息同时忠实地保留所需内容。Chen等人已经证明，NR对扰动具有鲁棒性，并且可以在没有任何特殊设计的情况下去噪。我们认为这是因为MLPs缺乏空间相关性先验，一些极端信息因此难以忠实再现。因此，这种欠拟合属性客观地限制了不可预测的退化。我们在实验中进一步发现，MLPs倾向于学习训练数据的平均亮度范围，而不是拟合不同图像的独特亮度。如图3所示，经过训练的MLP的输出亮度是相似的。通过在Y通道中可视化它们的像素值分布，我们进一步证明了我们NRN的归一化属性。更多结果详见补充材料（SM）。因此，我们设置L = 8，以在退化归一化和内容保真度之间取得平衡。通过减少这些多变的退化信号，我们降低了后续增强过程的难度。

## 3.3文本驱动外观判别器

## 3.2.1 动机

现有方法主要采用图像级别监督，即强制输出接近目标图像。然而，不同参考图像之间的亮度差异很大，使模型训练变得困难；一些参考图像视觉效果较差（例如，亮度不自然），导致视觉上不友好的结果。为了减少训练难度并弥合度量有利和视觉友好版本之间的差距，我们设计了一个文本驱动外观判别器（TAD），分别从语义级别和外观级别监督图像生成。

图表 4我们提出的文本驱动外观鉴别器（左侧区域）和我们的协作注意模块（右侧区域）的细节。前者使用文本和图像模态监督输入，并关注高频分量。后者将注意力调整到不同的通道并输出注意力图。

## 3.4文本驱动判别

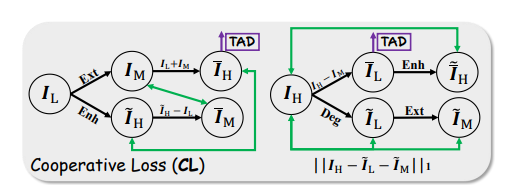
我们将低光照域表示为L，将高光域表示为H。如图4所示，我们引入多模态学习，以同时监督具有图像和文本模态的图像。具体来说，受Radford等人的启发，我们采用最近知名的CLIP模型来获得高效的先验。它由两个预训练编码器组成，即Enct用于文本和Enci用于图像。我们首先手动设计两个提示，即低光照图像和高光照图像，分别描述为TL和TH。在SM中提供了更多其他文本的实验。Enct从两个提示中提取大小为1×512的两个特征向量。类似地，Enci从我们的中间结果I中提取相同大小的向量。我们计算图像向量和文本向量之间的余弦相似度，以衡量它们的差异，公式如下：

## 3.4.1基于外观的判别

诚然，文本描述无法像图像那样提供低级别的指导。为了生成忠实的内容，图像监督是必要的。如图4的紫色区域所示，我们堆叠了一个判别器，用于区分预测结果和真实图像，鼓励图像级别的真实性（例如，颜色、纹理和结构）。考虑到图像处理中的细节失真，我们嵌入了一个由高通滤波器和相同结构的判别器组成的高频路径。该滤波器提取高频成分，判别器在边缘级别对其进行监督。基于这种双路径颜色边缘判别结构，我们实现了颜色和细节之间的权衡。

在训练过程中，TAD在学习L和H之间的双向映射关系中扮演对抗角色。我们在每个生成循环上开发对抗损失来实现它。如图2所示，对于增强操作GH(·)：IL → eIH，我们应用一个TAD模块，并将其外观判别称为DH。制定一个对抗目标函数如下：

其中，DH旨在确定图像是捕获还是生成的，即区分增强结果GH(IL)和真实高光域H。而GH(·)旨在欺骗DH，即生成接近H的结果。同时，我们还采用上述余弦约束来监督GH(·)。反向映射GL(·)：IH → eIL采用类似的目标函数，由另一个TAD进行监督。

在这两个函数中，生成器试图最小化目标，而TAD模块则试图最大化目标。在这种对抗学习过程中，我们的模型实现了更好的语义一致性，这在SM中的实验中得到了证明。

图表 5所提出的合作损失函数的详细信息。

## 3.5双循环生成过程

以前的方法主要将低光照图像IL直接映射到其高光照版本eIH。为了在没有配对数据的情况下提供稳定约束，我们堆叠了一个前向增强模块和一个后向退化模块，用于双向映射，以无监督方式运行。

## 3.5.1双循环

具体来说，前向增强过程旨在实现映射GH(·)：IL → eIH。而另一个则相反，即通过映射GL(·)：IH → eIL从干净图像IH中描述低光照场景。我们的生成循环由交替执行这两个操作组成。如图2左端所示，我们的输入是观察到的低光照图像。首先提取一个注意力引导IA，并通过神经表示模块进行归一化。然后，后续过程首先将其转换为高光照域，然后将增强图像eIH映射回低光照版本eeIL。这个增强-退化生成分支表达为：

其中，GH(·)和GL(·)分别表示增强和退化操作，NRN是第3.2节讨论的神经表示归一化模块。ME表示掩码提取器，如图2的绿色区域所示，在那里我们开发了一个协作注意力模块（CAM）来提取注意力图IA。CAM的细节显示在图4的右端。

退化-增强生成分支，如图2的右端所示，类似地表达为：

为了约束这种双向映射，在训练过程中，我们开发循环一致性，直接在像素级别施加监督。例如，对于图2中的左分支，我们确保：IL ≈ eeIL = GL(GH(IL))。相应地，另一个循环如下：IH ≈ eeIH = GH(GL(IH))。我们采用l1范数来衡量差异，并制定坐标一致性约束如下：

## 3.5.2协作损失

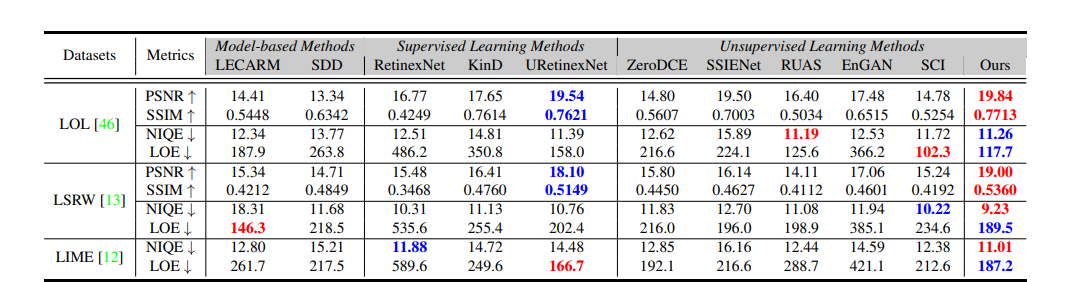
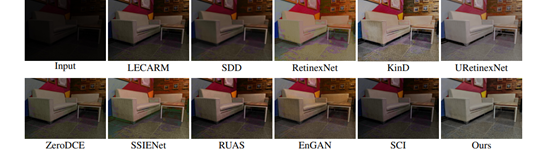
为了减少解决方案空间并增强注意力引导，我们详细阐述了协作损失（CL）。受Lee等人的启发，该函数通过施加相互约束来协同训练不同模块。如图5所示，我们间接监督注意力引导IA，并为所有模块提供更强的约束。具体来说，在图2的掩码提取器（ME）中，由于注意力图IA是基于提取的图像特征生成的，特征的内容严重影响IA的质量。为了获得更好的IA，我们从相同特征中生成一个亮度掩码IM，并与其他信息一起进行共同监督，包括我们增强的图像。

对于给定的低光照输入IL，一方面，我们使用ME提取IM，并通过将IM与IL相加获得伪高光照图像IH。另一方面，我们将IL与预测的高光照结果eIH在像素级别相减，得到一个伪亮度掩码IM。通过一致性损失，我们鼓励估计的IM与提取的IM保持一致，计算得到的IH与描绘的eIH保持相似，表达为：

对于给定的高光照输入IH的另一个分支，如图5的右侧所示，损失类似：

对于基于IM计算的图像，即IH和IL，我们进一步使用TAD的双路径判别器检查它们的真实性，表达为：

其中，H和L分别表示伪高光照和低光照图像域。最终的协作损失为：

结合第3.2节中的与神经表示相关的损失，第3.3节中的对抗约束以及第3.4节中的一致性损失，我们的最终目标函数表达为：

图表 6在三个基准上进行比较。最佳和第二佳结果分别以红色和蓝色突出显示。

图表 7在 LSRW 数据集上对最先进的低光图像增强算法进行主观比较

## 4实验

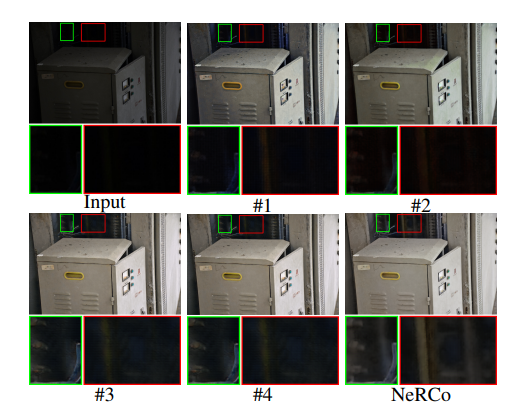
在本节中，我们首先介绍我们方法的实现细节。然后，我们通过多个基准测试将其与最先进的方法进行比较。为了确定每个组件的贡献，我们进一步进行消融分析。所有实验均使用PyTorch在单个NVIDIA Tesla V100 GPU上进行。

## 4.1 实现细节

**参数设置：**对于训练，我们采用Adam优化器[18]，超参数设置为β1 = 0.5，β2 = 0.999，ϵ = 10^-8。我们的模型训练300个epochs，初始学习率为2×10^-4，并在最后200个epochs线性衰减至0。批量大小设置为1，为了效率起见，将补丁大小调整为256×256。启发式地，我们采用具有3个隐藏层的MLP来归一化退化水平。

**基准和指标：** 为了验证我们方法的有效性，我们在LSRW数据集[13]上训练和测试模型，该数据集包含1000对低光照-正常光照图像用于训练，以及50对用于评估。每对图像包括一个退化图像和一个曝光良好的参考图像，这些图像是在现实世界中使用不同曝光时间拍摄的。为了进行更有说服力的比较，我们进一步将评估扩展到其他基准测试，如LOL[46]和LIME[12]。由于LOLv1仅包含15张图像用于评估，我们从LOLv2的测试集中随机抽取35张图像（未用于训练），并在这50张图像上进行评估。为了展示对真实际退化情景的泛化能力，我们在LSRW数据集上训练的模型上进行了LIME数据集的测试。需要注意的是，为了证明我们无监督学习方式的优越性，在训练过程中，我们仅采用配对训练数据的低光照部分，并用来自BSD300数据集[29]的300张图像替换参考图像。我们使用两种全参考指标，即峰值信噪比（PSNR）和结构相似性指数（SSIM）[45]，以及两种无参考指标，即NIQE [31]和LOE [44]，来客观评估不同算法的有效性。一般来说，更高的PSNR或SSIM意味着更真实的恢复结果，而更低的NIQE或LOE表示更高质量的细节、亮度和色调。

## 4.2 与最先进技术的比较

为了进行更全面的分析，我们将我们的方法与两种最近提出的基于模型的方法（即LECARM [39]和SDD [14]）、三种先进的监督学习方法（即RetinexNet [46]、KinD [55]和URetinex-Net [47]）以及五种无监督学习方法进行比较，包括ZeroDCE [11]、SSIENet [53]、RUAS [25]、EnGAN [16]和SCI [28]。

图表 8消融研究的视觉结果。全套表现最佳，尤其是在绿色和红色框出的区域。

**定量分析：** 我们通过采用官方预训练模型和运行它们各自的公开代码，获得其他方法的定量结果。如表1所示，在所有基准测试中，我们的方法在全参考和无参考指标上几乎达到了最先进的性能。这验证了所提出框架的优越效果。值得注意的是，我们的方法甚至表现优于监督方法。与一些最近竞争激烈的无监督方法（如EnGAN [16]和SCI [28]）相比，我们提供了比EnGAN更强的约束，包括来自提示的面向视觉的指导。此外，由于归一化属性，我们的方法在一些具有挑战性场景上的表现尤为出色，特别是在一些挑战性场景上，我们的方法优于SCI。下面给出了在一个严重亮度退化情景中的所有方法（包括NeRCo和SCI）的视觉结果。

**定性分析：** 为了进行更直观的比较，我们在图6中报告了所有方法的视觉结果。我们的输入是一个严重退化的图像。可以看到，最近的传统方法无法恢复足够的亮度。而基于先进深度学习的方法会过度平滑背景或引入未知的遮蔽物，导致糟糕的伪影和不自然的色调。特别是，SCI未能有效增强这种具有挑战性的情景。相比之下，我们的模型实现了最佳的视觉质量，具有明显的对比度和生动的颜色。细节也得到了很好的保留。更多结果请参见SM。

## 4.3 消融研究

图表 9对不同设置下获得的增强结果的定量评估。最佳结果和第二佳结果分别以红色和蓝色突出显示。

如图3所示，我们考虑了四种消融设置，逐步将提出的组件添加到双循环中。所有消融研究均在LSRW数据集上进行。

i）“#1”是一个简单的双循环，没有其他操作，仅采用普通颜色判别器。这个框架取得了可比较的分数，表明其有效性。

ii）“#2”在“#1”的基础上添加了掩码提取器（ME）和相关的协作损失（CL）。ME提供了注意力引导，CL减少了解决方案空间。需要注意的是，“#2”进一步采用边缘路径判别器，但没有文本监督。

iii）“#3”在“#2”的基础上添加了TAD，利用来自文本模态的视觉导向指导。关于TAD的进一步消融研究将在SM中给出。

iv）“#4”在“#2”中添加了NRN，但没有TAD，旨在比较TAD和NRN的性能增益。

v）最后，我们采用完整的NeRCo。由于对不同退化条件的改进鲁棒性和视觉友好的引导，这种设置实现了明显的性能增益。

我们在图7中报告了所有设置的定性结果。可以看到，在第一个样本中，“#1”增加了对比度，并在暗区域粗略恢复了物体，但增强的色调仍然不真实。“#2”在一定程度上减轻了色偏现象，但仍存在不受欢迎的遮蔽物。 “#3”和“#4” 生成了更清晰的结果，具有更真实的亮度。此外，NeRCo的结果保留了大部分细节，并表现出最佳的感知效果，特别是在绿色和红色框中的区域。

这归因于神经表示归一化和文本驱动的外观判别器，前者统一了退化水平并降低了增强任务的难度，而后者则引导了视觉友好的优化。

## 4.4结论

我们提出了一种隐式神经表示方法，用于协作低光照图像增强，命名为NeRCo，以无监督方式恢复视觉友好的结果。首先，对于输入的退化图像，我们采用神经表示来归一化退化水平（例如，暗亮度和自然噪声）。此外，对于输出的增强图像，我们配备了一个具有高频路径的判别器，并利用来自预训练视觉-语言模型的先验知识来提供感知导向的指导。最后，为了减轻对配对数据的依赖并以自监督方式增强低光照场景，我们开发了双闭环协作约束来训练增强模块。它鼓励所有组件相互约束，进一步减少解决方案空间。实验证明，与其他表现优异的方法相比，我们的方法具有优越性。所提出的组件为其他低级任务（如图像去雾、压缩感知和高光谱成像）提供了有价值的启示。

## 5. 结论

我们提出了一种隐式神经表示方法，用于协作低光照图像增强，命名为NeRCo，以无监督方式恢复视觉友好的结果。首先，对于输入的退化图像，我们采用神经表示来归一化退化水平（例如，暗亮度和自然噪声）。此外，对于输出的增强图像，我们配备了一个具有高频路径的判别器，并利用来自预训练视觉-语言模型的先验知识来提供感知导向的指导。最后，为了减轻对配对数据的依赖并以自监督方式增强低光照场景，我们开发了双闭环协作约束来训练增强模块。它鼓励所有组件相互约束，进一步减少解决方案空间。实验证明，与其他表现优异的方法相比，我们的方法具有优越性。所提出的组件为其他低级任务（如图像去雾、压缩感知和高光谱成像）提供了有价值的启示。

参考文献

[1] 凌云. PS7219显示驱动器及其在PLC中的应用[J]. 湖南冶金职业技术学院报, 2002, 28(3): 249-252.

[1] Yunpeng Bai, Chao Dong, 和 Cairong Wang. Ps-nerv: Patch-wise stylized neural representations for videos. arXiv preprint arXiv:2208.03742, 2022.

[2] Bolun Cai, Xianming Xu, Kailing Guo, Kui Jia, Bin Hu, 和 Dacheng Tao. A joint intrinsic-extrinsic prior model for retinex. In Proceedings of the IEEE International Conference on Computer Vision, 2017.

[3] Jianrui Cai, Shuhang Gu, 和 Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. IEEE Transactions on Image Processing, 2018.

[4] Hao Chen, Bo He, Hanyu Wang, Yixuan Ren, Ser Nam Lim, 和 Abhinav Shrivastava. Nerv: Neural representations for videos. Advances in Neural Information Processing Systems, 2021.

[5] Yinbo Chen, Sifei Liu, 和 Xiaolong Wang. Learning continuous image representation with local implicit image function. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2021.

[6] Xuxin Cheng, Qianqian Dong, Fengpeng Yue, Tom Ko, Mingxuan Wang, 和 Yuexian Zou. M3st: Mix at three levels for speech translation. In ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023.

[7] Xuxin Cheng, Zhihong Zhu, Hongxiang Li, Yaowei Li, 和 Yuexian Zou. Ssvmr: Saliency-based self-training for video-music retrieval. In ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023.

[8] Jaemin Cho, Jie Lei, Hao Tan, 和 Mohit Bansal. Unifying vision-and-language tasks via text generation. In Proceedings of International Conference on Machine Learning, 2021.

[9] Emilien Dupont, Yee Whye Teh, 和 Arnaud Doucet. Generative models as distributions of functions. In Proceedings of The 25th International Conference on Artificial Intelligence and Statistics, 2022.

[10] Xueyang Fu, Delu Zeng, Yue Huang, Xiao-Ping Zhang, 和 Xinghao Ding. A weighted variational model for simultaneous reflectance and illumination estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[11] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, 和 Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020.

[12] Xiaojie Guo, Yu Li, 和 Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. IEEE Transactions on Image Processing, 2017.

[13] Jiang Hai, Zhu Xuan, Ren Yang, Yutong Hao, Fengzhu Zou, Fang Lin, 和 Songchen Han. R2rnet: Low-light image enhancement via real-low to real-normal network. Journal of Visual Communication and Image Representation, 2023.

[14] Shijie Hao, Xu Han, Yanrong Guo, Xin Xu, 和 Meng Wang. Low-light image enhancement with semi-decoupled decomposition. IEEE Transactions on Multimedia, 2020.

[15] Md Jahidul Islam, Chelsey Edge, Yuyang Xiao, Peigen Luo, Muntaqim Mehtaz, Christopher Morse, Sadman Sakib Enan, 和 Junaed Sattar. Semantic segmentation of underwater imagery: Dataset and benchmark. In IEEE/RSJ International Conference on Intelligent Robots and Systems, 2020.

[16] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, 和 Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. IEEE Transactions on Image Processing, 2021.

[17] Chen Ju, Tengda Han, Kunhao Zheng, Ya Zhang, 和 Weidi Xie. Prompting visual-language models for efficient video understanding. In European Conference on Computer Vision, 2022.

[18] Diederik P. Kingma 和 Jimmy Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.

[19] Jaewon Lee, Kwang Pyo Choi, 和 Kyong Hwan Jin. Learning local implicit fourier representation for image warping. In European Conference on Computer Vision, 2022.

[20] Jaewon Lee 和 Kyong Hwan Jin. Local texture estimator for implicit representation function. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2022.

[21] Kanggeun Lee 和 Won-Ki Jeong. Iscl: Interdependent self-cooperative learning for unpaired image denoising. IEEE Transactions on Medical Imaging, 2021.

[22] Hongxiang Li, Meng Cao, Xuxin Cheng, Yaowei Li, Zhihong Zhu, 和 Yuexian Zou. G2l: Semantically aligned and uniform video grounding via geodesic and game theory. In Proceedings of the IEEE International Conference on Computer Vision, 2023.

[23] Weiqi Li, Bin Chen, 和 Jian Zhang. D3c2-net: Dual-domain deep convolutional coding network for compressive sensing. arXiv preprint arXiv:2207.13560, 2022.

[24] Yaowei Li, Bang Yang, Xuxin Cheng, Zhihong Zhu, Hongxiang Li, 和 Yuexian Zou. Unify, align and refine: Multi-level semantic alignment for radiology report generation. arXiv preprint arXiv:2303.15932, 2023.

[25] Risheng Liu, Long Ma, Jiaao Zhang, Xin Fan, 和 Zhongxuan Luo. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2021.

[26] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, 和 Alexander C Berg. Ssd: Single shot multibox detector. In European Conference on Computer Vision, 2016.

[27] Yun Liu, Zhongsheng Yan, Sixiang Chen, Tian Ye, Wenqi Ren, 和 Erkang Chen. Nighthazeformer: Single nighttime haze removal using prior query transformer. arXiv preprint arXiv:2305.09533, 2023.

[28] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, 和 Zhongxuan Luo. Toward fast, flexible, and robust low-light image enhancement. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2022.

[29] D. Martin, C. Fowlkes, D. Tal, 和 J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In Proceedings of the IEEE International Conference on Computer Vision, 2001.

[30] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, 和 Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In European Conference on Computer Vision, 2020.

[31] Anish Mittal, Rajiv Soundararajan, 和 Alan C. Bovik. Making a “completely blind” image quality analyzer. IEEE Signal Processing Letters, 2013.

[32] Nathan Moroney. Local color correction using non-linear masking. In Color and Imaging Conference, 2000.

[33] Michael K. Ng 和 Wei Wang. A total variation model for retinex. SIAM Journal on Imaging Sciences, 2011.

[34] Etta D Pisano, Shuquan Zong, Bradley M Hemminger, Marla DeLuca, R Eugene Johnston, Keith Muller, M Patricia Braeuning, 和 Stephen M Pizer. Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. Journal of Digital imaging, 1998.

[35] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, 和 Ilya Sutskever. Learning transferable visual models from natural language supervision. In Proceedings of International Conference on Machine Learning, 2021.

[36] Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, 和 Aaron Courville. On the spectral bias of neural networks. In Proceedings of International Conference on Machine Learning, 2019.

[37] Yongming Rao, Wenliang Zhao, Guangyi Chen, Yansong Tang, Zheng Zhu, Guan Huang, Jie Zhou, 和 Jiwen Lu. Denseclip: Language-guided dense prediction with context-aware prompting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022.

[38] Albert W. Reed, Hyojin Kim, Rushil Anirudh, K. Aditya Mohan, Kyle Champley, Jingu Kang, 和 Suren Jayasuriya. Dynamic ct reconstruction from limited views with

implicit neural representations and parametric motion fields. In Proceedings of the IEEE International Conference on Computer Vision, 2021.

[39] Yurui Ren, Zhenqiang Ying, Thomas H. Li, 和 Ge Li. Lecarm: Low-light image enhancement using the camera response model. IEEE Transactions on Circuits and Systems for Video Technology, 2019.

[40] Vishwanath Saragadam, Jasper Tan, Guha Balakrishnan, Richard G Baraniuk, 和 Ashok Veeraraghavan. Miner: Multiscale implicit neural representation. In European Conference on Computer Vision, 2022.

[41] Yu Sun, Jiaming Liu, Mingyang Xie, Brendt Egon Wohlberg, 和 Ulugbek S Kamilov. Coil: Coordinate-based internal learning for imaging inverse problems. IEEE Transactions on Computational Imaging, 2021.

[42] Matthew Tancik, Ben Mildenhall, Terrance Wang, Divi Schmidt, Pratul P. Srinivasan, Jonathan T. Barron, 和 Ren Ng. Learned initializations for optimizing coordinate-based neural representations. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2021.

[43] Zia ur Rahman, Daniel J. Jobson, 和 Glenn A. Woodell. Retinex processing for automatic image enhancement. Journal of Electronic Imaging, 2004.

[44] Shuhang Wang, Jin Zheng, Hai-Miao Hu, 和 Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. IEEE Transactions on Image Processing, 2013.

[45] Zhou Wang, A.C. Bovik, H.R. Sheikh, 和 E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 2004.

[46] Chen Wei, Wenjing Wang, Wenhan Yang, 和 Jiaying Liu. Deep retinex decomposition for low-light enhancement. arXiv preprint arXiv:1808.04560, 2018.

[47] Wenhui Wu, Jian Weng, Pingping Zhang, Xu Wang, Wenhan Yang, 和 Jianmin Jiang. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2022.

[48] Li Xu, Qiong Yan, Yang Xia, 和 Jiaya Jia. Structure extraction from texture via relative total variation. ACM Trans. Graph., 2012.

[49] Yuhang Yang, Wei Zhai, Hongchen Luo, Yang Cao, Jiebo Luo, 和 Zheng-Jun Zha. Grounding 3d object affordance from 2d interactions in images. In Proceedings of the IEEE International Conference on Computer Vision, 2023.

[50] Tian Ye, Yunchen Zhang, Mingchao Jiang, Liang Chen, Yun Liu, Sixiang Chen, 和 Erkang Chen. Perceiving and modeling density for image dehazing. In European Conference on Computer Vision, 2022.

[51] Xuanyu Zhang, Bin Chen, Wenzhen Zou, Shuai Liu, Yongbing Zhang, Ruiqin Xiong, 和 Jian Zhang. Progressive content-aware coded hyperspectral compressive imaging. arXiv preprint arXiv:2303.09773, 2023.

[52] Xuanyu Zhang, Yongbing Zhang, Ruiqin Xiong, Qilin Sun, 和 Jian Zhang. Herosnet: Hyperspectral explicable reconstruction and optimal sampling deep network for snapshot compressive imaging. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2022.

[53] Yu Zhang, Xiaoguang Di, Bin Zhang, 和 Chunhui Wang. Self-supervised image enhancement network: Training with low light images only. arXiv preprint arXiv:2002.11300, 2020.

[54] Yunfan Zhang, Ties van Rozendaal, Johann Brehmer, Markus Nagel, 和 Taco Cohen. Implicit neural video compression. In ICLR Workshop on Deep Generative Models for Highly Structured Data, 2022.

[55] Yonghua Zhang, Jiawan Zhang, 和 Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In Proceedings of the 27th ACM International Conference on Multimedia, 2019.

[56] Kaiyang Zhou, Jingkang Yang, Chen Change Loy, 和 Ziwei Liu. Learning to prompt for vision-language models. International Journal of Computer Vision, 2022.

隐式神经表示用于协作低光图像增强 - 补充文档

摘要

这是论文“隐式神经表示用于协作低光图像增强”的补充材料。首先，我们在A部分提供了更多关于NRN模块的结果以全面展示。此外，在B部分，我们使用不同的提示词训练框架并与其他消融设置进行比较，以验证文本驱动监督的效果。在C部分，我们进一步进行TAD的消融实验以确定每条路径的作用。在D部分，为了展示我们方法的语义优势，我们使用预训练的视觉语言模型对不同方法的增强结果进行分类并报告其准确性。我们的结果在高光图像的文本描述方面表现最佳。最后，在E部分，我们在三个著名的基准上展示了更多的定性分析，包括LSRW数据集、LOL数据集和LIME数据集。显然，所提出的NeRCo取得了最佳性能，进一步验证了我们的优越性。

## A. 归一化结果

深度学习模型通过学习将输入域的样本映射到目标域。然而，在实际应用中，退化条件各不相同。对于某些远离训练输入域的输入，训练模型很难表现出稳定的优异性能。因此，我们开发了神经表示归一化（NRN）模块以归一化不同条件，这在第3.2节中已说明。为了提供更有说服力的证据，我们在这里添加了更多的实验结果。

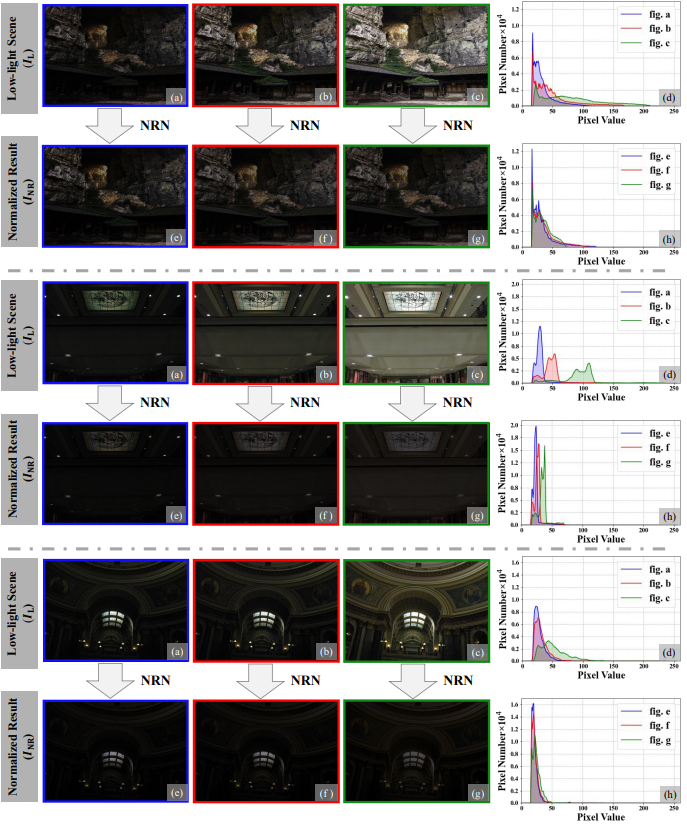
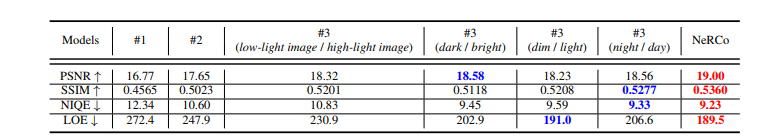
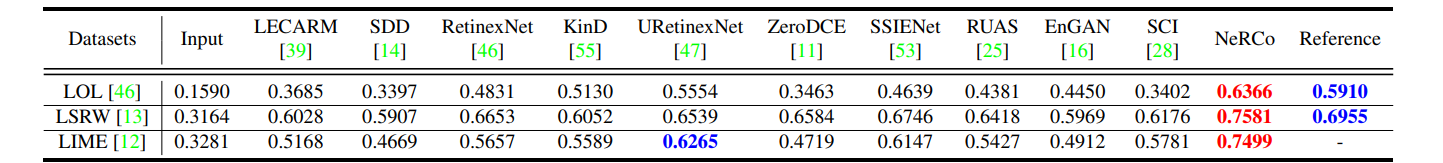
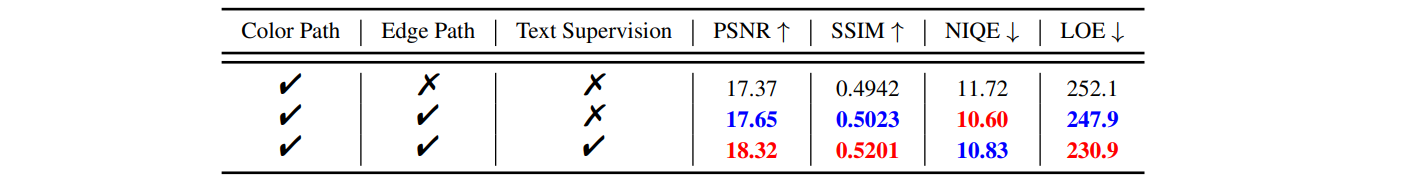
如图8所示，我们采用了SICE [3]数据集中的三组图像，每组包含三张亮度不同但内容相同的图像。我们分别用蓝色、红色和绿色线框将它们框起来。所有这些图像都经过我们的NRN模块处理，相应的结果也用相同颜色的框表示。可以看出，原始输入的亮度差异很大，而NRN的输出亮度则相似。为了更直观地展示，在右侧，我们提供了它们在Y通道上的像素分布的可视化。很明显，NRN收缩了亮度变化的范围并归一化了退化水平。

## B. 使用替代文本提示的实验

为了研究我们提出的文本驱动监督的贡献，我们比较了在不同提示词上训练的模型的性能。具体来说，我们考虑了三对替代提示词来指导模型训练：i) dark和bright。ii) dim和light。iii) night和day。这些替代实验是在第4.3节提到的“#3”消融设置上进行的，该设置从NeRCo中移除了神经表示函数。

如表3所示，我们报告了不同设置在LSRW数据集[13]上的结果。我们设计了不同的提示词来研究不同文本对模型性能的影响。可以看到，带有不同提示词的“#3”设置在所有四个指标上都实现了不错的分数。尽管它们的值有所不同，但它们在一个稳定的范围内，即优于其他消融设置且低于NeRCo。一方面，我们可以看到文本驱动的监督确实提升了性能。另一方面，这也间接证明了我们提出的神经表示归一化（NRN）模块的贡献。由于低光图像和高光图像是描述此任务中图像的常用文本，我们在其他实验中使用了这一对提示词以进行更直观的验证。

## C. TAD消融

TAD包含三个路径：颜色判别、边缘监督和文本驱动判别。为了进一步明确每个组件的作用，我们对TAD进行了消融研究，去除了提交论文中消融设置“#3”中的不同路径。注意，由于至少需要一种监督，我们采用颜色判别作为基础判别器。结果如表5所示，可以看到边缘路径和文本监督都提高了效果。

图表 10捕获的低光场景 (IL) 与 NRN (INR) 结果之间的比较。低光样本来自 SICE [3] 数据集。它包含大量图像集，每个图像集都有共同的内容和不同的退化条件。

表格 5 基于消融设置“#3”对 TAD 进行消融研究。我们在 LSRW 数据集上进行实验。最佳结果和第二佳结果分别以红色和蓝色突出显示

表格 4三个基准上不同设置的平均语义得分。最佳和第二佳结果分别以红色和蓝色突出显示。可以看出，预训练的视觉语言模型对我们的结果的分类比其他方法更准确，这表明我们的方法具有更好的语义一致性。

表格 3：使用不同提示选项的消融研究。最佳结果和第二佳结果分别以红色和蓝色突出显示。可以看到，使用提示训练的设置优于其他版本，并且提示可以用同义词替换。这证明了文本驱动的监督在模型性能方面有所提升

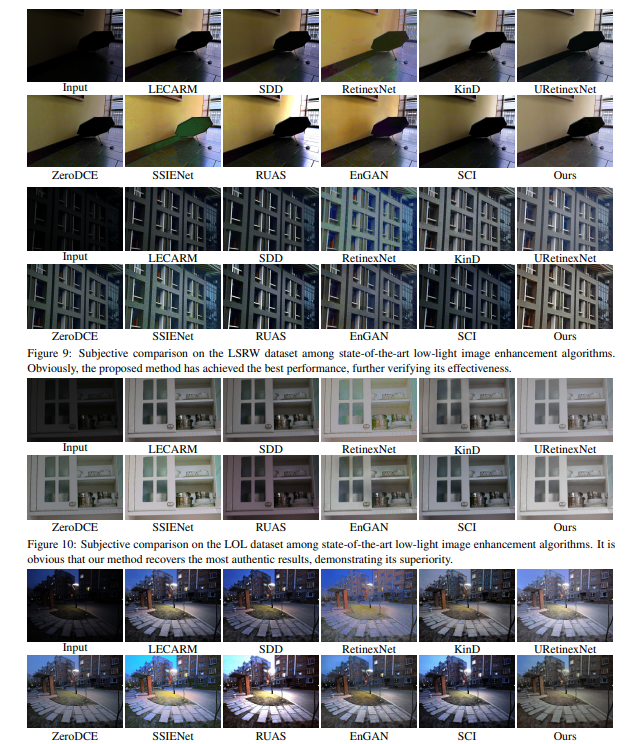
## D. 语义评估

为了展示我们的NeRCo在语义层面的优越性，我们采用预训练的CLIP模型[35]来计算不同方法的语义评分。具体来说，我们首先设计了一个提示词“high-light image”。然后，通过CLIP模型生成图像向量和文本向量。我们计算它们的余弦差异，并使用softmax函数获得语义评分，评分范围从0到1。较高的分数表示增强图像与提示词“high-light image”之间的语义一致性更好。

我们在表4中报告了不同方法结果的平均预测准确率。可以看到，预训练的CLIP认为输入的低光图像最不可能是高光图像，而高光参考图像被准确分类，除了LIME[12]仅包含退化场景。尽管某些方法输出的结果在语义上令人印象深刻，但我们的NeRCo获得了最高分，甚至超过了真实值。这证明了数据集中的参考图像在语义上是良好的，因为它们获得了第二高的分数。由于在训练期间的文本驱动判别，我们的方法比参考图像产生了更多的感知友好结果。

## E. 定性分析

我们在论文中提供了充分的定量结果（表1）。然而，由于空间限制，仅给出了部分视觉比较（图6）。在这里，我们补充了与其他SOTA方法的更多定性分析，包括LECARM[39]、SDD[14]、RetinexNet[46]、KinD [55]、URetinex-Net[47]、ZeroDCE[11]、SSIENet[53]、RUAS[25]、EnGAN[16]和SCI[28]。

图9展示了在LSRW数据集上的增强结果。可以看到，传统的基于模型的方法无法恢复足够的亮度，而其他比较方法则存在颜色失真问题。RetinexNet、KinD、ZeroDCE和RUAS等方法开发了后处理去噪操作以去除黑暗区域的固有噪声，但它们往往丢失细节。总体而言，我们的NeRCo能够进行颜色调整和细节保留，展示了其优越性。此外，我们还提供了我们提出的NeRCo与其他SOTA方法在其他著名基准上的视觉比较。图10展示了在LOL数据集上的比较，图11展示了在LIME数据集上的定性结果。显然，在所有这些比较中，我们的方法恢复了最真实的色调，并提供了视觉友好的结果，证明了其有效性。

图表 11：在 LIME 数据集上，最新低光图像增强算法的主观比较。我们的模型在这个低光图像数据集上仍然表现最佳，这证明了它的有效性