西北工业大学

数字图像处理–论文翻译

原论文标题：
FeatEnHancer: Enhancing Hierarchical Features for Object Detection and Beyond Under Low-Light Vision

秦紫玄

计算机学院

计算机科学与技术

2024 年 11 月

学号：2022300841

Northwestern Polytechnical University

FeatEnHancer: 增强低光照视觉下物体检测的层次特征

Khurram Azeem Hashmi[1,2], Goutham Kallempudi[2], Didier Stricker[1,2] and Muhammamd Zeshan Afzal[1,2]

[1]DFKI - German Research Center for Artificial Intelligence, [2]RPTU Kaiserslautern,
{khurram_azeem.hashmi, didier.stricker, muhammad_zeshan.afzal}@dfki.de, kallempu@rptu.de

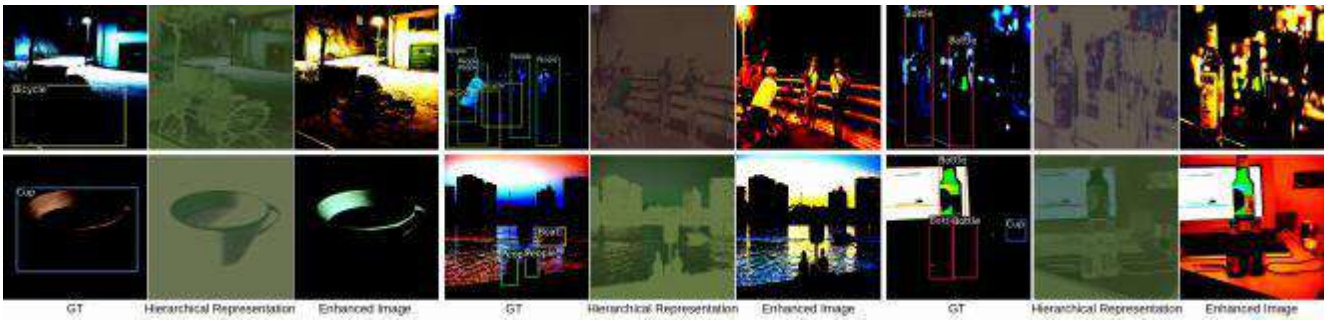

图1: 我们的FeatEnHancer学习到的层次化表示和增强后的图像。 我们在下游目标检测任务上训练我们的FeatEnHancer，并从验证集中可视化这些图像。这些映射和增强后的图像显示，尽管生成的图像在视觉上吸引力较小，但我们的模型增强了与任务相关的特征。最好在屏幕上查看。

摘要

在低光照条件下提取对下游任务有用的视觉线索尤其具有挑战性。先前的工作通过将视觉质量与机器感知能力相关联或设计需要在合成数据集上预训练的光照退化变换方法来创建增强表示。我们认为，优化与下游任务损失相关的增强图像表示可以产生更具表现力的表示。因此，在这项工作中，我们提出了一个新颖的模块，FeatEnHancer，它通过使用与任务相关的损失函数引导的多头注意力来分层组合多尺度特征，以创建合适的表示。此外，我们的内部尺度增强提高了每个尺度或级别提取的特征的质量，并以反映它们对当前任务的相对重要性的方式结合了不同尺度的特征。FeatEnHancer是一个通用的即插即用模块，可以集成到任何低光照视觉管道中。我们通过广泛的实验表明，使用FeatEnHancer产生的增强表示在几个低光照视觉任务中显著且一致地改善了结果，包括暗物体检测（在ExDark上+5.7 mAP）、面部检测（在DARK FACE上+1.5 mAP）、夜间语义分割（在ACDC上+5.1 mIoU）和视频物体检测（在DarkVision上+1.8 mAP），突出了在低光照条件下增强层次特征的有效性。

1. 介绍

最近在高级视觉任务中取得的显著进步表明，给定高质量的图像，当前的视觉主干网络[20, 15, 12, 32, 31]、物体检测器[42, 28, 43, 19, 2, 3, 49, 4, 71, 64, 65]和语义分割模型[34, 48, 57, 7, 58]能够有效地学习执行视觉任务所需的特征。同样，现代低光照图像增强（LLIE）方法[44, 67, 14, 21, 17, 25]能够将低光照图像转换为视觉友好的表示。然而，简单地将两者结合起来，在低光照条件下的高级视觉任务中带来的增益是次优的。

本项目探讨了LLIE与高级视觉方法结合性能低下的根本原因，并观察到以下限制：1）尽管现有的LLIE方法在人类视觉感知方面取得了突破，但由于缺乏多尺度特征，它们与视觉主干网络[20, 12, 15, 32, 31]不一致。例如，增强方法可能增加了某些区域的亮度，但同时也破坏了物体的边缘和纹理信息。2）不同低光照图像之间的像素分布可能因光照环境的差异而有很大差异[17, 25, 68]。这在某些情况下增加了类内差异（见图3，[17]只识别出一辆自行车，而不是地面真实情况中的两辆自行车）。3）当前的LLIE方法[14, 17, 25, 21, 44, 56, 67]采用增强损失函数来优化增强网络。这些损失函数迫使网络平等地关注所有像素，缺乏学习对高级下游视觉任务（如物体检测中的物体姿态和形状）必要的信息细节。此外，为了训练这些增强网络，它们中的大多数[44, 14, 67, 56]需要一组高质量的图像，这在现实世界环境中很难获得。

受到这些观察结果的启发，并受到最近LLIE[17, 25]和基于视觉的主干网络[15, 32, 31]的发展，本文旨在通过探索一个端到端可训练的方法来弥合这一差距，该方法在单个网络中联合优化增强和下游任务目标。为此，我们提出了FeatEnHancer，一个通用的特征增强器，它学习丰富有利于低光照环境中下游视觉任务的多尺度层次特征。图1中展示了学习到的层次表示和增强图像的示例。

特别地，我们的FeatEnHancer首先对低光照RGB输入图像进行下采样以构建多尺度层次表示。随后，这些表示被送入我们的特征增强网络（FEN），这是一个深度卷积网络，用于增强内部尺度的语义表示。请注意，FEN的参数可以通过与任务相关的损失函数进行调整，这推动了FEN只增强与任务相关的特征。这种多尺度学习允许网络从更高和更低分辨率的特征中增强全局和局部信息。一旦获得了不同尺度上的增强表示，剩下的障碍就是有效地融合它们。为了实现这一点，我们选择了两种不同的策略来捕获来自更高和更低分辨率特征的全局和局部信息。首先，为了合并高分辨率特征，受到[50]中多头注意力的启发，我们设计了一种尺度感知注意力特征聚合（SAFA）方法，它共同关注不同尺度的信息。其次，对于低分辨率特征，采用了跳跃连接[20]方案将从SAFA增强的表示与低分辨率特征合并。有了这些共同学习到的层次特征，我们的FeatEnHancer提供了可以被高级方法（如特征金字塔网络[27]用于物体检测[43]和实例分割[19]，或UNet[45]用于语义分割[34]）利用的语义强大的表示。

这项工作的主要贡献可以总结如下：

1. 我们提出了FeatEnHancer，一个新颖的模块，它增强层次特征以提升低光照条件下的下游视觉任务。我们的内部尺度特征增强和尺度感知注意力特征聚合方案与视觉主干网络一致，并产生强大的语义表示。FeatEnHancer是一个通用的即插即用模块，可以与任何高级视觉任务一起端到端训练。

2. 据我们所知，这是第一个在低光照场景中充分利用多尺度层次特征并将这些特征推广到多个下游视觉任务（如物体检测、语义分割和视频物体检测）的工作。

3. 在四个不同的下游视觉任务(包括图像和视频)上进行的大量实验表明，我们的方法相对于基线、LLIE方法和任务特定最先进的方法上都带来了一致且显著的改进。

2. 相关工作

2.1. 增强低光照图像

基于深度学习的LLIE方法专注于改善低光照图像的视觉质量，以满足人类视觉感知[23, 22]。大多数LLIE方法[14, 44, 56, 67]在监督学习范式下操作，在训练期间需要成对的数据。无监督的基于GAN的方法[21]消除了训练过程中成对数据的需要。然而，它们的性能依赖于对未成对数据的仔细选择。最近，零参考方法[17，25，68]通过设计一组非参考损失函数来增强低光照图像，无需成对或未配对数据。受这些最新发展的启发，这项工作旨在通过增强多尺度层次特征来弥合低光照增强和下游视觉任务（如物体检测[10, 33, 62]、语义分割[60, 47]和视频物体检测[63]）之间的差距，无需配对数据即可提升性能。

2.2. 增强低光照以改善下游视觉任务

这些方法将机器感知作为成功的标准，同时增强图像以改善下游视觉任务。实现这一目标的一个明显方法是将LLIE方法作为初始步骤[70, 17]。然而，这导致了不尽如人意的结果(见表2、4和5)。近期，另一项工作探索了端到端的管道，优化增强和个别任务，我们的工作也遵循了同样的精神。

面部检测. Liang等人[26]提出了一种通过利用多重曝光生成从弱光图像中提取信息的有效方案。此外，提出了双向域自适应[52，51]和联合执行增强和检测的并行架构[37]来推进研究。然而，这些方法专门设计用于面部检测[62, 53]，当应用于通用物体检测[51]时只能提供微小的改进。相反，我们的FeatEnHancer是一个通用模块，它显著改善了几个下游视觉任务。因此，我们避免将我们的方法与仅针对面部检测评估的架构进行比较。

暗物体检测. 由于真实世界的低照明数据集[33, 39]的出现，暗（低光照）物体检测方法最近出现了[10, 30]。IA-YOLO [30]引入了基于卷积神经网络(CNN)的参数预测器，该参数预测器学习差分图像处理模块中采用的滤波器的最优配置。与我们的工作最相关的是MAET [10]，它研究了低照度下的物理噪声模型和图像信号处理(ISP)管道，并学习该模型以预测退化参数和对象特征。为了避免特征纠缠，他们采用正交切线规则来分离物体之间的余弦相似性和退化特征。然而，由于[30]中的特定天气超参数和[10]中的退化参数，这些工作依赖于大型合成数据集来实现期望的性能。与它们不同，我们的FeatEnHancer是从任务相关的损失函数优化而来的，不需要任何预训练，也不需要模仿低光照或恶劣天气条件的合成数据集。

其他高级视觉任务. 除了面部和物体检测，最近的研究还探索了高级计算机视觉任务，如语义分割[6，34]。Xue等人[60]设计了一种对比学习策略，以同时改善视觉和机器感知，在具有对应关系的不利条件数据集(ACDC)的夜间语义分割数据集上实现了令人印象深刻的性能[47]。此外，黑暗视觉[63]最近出现，以解决弱光视觉下的视频对象检测。在这项工作中，由于[47，63]，我们应用了FeatEnHancer对弱光视觉下的语义分割和视频对象检测进行研究，考察其泛化能力。

2.3. 学习多尺度层次特征

以不同的比例表示物体是计算机视觉的主要困难之一。因此，这一领域的工作可以追溯到手工设计特征的时代[36，11，38，24]。现代物体探测器[43，28，2，71，49，40，65]利用多尺度特性来应对这一挑战。类似地，多尺度表示[34]和金字塔池化方案[69]已经被提出用于有效的语义分割。此外，当前基于视觉的主干网络[15, 31, 32]的改进表明，在特征提取过程中直接学习层次特征可以提升下游视觉任务[19, 57, 2]。然而，CNN的多尺度和层次结构尚未在低光照视觉任务中得到充分探索。

在恶劣的天气条件下，DENet [41]采用拉普拉斯金字塔[1]将图像分解为低频和高频分量用于对象检测。尽管结果令人鼓舞，但DENet中的多尺度特征学习依赖于拉普拉斯金字塔，这容易受噪声影响，并且可能在具有高对比度或尖锐边缘的区域中产生不一致性。与现代视觉主干网络[27，32，31]中的多尺度学习相一致，我们的FeatEnHancer采用CNN来生成多尺度特征表示，这些特征表示通过尺度感知注意力特征聚合和跳跃连接来融合。我们的方法更加灵活，并与下游视觉任务对齐，提升了多个下游视觉任务的最新成果。

3. 提出的方法

本文的核心思想是设计一个增强弱光视觉下机器感知的通用可插拔模块，以解决多个下游视觉任务，如对象检测、语义分割和视频对象检测。FeatEnHancer的整体架构如图2所示。我们的FeatEnHancer以低光照图像作为输入，通过丰富任务相关的层次特征来自适应地增强其语义表示。现在我们详细讨论FeatEnHancer的关键组件

3.1. 分级特征增强

受基于视觉的主干网络[15，31，32]最近改进的启发，我们通过联合优化特征增强和下游任务，在低光照视觉下引入层次特征增强。与[15，31，32]不同，我们的目标是从低光照图像中提取空间特征，并生成有意义的语义表示。为了增强层次特征，我们首先从低光照输入图像构建多尺度表示。 然后，我们将这些多尺度表示送入我们的特征增强网络。

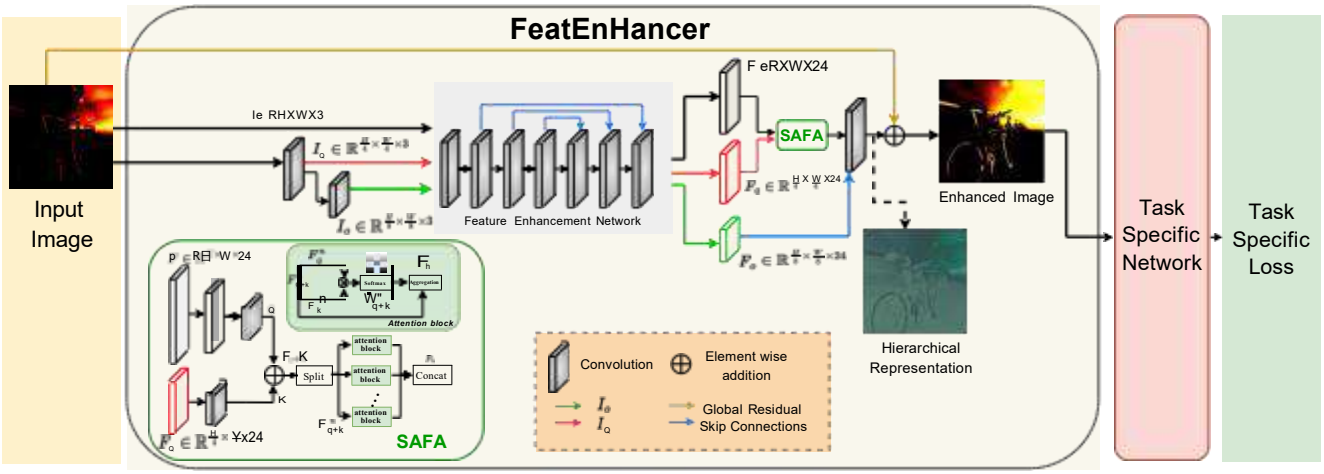

图2：提出的FeatEnHancer在下游视觉任务中使用的网络架构 我们的FeatEnHancer接收一个低光照图像，并通过丰富与任务相关的层次特征来适应性地增强其语义表示。放大图片以获得最佳视图。

构建多尺度表示. 我们以低光照RGB图像 $I \in R^{H\times W\times C}$ 作为输入，并使用常规卷积算子Conv(.)对I进行操作，生成 $I_q \in R^{H/4\times W/4\times 3}$ 和 $I_O \in R^{H/8\times W/8\times 3}$,分别代表输入图像的四分之一和八分之一尺度。简而言之，可以表示为：

$$I_q = Conv(I; K=7, S=4)$$
$$I_o = Conv(I_q; K=3, S=2)$$

其中KK和SS分别表示卷积核大小和步长，H,W,H,W,和CC分别代表图像的高度、宽度和通道数。

特征增强网络. 为了在每个尺度上增强特征，我们需要一个增强网络，该网络学习增强对下游任务重要的空间信息。受低光照图像增强网络[17, 25]的启发，我们设计了一个全卷积的内部尺度特征提取网络（FEN）。然而，与[17, 25]不同，我们的FEN在开始时引入了一个单独的卷积层，该层生成一个特征图 $F \in R^{H\times W\times C}$ ，其中C从3变换到32，同时保持分辨率（H×W）与输入相同。然后应用一系列六个卷积层，每个卷积层都伴随着ReLU激活函数，每个卷积层都具有K=3和S=1。我们在每个尺度I,Iq,和Io上分别应用FEN，获得多尺度特征表示，分别记为F,Fq,和Fo。这种多尺度学习使得网络能够从更高和更低分辨率的特征中增强全局和局部信息。因此，我们忽略了下采样和批量归一化，以保持邻近像素之间的语义关系，这与[17]类似。然而，我们在我们的FEN中丢弃了DCENet[17]的最后一个卷积层，并将每个尺度的最终增强特征表示传播出去，用于多尺度特征融合。请注意，FEN在FeatEnHancer中的实现细节与所提出的模块是独立的，甚至可以应用更高级的图像增强网络，如[68]，以提高性能。现在，我们将详细讨论多尺度特征融合。

3.2.多尺度特征融合

由于我们已经从特征增强网络（FEN）获得了多尺度特征表示（F, Fq, 和 Fo），剩下的挑战是如何有效地融合它们。低尺度特征（Fo）包含了细节和边缘信息。相比之下，高分辨率特征（Fq）捕获了更抽象的信息，如形状和模式。因此，简单的聚合会导致性能不佳（见表6a）。因此，我们采用了两种不同的策略来捕获更高和更低分辨率特征的全局和局部信息。首先，受到[50]中多头注意力的启发，这种注意力机制使得网络能够联合学习不同通道的信息，我们设计了一个尺度感知注意力特征聚合（SAFA）模块，它能够同时关注不同尺度的特征。其次，我们采用了跳跃连接（SC）方案，以整合来自Fo的低层信息和来自SAFA的增强表示，从而获得最终的增强层次表示。采用SAFA来合并高分辨率特征和SC来处理低分辨率特征，可以导致更强大的层次化表示（见表6b）。现在，我们将详细讨论SAFA。

尺度感知注意力特征聚合（SAFA）. 尽管高分辨率特征有助于捕获细节，例如识别小物体，

但对它们应用注意力操作计算成本很高。因此，在SAFA中，我们提出了一个高效的多尺度聚合策略，其中增强的高分辨率层次特征在注意力特征聚合之前被投影到更小的分辨率。如图2所示，SAFA将F转换为Q，Fq转换为K，然后Q和K被连接形成一组层次特征Fq+k，它们沿通道维度C被分割成N个块：

$$F_{q+k}^{n} = F_{q+k}[:,:,(n-1)\frac{C}{N}:n\frac{C}{N}]$$

其中n ∈ {1, 2, ..., N}，N是注意力块的总数。$F_{q+k}^{n} \in R^{\frac{H}{8}\times\frac{W}{8}\times\frac{C}{N}}$用于计算单个注意力块中的注意力权重W：

$$W_{q+k}^{n} = F_{q}^{n} \cdot F_{k}^{n}$$

$$\overline{W}_{q+k}^{n} = \frac{\exp(W_{q+k}^{n})}{\sum_{l-1}^{L}\exp(W_{q+k}^{n})}$$

其中W_{q+k}^{n}是F_{q}^{n}和F_{k}^{n}对于第n个块的注意力权重，而$\overline{W}_{q+k}^{n}$是W_{q+k}^{n}的归一化形式。根据第n个块的归一化注意力权重，我们应用加权求和来计算第n个块的增强层次表示$\overline{F}_{h}^{n} \in R^{\frac{H}{8}\times\frac{W}{8}\times\frac{C}{N}}$：

$$\overline{F}_{h}^{n} = \sum_{l=1}^{L}\overline{W}_{q+k}^{n} \cdot F_{q+k}^{n}$$

现在我们将所有$\overline{F}_{h}^{n}$沿通道维度连接起来，得到$\overline{F}_{h}$。注意，尽管$\overline{F}_{h}$与Q和K大小相同，但它包含了更丰富的表示，包含了多尺度高分辨率特征的信息。随后，如前所述，借助跳跃连接（SC），我们将Fo和$\overline{F}_{h}$整合起来，获得最终的增强层次表示，涵盖了全局和局部特征，如图1和2所示。请注意，在跳跃连接之前，我们对$\overline{F}_{h}$和Fo进行上采样，其中上采样操作是通过简单的双线性插值执行的，这比使用转置卷积要快得多。与现有工作中的图像增强不同，通过多尺度层次特征增强策略，我们的FeatEnHancer学习了强大的语义表示，捕获了局部和全局特征。这使其成为一个通用模块，增强层次特征，提升低光照视觉下的机器感知能力。

Dataset	Task	#Cls	#Train	#Val
ExDark [33]	Dark object detection	12	4800	2563
DARK FACE [62]	Face detection	1	5400	600
ACDC Nightime [47]	Semantic segmentation	19	400	106
DarkVision [63]	Video object detection	4	26	6

表 1: 用于报告四个不同下游视觉任务结果的数据集统计信息。#Cls表示类别数量，而#Train和#Val分别表示每个数据集的训练样本和验证样本数量。

4. 实验

我们对提出的 FeatEnHancer 模块进行了广泛的实验评估，涵盖了多个下游任务，包括通用物体检测[33,39]、面部检测[62]、语义分割[47]和视频物体检测[63]。表1总结了所使用的数据集的关键统计信息。本节首先将所提方法与强大的基线、现有的LLIE方法和特定任务的最新方法进行比较。然后，我们将对FeatEnHancer的重要设计选择进行消融研究。我们在附录A中提供了每个实验的完整实现细节。

4.1. 暗物体检测

设置. 对于真实世界数据上的暗物体检测实验，我们考虑了专门的暗（ExDark）[33]数据集（见表1）。我们采用RetinaNet[28]作为典型的检测器，Featurized Query R-CNN[65]作为高级物体检测框架来报告结果。在这两种检测器的情况下，都在COCO[29]数据集上预训练的模型在每个数据集上进行微调。对于RetinaNet，图像被调整到640×640，我们使用mmdetection[5]中的1×调度（使用SGD优化器[46]训练12个周期，初始学习率为0.001）。对于Featurized Query R-CNN，我们采用多尺度训练[4,49,65]（短边从400到800不等，长边为1333）。FQ R-CNN使用ADAMW[35]优化器训练50000次迭代（初始学习率为0.0000025，权重衰减为0.0001，批量大小为8）。请注意，对于每个物体检测框架，我们都采用相同的设置来复现我们工作的结果、基线、LLIE方法和特定任务的最新方法。

我们将FeatEnHancer与几种最先进的LLIE方法进行比较，包括KIND[67]、RAUS[44]、EnGAN[21]、MBLLEN[14]、Zero-DCE[17]、ZeroDCE++[17]和最先进的暗物体检测方法MAET[10]。对于LLIE方法，所有图像都从它们发布的检查点增强并传递给检测器。在MAET[10]的情况下，我们使用他们提出的退化管道预训练检测器，然后将其微调在两个数据集上以建立直接比较。

Methods	RetinaNet		FQ R-CNN	
	mAP50	mAP	mAP50	mAP
Baseline	72.1	46.3	74.5	47.0
RAUS [44]	64.7	44.0	77.0	48.1
KIND [67]	70.7	45.1	80.5	51.5
Zero-DCE++ [25]	70.3	45.2	79.5	49.2
EnGAN [21]	70.4	44.9	80.0	51.9
MBLLEN [14]	70.6	45.1	80.0	51.0
Zero-DCE [17]	71.0	45.2	80.6	52.0
MAET [10]	71.8	45.7	81.6	52.4
FeatEnHancer	**72.6**	**46.4**	**86.3**	**56.5**

表2： 在ExDark数据集上的定量比较. 突出显示了在常用评估指标上获得的结果。我们的FeatEnHancer带来了一致的改进，并在FQ R-CNN上取得了新的最先进结果。

Methods	RetinaNet		FQ R-CNN	
	AP50	AP	AP50	AP
Baseline	47.3	19.9	67.5	28.6
RAUS [44]	42.1	17.6	65.5	27.4
KIND [67]	47.2	19.8	65.0	27.5
Zero-DCE++ [25]	47.3	20.1	66.2	28.2
EnGAN [21]	45.1	19.3	67.4	28.4
MBLLEN [14]	47.1	19.8	67.3	27.1
Zero-DCE [17]	**47.4**	**20.1**	66.9	27.5
MAET [10]	44.3	18.7	66.1	27.1
FeatEnHancer	47.2	19.9	**69.0**	**29.4**

表 3: 在 DARKFACE 数据集上比较 FeatEnHancer. 使用 RetinaNet 时，FeatEnHancer与其他方法表现相当。然而，使用FQ R-CNN时，FeatEnHancer超越了所有其他方法。

Method	mIoU
Baseline [7]	45.7
RetinexNet [54]	41.9
DRBN [59]	43.3
FIDE [61]	43.4
KIND [67]	43.0
EnGAN [21]	43.8
ZeroDCE [17]	43.4
SSIENet [66]	41.4
Xue *et al.* [60]	49.8
FeatEnHancer	**54.9**

表4： 在ACDC数据集上的定量比较. 我们的FeatEnHancer带来了巨大的性能提升，达到了新的最先进结果。

ExDark上的结果. 表2 列出了LLIE作品、MAET以及所提出方法在两种物体检测框架上的结果。显然，我们的FeatEnHancer相较于之前的方法带来了一致且显著的性能提升。请注意，尽管MAET和我们的方法在RetinaNet上的性能相当（大约72 AP_{50}），但在FQ R-CNN上，所提出的FeatEnHancer以显著的优势超越了MAET，达到了新的最高水平AP_{50}为86.3。此外，图3 展示了我们的方法和两个最佳竞争者使用FQ R-CNN作为检测器的四个检测示例。这些结果说明，尽管视觉质量较差，我们的FeatEnHancer增强了有利于暗物体检测的层次特征，产生了最先进的结果。

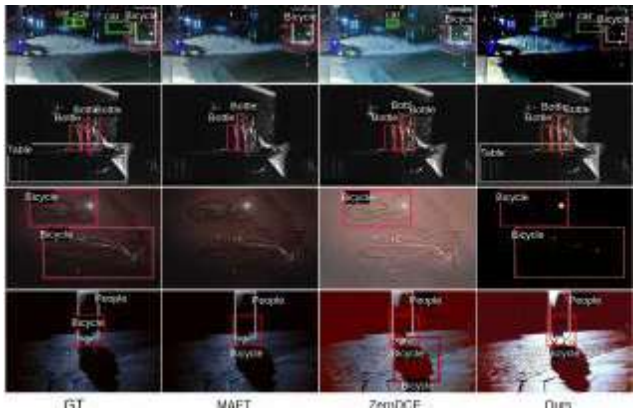

图3：在ExDark数据集上FeatEnHancer与前两个最佳竞争者的视觉比较。放大图片以获得最佳视图。

4.2. DARK FACE上的面部检测

设置. DARK FACE[53, 62]是一个为UG²比赛发布的具有挑战性的面部检测数据集。在DARK FACE数据集（见表1）上的实验中，所有方法的图像都被调整为更大的分辨率1500×1000。我们采用了相同的物体检测框架RetinaNet和FQ R-CNN，并遵循第4.1节中解释的相同的实验设置。

结果. FeatEnHancer、MAET以及六种LLIE方法使用RetinaNet和Featurized Query R-CNN的性能总结在表3中。请注意，一些LLIE方法[17,25,67]在RetinaNet的情况下比我们的方法取得了更好的结果。我们认为，由于DARK FACE数据集中的面部非常小且图像非常暗，RetinaNet甚至无法从增强的层次特征中捕获信息。我们在附录B中讨论了这种行为的一个例子。另一方面，LLIE方法直接提供了更明亮的图像，在这种情况下带来了轻微的增益（+0.1mAP_{50}）。然而，请注意，使用更强大的检测器，我们的FeatEnHancer超越了所有LLIE方法和MAET，显著提高了性能（+1.5 mAP_{50}），达到了69.0 mAP_{50}。

4.3. ACDC上的夜间语义分割

设置. 我们使用ACDC数据集[47]（见表1）中的夜间图像来报告在低光照环境下的语义分割结果。DeepLabV3+[7]作为语义分割的基线模型，从mmseg[8]中选取，以便与同期工作[60]进行直接比较。我们遵循[60]中相同的实验设置。完整的实现细节请参考附录A。

结果. 我们将我们的方法与几种最先进的LLIE方法进行了比较，包括RetinaNet[54]、KIND[67]、FIDE[61]、DRBN[59]、EnGAN[21]、SSIENet[66]、ZeroDCE[17]，

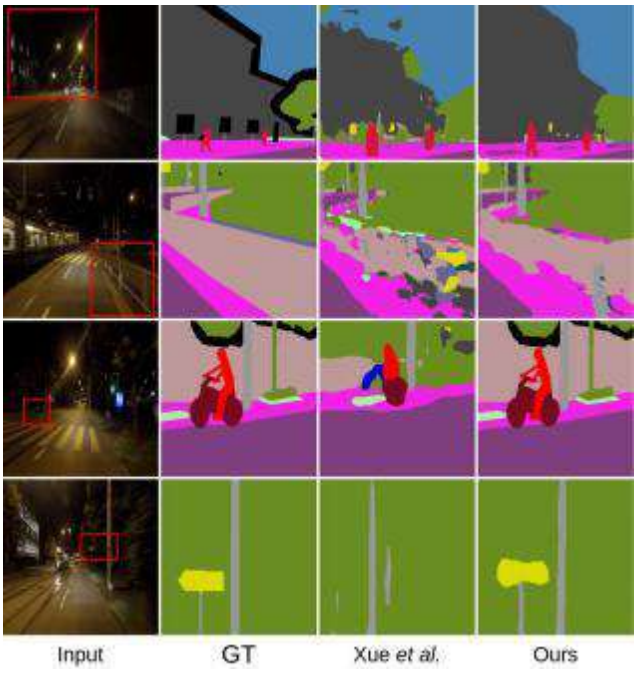

图4： 在ACDC夜间语义分割任务上，FeatEnHancer与之前最佳工作[60]的定性比较。FeatEnHancer提供了更准确的分割。

Method	Illumination (3.2) mAP	Illumination (0.2) mAP
Baseline [55]	32.8	10.4
RAUS [44]	7.42	5.19
EnGAN [21]	7.83	5.41
MBLLEN [14]	7.82	5.39
KIND [67]	7.43	5.25
Zero-DCE++ [25]	7.51	5.02
Zero-DCE [17]	7.83	5.43
FeatEnHancer	**34.6**	**11.2**

表5： 在DarkVision数据集上比较FeatEnHancer与LLIE方法。FeatEnHancer是唯一一种在两个照明水平上都能提升强大基线方法性能的方法。

.

以及当前最先进的夜间语义分割方法Xue等人的方法[60]。如表4所示，我们的FeatEnHancer在基线上带来了显著的改进，mIoU达到了54.9，超过了之前最好的结果5.1个百分点。此外，我们在图4中展示了与之前最佳竞争者[60]的定性比较。显然，我们的FeatEnHancer为更大和更小的物体（如最后一行中的地形和交通标志）提供了更准确的分割。这些结果证实了FeatEnHancer作为一个通用模块的有效性，它在暗物体检测和夜间语义分割中都取得了最新的结果。

4.4. DarkVision上的视频物体检测

设置. 我们将实验从静态图像扩展到视频领域，以测试我们方法的泛化能力。在低光照视觉下的视频物体检测在最近出现DarkVision数据集[63]上进行评估（见表1以获取数据集详细信息）。尽管数据集尚未公开，但我们感谢[63]的作者提供及时访问。为了在低光照设置下评估我们的FeatEnHancer，我们采用了低端相机分割的两个不同光照水平，即0.2和3.2。为了消融研究，我们采用了3.2%的光照水平分割。我们将SELSA[55]作为我们的基线，并遵循mmtracking[9]中的ResNet-50主干网络的相同实验设置。为了直接比较，我们首先通过LLIE方法增强所有视频帧，然后将这些帧输入到基线中，如第4.1节中所做的。作为视频物体检测的常见做法[16,18,55]，mAP@IoU=0.5被用作评估指标来报告结果。更多细节请参见附录A。

结果. 表5比较了我们的FeatEnHancer与几种LLIE方法[44,21,14,67,17,25]和强大的视频物体检测基线[55]。显然，我们的FeatEnHancer为基线提供了相当大的增益，在3.2和0.2光照水平下的mAP分别为34.6和11.2。请注意，我们的FeatEnHancer是唯一一种在图像和视频模态下都能提升性能的方法。相比之下，如表5所示，现有的LLIE方法不仅未能协助基线方法，还恶化了性能。LLIE方法的这种较差的泛化性突显了从特定领域成对数据[14, 67, 44]、未配对数据[21]和无数据曲线估计[17, 25]中学习并不是通用增强方法的最优解决方案。因此，需要更多的研究。

4.5. 消融研究

本节将对FeatEnHancer的关键设计选择进行消融研究，这些研究涉及到RetinaNet、DeepLabV3+和SELSA在ExDark（暗物体检测）、ACDC（夜间语义分割）和DarkVision在3.2%光照水平（视频物体检测）上的应用。

SAFA在FeatEnHancer中的作用. 提出的FeatEnHancer中的一个重要组成部分是尺度感知注意力特征聚合（SAFA），它用于聚合高分辨率特征。为了验证其有效性，我们进行了一系列实验，其中SAFA被替换为简单的平均值或跳跃连接（SC）[20]来融合增强的多尺度特征F和Fq。

Method	ExDark (mAP)	ACDC (mIoU)	DarkVision (mAP)
simple averaging	69.5	50.3	32.9
skip connections [20]	70.3	51.7	33.1
SAFA	72.6	54.9	34.6

(a) SAFA的有效性.

Method	ExDark (mAP)	ACDC (mIoU)	DarkVision (mAP)
SC , SC	69.7	51.7	32.8
SAFA , SAFA	70.2	52.6	33.4
SC, SAFA	70.9	52.9	33.8
SAFA, SC	72.6	54.9	34.6

(b) 多尺度融合的各种组合.

Method	ExDark (mAP)	ACDC (mIoU)	DarkVision (mAP)
maxpool	69.3	51.3	32.9
adavgpool [68]	69.9	50.7	32.9
interpolation [30]	70.7	51.5	33.1
Convolution	72.6	54.9	34.6

(c) 下采样技术.

Scale	ExDark (mAP)	ACDC (mIoU)	DarkVision (mAP)
(2, 4)	71.8	52.7	34.1
(4, 8)	72.6	54.9	34.6
(4, 16)	71.5	51.4	33.9
(8, 16)	68.7	45.6	31.9

(d) Iq 和Io各自不同的尺度.

N	ExDark (mAP)	ACDC (mIoU)	DarkVision (mAP)
2	72.1	53.9	34.2
4	72.4	54.3	34.5
8	72.6	54.9	34.6
12	72.4	54.3	34.1

(e) # SAFA的注意力块数量.

表6：在三个基准测试上对提出的FeatEnHancer进行消融研究. (a) 我们通过用不同的聚合方法替换SAFA来融合F和Fq，以研究其有效性。(b) 我们尝试了SAFA和跳跃连接（SC）的各种组合，以证明最优的设计选择。在这里，(SC, SC)意味着仅使用跳跃连接来合并Fq和Fo与F。(c) 除了卷积，我们还尝试了其他下采样技术来生成更低分辨率的表示。在这里，adavgpool表示在[68]中所做的自适应平均池化。(d) 我们改变尺度大小以生成更低尺度的表示。在这里，(2, 4)意味着 $I_q \in R^{\frac{H}{2}\times\frac{W}{2}\times 3}$ 和 $I_o = R^{\frac{H}{4}\times\frac{W}{4}\times 3}$ (e) 我们改变FeatEnHancer中SAFA的注意力块数量N。默认设置被突出显示。

实验结果总结在表6a中。显然，SAFA在ExDark上比平均值和SC策略分别高出+2.3 mAP，在ACDC上高出+3.2 mIoU，在DarkVision上高出+1.5 mAP。这些显著的提升表明，尺度感知注意力在FeatEnHancer中实现了最佳的多尺度特征聚合。

多尺度特征融合. 我们尝试了SAFA和SC的各种组合，以找到融合Fq和Fo与F的最佳设计选择。如表6b所示，当SAFA首先应用于融合F和Fq，然后使用跳跃连接将Fo与SAFA的输出合并时，性能有明显提升，达到了ExDark上的72.6 mAP，ACDC上的54.9 mIoU和DarkVision上的34.6 mAP。因此，我们将这种方法作为默认设置。

卷积下采样. 表6c总结了在输入图像I上应用的不同下采样技术的结果，以生成更低分辨率的Iq和Io（见第3.1节）。我们提出的卷积下采样与最大池化、自适应平均池化[68]和双线性插值[30]相比，在ExDark上获得了+1.9 mAP的显著提升，在DarkVision上获得了+3.4 mIoU和+1.5 mAP的提升。这些结果证明了卷积下采样的有效性，因为它与各种视觉主干网络[32, 15, 27]更为一致。

不同尺度大小. 我们在表6d中分析了生成低分辨率时不同尺度大小的影响。例如，(2, 4)意味着输入图像I的分辨率被降低2倍和4倍，以生成Iq和Io。所有这些尺度都是通过常规卷积算子Conv(.)生成的，如方程1中所解释的。观察表6d中的结果，所有三个任务的最佳性能是在尺度大小为(4, 8)时实现的，因此被选为默认设置。

SAFA中注意力块的数量. 表6e研究了我们SAFA中注意力块数量N的影响。随着N的增加，所有三个任务的性能都有所提升。这表明SAFA中更多的注意力块可以带来额外的增益。当N达到8时，实现了最佳性能，ExDark上的72.6 mAP，ACDC上的54.9 mIoU，以及DarkVision上的34.6 mAP，之后性能趋于饱和。因此，N=8被用作默认设置。

5. 结论

本文提出了FeatEnHancer，这是一个新颖的通用特征增强模块，旨在丰富低光照条件下有利于下游任务的层次特征。我们的内部尺度特征增强和尺度感知注意力特征聚合方案与视觉主干网络一致，并产生强大的语义表示。此外，我们的FeatEnHancer不需要在合成数据集上预训练，也不依赖于增强损失函数。这些架构创新使FeatEnHancer成为一个即插即用模块。在涵盖图像和视频的四个不同下游视觉任务上的广泛实验表明，我们的方法在基线、LLIE方法和任务特定最先进方法上都带来了一致且显著的改进。

参考

[1] Peter J Burt and Edward H Adelson. The laplacian pyramid as a compact image code. In *Readings in computer vision*, pages 671–679. Elsevier, 1987.

[2] Zhaowei Cai and Nuno Vasconcelos. Cascade R-CNN: delving into high quality object detection. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pages 6154–6162. Computer Vision Foundation / IEEE Computer Society, 2018.

[3] Zhaowei Cai and Nuno Vasconcelos. Cascade r-cnn: High quality object detection and instance segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(5):1483–1498, 2021.

[4] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part I*, volume 12346 of *Lecture Notes in Computer Science*, pages 213–229. Springer, 2020.

[5] Kai Chen, Jiaqi Wang, Jiangmiao Pang, Yuhang Cao, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jiarui Xu, Zheng Zhang, Dazhi Cheng, Chenchen Zhu, Tianheng Cheng, Qijie Zhao, Buyu Li, Xin Lu, Rui Zhu, Yue Wu, Jifeng Dai, Jingdong Wang, Jianping Shi, Wanli Ouyang, Chen Change Loy, and Dahua Lin. MMDetection: Open mmlab detection toolbox and benchmark. *arXiv preprint arXiv:1906.07155*, 2019.

[6] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs, 2014.

[7] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

[8] MMSegmentation Contributors. MMSegmentation: Openmmlab semantic segmentation toolbox and benchmark. https://github .com/open-mmlab/mmsegmentation, 2020.

[9] MMTracking Contributors. MMTracking: OpenMMLab video perception toolbox and benchmark. https://github.com/open-mmlab/mmtracking, 2020.

[10] Ziteng Cui, Guo-Jun Qi, Lin Gu, Shaodi You, Zenghui Zhang, and Tatsuya Harada. Multitask aet with orthogonal tangent regularity for dark object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2553–2562, October 2021.

[11] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 886–893. Ieee, 2005.

[12] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021.

[13] Vincent Dumoulin and Francesco Visin. A guide to convolution arithmetic for deep learning, 2016.

[14] Yu Li Feifan Lv and Feng Lu. Attention-guided low-light image enhancement. *arXiv preprint arXiv:1908.00682*, 2019.

[15] Shang-Hua Gao, Ming-Ming Cheng, Kai Zhao, Xin-Yu Zhang, Ming-Hsuan Yang, and Philip Torr. Res2net: A new multi-scale backbone architecture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(2):652–662, 2021.

[16] Tao Gong, Kai Chen, Xinjiang Wang, Qi Chu, Feng Zhu, Dahua Lin, Nenghai Yu, and Huamin Feng. Temporal roi align for video object recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 1442–1450, 2021.

[17] 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. *CoRR*, abs/2001.06826, 2020.

[18] Khurram Azeem Hashmi, Didier Stricker, and Muhammamd Zeshan Afzal. Spatio-temporal learnable proposals for end-to-end video object detection, 2022.

[19] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[20] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 770–778. IEEE Computer Society, 2016.

[21] 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. *IEEE Transactions on Image Processing*, 30:2340–2349, 2021.

[22] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image processing*, 6(7):965–976, 1997.

[23] Edwin H Land. An alternative technique for the computation of the designator in theretinex theory of color vision. *Proceedings of the national academy of sciences*, 83(10):3078–3080, 1986.

[24] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)*, volume 2, pages 2169–2178. IEEE, 2006.

[25] Chongyi Li, Chunle Guo Guo, and Chen Change Loy. Learning to enhance low-light image via zero-reference deep curve estimation. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[26] Jinxiu Liang, Jingwen Wang, Yuhui Quan, Tianyi Chen, Jiaying Liu, Haibin Ling, and Yong Xu. Recurrent exposure generation for low-light face detection. *IEEE Transactions on Multimedia*, 24:1609–1621, 2022.

[27] Tsung-YiLin, Piotr Dollár, Ross B. Girshick, Kaiming He, Bharath Hariharan, and Serge J. Belongie. Feature pyramid networks for object detection. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 936–944. IEEE Computer Society, 2017.

[28] Tsung-Yi Lin, Priya Goyal, Ross B. Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. *CoRR*, abs/1708.02002, 2017.

[29] Tsung-YiLin, Michael Maire, Serge J. Belongie, Lubomir D. Bourdev, Ross B. Girshick, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: common objects in context. *CoRR*, abs/1405.0312, 2014.

[30] Wenyu Liu, Gaofeng Ren, Runsheng Yu, Shi Guo, Jianke Zhu, and Lei Zhang. Image-adaptive yolo for object detection in adverse weather conditions. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(2):1792–1800, June. 2022.

[31] Ze Liu, Han Hu, YutongLin, Zhuliang Yao, Zhenda Xie, Yixuan Wei, Jia Ning, Yue Cao, Zheng Zhang, Li Dong, FuruWei, and Baining Guo. Swin transformer v2: Scaling up capacity and resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12009–12019, June 2022.

[32] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10012–10022, October 2021.

[33] Yuen Peng Loh and Chee Seng Chan. Getting to know low-light images with the exclusively dark dataset. *CoRR*, abs/1805.11227, 2018.

[34] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[35] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.

[36] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60:91–110, 2004.

[37] TengyuMa, LongMa,Xin Fan, Zhongxuan Luo, andRisheng Liu. PIA: parallel architecture with illumination allocator for joint enhancement and detection in low-light. In João Magalhães, Alberto Del Bimbo, Shin'ichi Satoh, Nicu Sebe, Xavier Alameda-Pineda, Qin Jin, Vincent Oria, and Laura Toni, editors, *MM '22: The 30th ACM International Conference on Multimedia, Lisboa, Portugal, October 10 - 14, 2022*, pages 2070–2078. ACM, 2022.

[38] Krystian Mikolajczyk and Cordelia Schmid. Scale & affine invariant interest point detectors. *International journal of computer vision*, 60:63–86, 2004.

[39] Igor Morawski, Yu-An Chen, Yu-Sheng Lin, and Winston H. Hsu. NOD: taking a closer look at detection under extreme low-light conditions with night object detection dataset. *CoRR*, abs/2110.10364, 2021.

[40] Siyuan Qiao, Liang-Chieh Chen, and Alan Yuille. Detectors: Detecting objects with recursive feature pyramid and switchable atrous convolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10213–10224, June 2021.

[41] Qingpao Qin, Kan Chang, Mengyuan Huang, and Guiqing Li. Denet: Detection-driven enhancement network for object detection under adverse weather conditions. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, pages 2813–2829, December 2022.

[42] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *CoRR*, abs/1804.02767, 2018.

[43] ShaoqingRen, Kaiming He, Ross B. Girshick, and Jian Sun. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(6):1137–1149, 2017.

[44] Liu Risheng, Ma Long, Zhang Jiaao, Fan Xin, and Luo Zhongxuan. 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.

[45] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[46] Sebastian Ruder. An overview of gradient descent optimization algorithms, 2016.

[47] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Acdc: The adverse conditions dataset with correspondences for semantic driving scene understanding. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10765–10775, October 2021.

[48] Robin Strudel, Ricardo Garcia, Ivan Laptev, and Cordelia Schmid. Segmenter: Transformer for semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7262–7272, October 2021.

[49] Peize Sun, Rufeng Zhang, Yi Jiang, Tao Kong, Chenfeng Xu, Wei Zhan, Masayoshi Tomizuka, Lei Li, Zehuan Yuan, Changhu Wang, and Ping Luo. Sparse R-CNN: end-to-end object detection with learnable proposals. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021*, pages 14454–14463. Computer Vision Foundation / IEEE, 2021.

[50] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[51] Wenjing Wang, Xinhao Wang, Wenhan Yang, and Jiaying Liu. Unsupervised face detection in the dark. *IEEE Transactions*

on Pattern Analysis and Machine Intelligence, 45(1):1250–1266, 2023.

[52] Wenjing Wang, Wenhan Yang, and Jiaying Liu. Hla-face: Joint high-low adaptation for low light face detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16195–16204, June 2021.

[53] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *CoRR*, abs/1808.04560, 2018.

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

[55] Haiping Wu, Yuntao Chen, Naiyan Wang, and Zhaoxiang Zhang. Sequence level semantics aggregation for video object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

[56] Wenhui Wu, Jian Weng, Pingping Zhang, Xu Wang, Wenhan Yang, and Jianmin Jiang. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5901–5910, June 2022.

[57] Tete Xiao, Yingcheng Liu, Bolei Zhou, Yuning Jiang, and Jian Sun. Unified perceptual parsing for scene understanding. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

[58] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M. Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 12077–12090. Curran Associates, Inc., 2021.

[59] Ke Xu, Xin Yang, Baocai Yin, and Rynson W.H. Lau. Learning to restore low-light images via decomposition-and-enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[60] Xinwei Xue, Jia He, Long Ma, Yi Wang, Xin Fan, and Risheng Liu. Best of both worlds: See and understand clearly in the dark. In *Proceedings of the 30th ACM International Conference on Multimedia*, MM '22, page 2154–2162, New York, NY, USA, 2022. Association for Computing Machinery.

[61] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[62] Wenhan Yang, Ye Yuan, Wenqi Ren, Jiaying Liu, Walter J. Scheirer, Zhangyang Wang, Taiheng Zhang, Qiaoyong Zhong, DiXie, ShiliangPu, Yuqiang Zheng, Yanyun Qu, Yuhong Xie, Liang Chen, Zhonghao Li, Chen Hong, Hao Jiang, Siyuan Yang, Yan Liu, Xiaochao Qu, Pengfei Wan, Shuai Zheng, Minhui Zhong, Taiyi Su, Lingzhi He, Yandong Guo, Yao Zhao, Zhenfeng Zhu, Jinxiu Liang, Jingwen Wang, Tianyi Chen, Yuhui Quan, Yong Xu, BoLiu,Xin Liu, Qi Sun, Tingyu Lin, Xiaochuan Li, Feng Lu, Lin Gu, Shengdi Zhou, Cong Cao, Shifeng Zhang, Cheng Chi, Chubing Zhuang, Zhen Lei, Stan Z. Li, Shizheng Wang, Ruizhe Liu, Dong Yi, Zheming Zuo, Jianning Chi, Huan Wang, Kai Wang, Yixiu Liu, Xingyu Gao, Zhenyu Chen, Chang Guo, Yongzhou Li, Huicai Zhong, Jing Huang, Heng Guo, Jianfei Yang, Wenjuan Liao, Jiangang Yang, Liguo Zhou,Mingyue Feng, and Likun Qin. Advancing image understanding in poor visibility environments: A collective benchmark study. *IEEE Transactions on Image Processing*, 29:5737–5752, 2020.

[63] Bo Zhang, Yuchen Guo, Runzhao Yang, Zhihong Zhang, Jiayi Xie, Jinli Suo, and Qionghai Dai. Darkvision: A benchmark for low-light image/video perception, 2023.

[64] Hao Zhang, Feng Li, Shilong Liu, Lei Zhang, Hang Su, Jun Zhu, Lionel Ni, and Heung-Yeung Shum. DINO: DETR with improved denoising anchor boxes for end-to-end object detection. In *The Eleventh International Conference on Learning Representations*, 2023.

[65] Wenqiang Zhang, Tianheng Cheng, Xinggang Wang, Shaoyu Chen, Qian Zhang, and Wenyu Liu. Featurized query r-cnn, 2022.

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

[67] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. *CoRR*, abs/1905.04161, 2019.

[68] Zhaoyang Zhang, Yitong Jiang, Jun Jiang, Xiaogang Wang, Ping Luo,and Jinwei Gu. Star: A structure-aware lightweight transformer for real-time image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4106–4115, October 2021.

[69] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[70] Ziqiang Zheng, Yang Wu, Xinran Han, and Jianbo Shi. Forkgan: Seeing into the rainy night. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision – ECCV 2020*, pages 155–170, Cham, 2020. Springer International Publishing.

[71] Xizhou Zhu, Weijie Su, LeweiLu, Bin Li, Xiaogang Wang, and Jifeng Dai. Deformable {detr}: Deformable transformers for end-to-end object detection. In *International Conference on Learning Representations*, 2021.