2 月工作汇报

Ku Jui

Feb 2024

Content

Part I

Pre-Knowledge

1 引言

低光图像增强（Low-Light Image Enhancement, LLIE）是图像处理领域的一个关键任务，旨在提升低光环境下拍摄图像的视觉质量。该领域的最新进展主要由深度学习技术推动，包括多种学习策略、网络架构、损失函数和训练数据集的应用。低光图像增强在视觉监控、自动驾驶和计算摄影等多个领域具有广泛应用。尤其在智能手机摄影领域，由于相机光圈大小、实时处理需求和存储限制，低光环境下的图像拍摄面临着显著挑战。

传统的低光增强方法，如基于直方图均衡化 [1] 和 Retinex 理论 [2–4] 的方法，虽然能够在一定程度上改善图像质量，但存在一些局限性。直方图均衡化能够提高图像的全局对比度，但可能增加背景噪声的对比度并降低有用图像内容的对比度，导致视觉效果不佳。Retinex 算法旨在消除图像照度分量的干扰并还原图像真实色彩，但通常忽略噪声问题，可能导致增强结果中噪声的保留或放大，并且其复杂的优化过程增加了模型的复杂度。

近年来，基于深度学习的低光图像增强技术取得了显著成就，特别是卷积神经网络（CNN）在多个计算机视觉任务中展现出卓越的性能。CNN 通过利用注意力机制 [5,6] 和上下文信息，能够从原始图像中有效提取局部特征 [7–9]。在低光图像中，亮度较低、对比度较弱的区域之间存在一定的关联性和相互作用，如果模型能够捕获全局光照，将有助于恢复图像的整体亮度和对比度。

同样，即使在低光条件下，对象的轮廓和边缘仍然是重要的视觉特征。捕获这些长距离的边缘信息对于保持图像的结构完整性至关重要。此外，低光图像中的纹理和模式可能难以辨识，但它们对于理解图像内容仍然很重要。长距离特征可以帮助模型识别和恢复这些纹理和模式。

不同对象和场景元素之间的空间关系是理解图像的关键。在低光图像中，捕获这些长距离的上下文关系对于正确解释图像至关重要。然而，现有的基于 CNN 的方法在处理局部光照不均匀、颜色信息和细节信息丢失问题时，仍存在过增强或增强不足的挑战，并且增强结果受到感受野大小的限制。因此，开发能够克服这些限制并有效提升低光图像质量的深度学习模型仍然是一个重要的研究方向。

为了解决这些问题，我们提出了一种新颖的方法（方法名待定）。受到 ULite [10] 中深度可分离卷积的启发，我们将 U-Net 网络中的卷积改为轴向深度可分离卷积，以减少模型冗余同时保持增强效果。此外，我们借鉴 Swin-Unet [11] 在 BottleNeck 中使用连续的 Swin Transformer 块以捕获图像长距离特征，相较于传统的 Transformer 块，Swin Transformer 能够减少参数量和模型复杂度。通过以上改进，我们的方法能够有效提升低光图像的视觉质量，同时保持较低的计算复杂度，具有实际应用的潜力。

本研究提出了三个主要的创新点：(若并行架构未实现或性能不佳，则可尝试仅通过 CNN 分支进行图像的弱光恢复，并相应调整创新点，去除并行架构这一创新点。)

- 首先，提出了一种结合卷积神经网络（CNN）和 Transformer 的并行架构用于弱光图像增强。在该架构中，UNet 网络用于捕获图像的局部特征并进行初步恢复，而 Swin Transformer 块用于捕获图像的长距离特征。最后，通过特征融合块将两者的特征进行融合，以实现更全面的图像增强效果。

- 其次，提出了一个深度语义模块，该模块融合了 Swin Transformer 分支，使 CNN 分支能够有效捕获图像的长距离特征。这种设计增强了 CNN 分支的能力，使其能够更好地理解图像的全局上下文信息。

- 最后，将深度可分离卷积融合进 CNN 分支中，应用于轻量级网络用于提取图像的局部特征。这种设计

旨在减少模型的参数量和计算复杂度，同时保持对局部特征的有效提取能力。

2 实验计划

实验的过程中，确保所有的实验在相同的硬件和软件环境下进行，并且为了确保结果的可靠性，可能需要多次运行实验并取平均值。我们主要基于 PyTorch 进行模型的搭建、训练和评估。基于 scikit-image 库计算 PSNR、SSIM 等评价指标。首先构建 U-Net 基本架构模型，然后实现 Swin Transformer 块中的 LocalselfAttention 类，PositionEncoding 类，PositionEmbedding 类。

2.0.1 Dataset

Tab. 1 展示了我们在实验中会使用到的弱光数据集，这些数据集包含真实数据与合成数据。对于每个数据集，我们需要进行如下操作：

- 预处理: 确保所有图像都经过相同的预处理步骤，如尺寸调整、归一化等。
- 分割: 将每个数据集分为训练集、验证集和测试集。

Name	Number	Format	Real/Syn	Video
LOL [12]	500	RGB	Real	
SCIE [13]	4,413	RGB	Real	
VE-LOL-L [14]	2,500	RGB	Real+Syn	

Table 1: Summary of paired training datasets. 'Syn' represents Synthetic.

2.0.2 Train

- 基线模型: 首先，训练基线模型。
- 消融研究: 接着，训练正常 BottleNeck 的模型，以进行消融实验。

2.0.3 Performance Evaluation

对于每个数据集，使用以下指标评估模型性能：

- 峰值信噪比 (PSNR)
- 结构相似性指数 (SSIM)
- 感知图像质量评估 (LPIPS)

2.0.4 Loss Function

$$J_{Huber}(\delta) = \frac{1}{N}\sum_{i=1}^{N}\begin{cases}\frac{1}{2}\|\hat{y}_i - y_i\|_2^2, \|\hat{y}_i - y_i\| < \delta, \\ \delta\left(\|\hat{y}_i - y_i\|_1 - \frac{1}{2}\delta\right), \|\hat{y}_i - y_i\| \geq \delta.\end{cases} \tag{1}$$

$$\ell_{feat}^{\phi,j}(\hat{y}, y) = \frac{1}{C_j H_j W_j}\|\phi_j(\hat{y}) - \phi_j(y)\|_2^2 \tag{2}$$

$$\mathcal{L}^{\mathrm{SSIM}}(P) = 1 - \mathrm{SSIM}(\tilde{p}). \tag{3}$$

一共尝试以下两种损失函数的搭配方式:

- 休伯损失函数和 SSIM 损失函数

$$L_{loss} = \alpha J_{Huber}(\delta) + \beta \mathcal{L}^{\mathrm{SSIM}}(P) \tag{4}$$

- 休伯损失函数，SSIM 损失函数，Perceptual 损失函数 (耗费更多训练时间)

$$L_{loss} = \alpha J_{Huber}(\delta) + \beta \mathcal{L}^{\mathrm{SSIM}}(P) + \gamma \ell_{feat}^{\phi,j}(\hat{y}, y) \tag{5}$$

Part II

Paper Reading

3 Lightweight Model

3.1 (2023.8)1M parameters are enough? A lightweight CNN-based model for medical image segmentation

1M 的参数就足够了吗？一种基于 CNN 的轻量级医学图像分割模型

(APSIPA ASC 2023 2 区) doi: 10.1109/APSIPAASC58517.2023

3.1.1 Research Background

卷积神经网络 (CNNs) 和基于 Transformer 的模型由于能够提取图像的 High-Level 特征和捕捉图像的重要方面而被广泛应用于医学图像分割。然而，在对高精度的需求和对低计算成本的期望之间往往存在权衡。具有更高参数的模型理论上可以获得更好的性能，但也会导致更高的计算复杂性和更高的内存使用率，因此实现起来并不实用。

在本文中，作者寻找一种轻量级的基于 U-Net 的模型，它可以实现几乎相当甚至更好的性能，即 U-Lite。作者基于深度可分离卷积的原理设计了 U-Lite，这样该模型既可以利用神经网络的强度，又可以减少大量的计算参数。

3.1.2 Contribution

具体来说，作者提出了在编码器和解码器中都具有 7×7 的轴向深度卷积，以扩大模型的感受野。为了进一步提高性能，作者使用了几个具有 3×3 的轴向空洞深度卷积作为作者的分支之一。

总体而言，U-Lite 仅包含 878K 参数，比传统 U-Net 少 35 倍。与其他最先进的架构相比，所提出的模型降低了大量的计算复杂性，同时在医学分割任务上获得了令人印象深刻的性能。

在本文中，作者重新思考了一种用于医学分割任务的高效轻量级架构，以进一步探索一种能够有效解决这一问题的高性能模型。

简而言之，本文的主要贡献有 3 个方面：

(1) 基于深度可分离卷积的概念，提出了轴向深度卷积模块的使用方法。该模块帮助模型解决每一个复杂的体系结构问题：扩大模型的感受野，同时减少沉重的计算负担。

(2) 提出 U-Lite，一种基于 CNN 的轻量级、简单的架构。据作者所知，U-Lite 是为数不多的在性能和参数数量方面超过最近高效紧凑型网络 UneXt 的型号之一。

(3) 作者已经在医学分割数据集上成功地实现了该模型，并取得了可观的效果。

3.1.3 Approach

作者提出的 U-Lite 模型的概述如 Fig. 1 所示。作者遵循 U-Net 的对称编码器-解码器架构，并以一种有效的方式设计 U-Lite，以便该模型能够利用 CNN 的强度，同时保持计算参数的数量尽可能少。

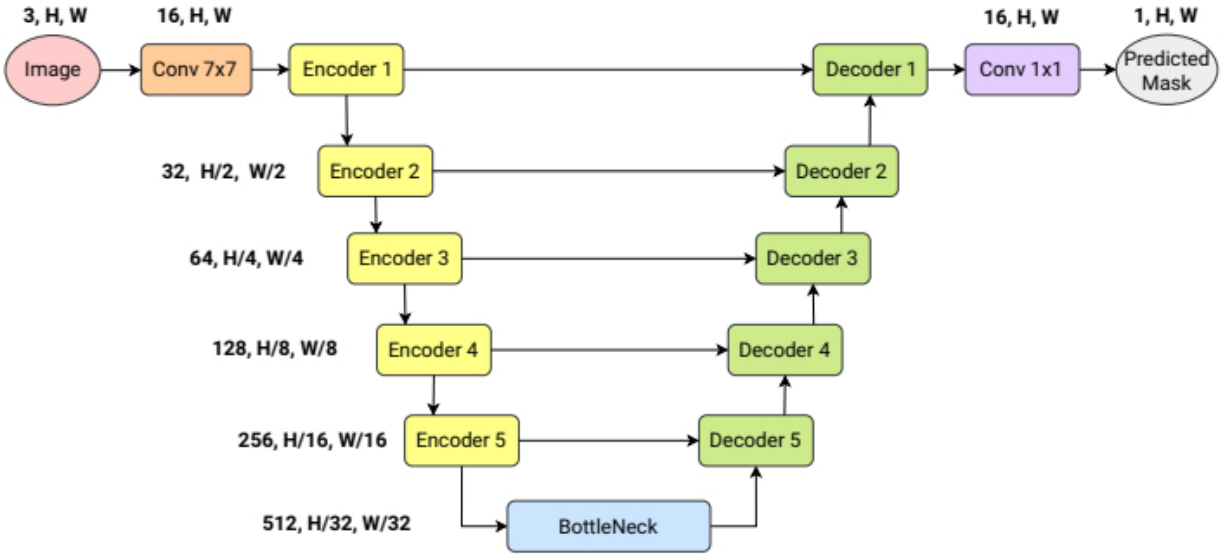

Figure 1: The proposed U-Lite architecture.

为此，作者深思熟虑地提出了一个轴向深度卷积模块，如 Fig. 2 所示。描述 U-Lite 的操作，形状为 $(3, H, W)$ 的输入图像通过 3 个阶段被馈送到网络：编码器阶段、Bottleneck 阶段和解码器阶段。U-Lite 遵循分层结构，其中编码器提取形状 $\left(C_i, \frac{H}{2^i}, \frac{W}{2^i}\right)$ 中的 6 个不同 Level 的特征，其中 $i \in \{0, 1, 2, \ldots, 5\}$。

Bottleneck 和解码器参与处理这些特征，并将它们放大到原始形状以获得分割 Mask。作者还在编码器和解码器之间使用 skip connections 连接。值得注意力的是，尽管 U-Lite 的设计很简单，但由于轴向深度卷积模块的贡献，该模型在分割任务上仍然表现良好。

Axial Depthwise Separable Convolution Module

轴线深度可分离卷积是一种特殊的卷积操作，它将传统的深度可分离卷积进一步分解为两个独立的轴向卷积：水平轴向卷积和垂直轴向卷积。

(1) **深度可分离卷积 (Depthwise Separable Convolution):** 这是一种将标准卷积分解为两个较小的卷积操作的技术。首先是深度卷积（Depthwise Convolution），它对每个输入通道单独应用卷积核，然后

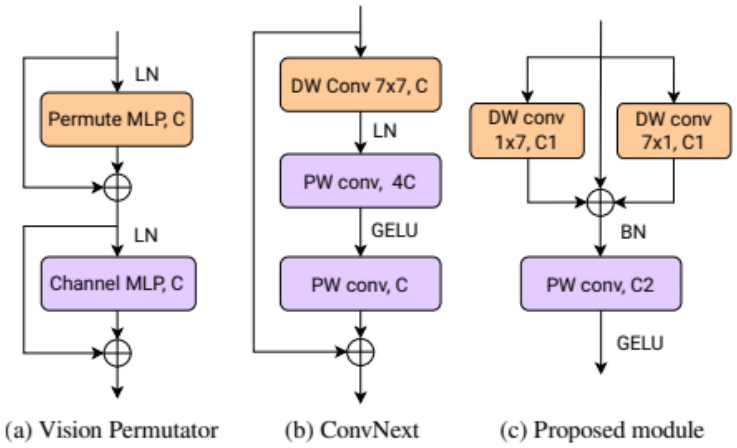

Figure 2: Architectures of (a) Vision Permutator, (b) ConvNext, and (c) Proposed Axial DW Convolution module. The proposed module is inspired by Vision Permutator's and ConvNext's designs.

是逐点卷积（Pointwise Convolution），它使用 1x1 的卷积核将深度卷积的结果合并在一起。这种分解减少了计算量和参数数量。

(2) **轴向卷积 (Axial Convolution):** 在轴向深度可分离卷积中，深度卷积被进一步分解为两个轴向操作：一个沿水平方向（宽度），另一个沿垂直方向（高度）。这样做可以进一步减少计算量，同时保持良好的特征提取能力。

视觉 Transformer 的成功推动了研究和改进这种特殊结构的各种工作。Swin-Transformer 通过将自注意力计算限制在大小为 7×7 的非重叠局部窗口，降低了 Transformer 的计算复杂性。ConvNext 实现了这一修改，并在 CNN 架构中采用了 kernel 大小为 7×7 的卷积，使 ResNet 在 ImageNet 上的最高精度达到 86.4%。

Vision Permutator 利用线性投影来沿着高度和宽度维度分别编码特征表示。

论文提及：如果用局部感受野[1]取代 ViT 的十字形感受野[2]，就像 Swin Transformer[3] 对 ViT 所作的那样，会发生什么？

感受野是一个视觉模型效果的至关重要的属性之一，因为模型是无法建模它感知不到的区域的。在 DeepLab 系列算法中，空洞卷积在不增加参数量的同时可以快速增大感受野。Swin-Transformer [15] 仅仅通过移动窗口的方式来增加感受野的方式在增加感受野上仍然过于缓慢，因为这个算法仍然需要大量堆叠网络块的形式来增加感受野。CSwin-Transformer(Cross-shape window) [16] 是 swin-Transformer 的改进版，它提出了通过十字形的窗口来进行自注意力计算，它不仅计算效率非常高，而且能够通过两层计算就获得全局的感受野。

因此，作者提出了轴向深度卷积模块，作为 Vision Permutator 和卷积设计的组合。该算子的数学公式表示为

[1]局部感受野 (Local Receptive Field) 指的是一个神经元只对输入数据的一个局部区域进行响应，这有助于捕捉图像中的局部特征，如边缘、角点等。

[2]十字形感受野 (Cross-shaped Receptive Field) 指的是在 ViT 中，每个注意力头计算的是全局的自注意力，这意味着每个输出位置的注意力是基于整个输入图像的。但在一些变体中，例如 CrossViT，会使用十字形的感受野，即在水平和垂直方向上分别计算注意力，以减少计算量并捕捉不同方向的特征。

[3]Swin Transformer 利用局部感受野取代十字感受野，它将输入图像划分为多个小块，然后在这些小块内部计算自注意力，从而实现局部感受野。其可以减少计算量，增强模型的泛化能力，并提高效率。

$$
\begin{aligned}
x' &= x + \mathrm{DW}_{1\times n}(x) + \mathrm{DW}_{n}(x) \\
y &= \mathrm{GELU}\left(PW_{C_1\to C_2}\left(\mathrm{BN}\left(x'\right)\right)\right)
\end{aligned}
\tag{6}
$$

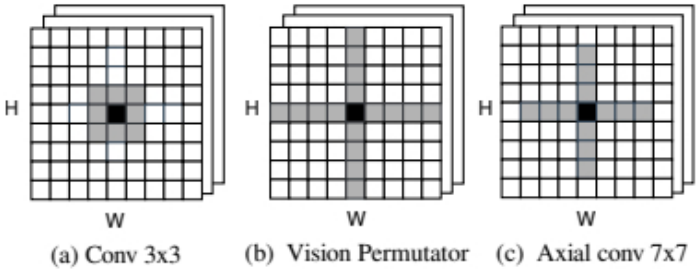

Figure 3: The receptive field comparison between Convolution 3×3, Vision Permutator, and Axial convolution 7×7. Axial convolution 7×7 offers a large receptive field compared with Convolution 3×3 while using fewer computational parameters than Vision Permutator.

其中，x 为输入特征，y 为输出特征；DW、PW 和 BN 分别代表深度卷积、点卷积和批量归一化，$1 \times n$ 和 $n \times 1$ 是卷积的 kernel 大小；C_1 和 C_2 表示特征图的输入和输出通道的数量。在作者的实验中，$n = 7$。为了实现最小和灵活的设计，作者使用了一种独特的逐点卷积，而不添加残差连接，允许自适应地改变输入通道的数量。

Encoder Block and Decoder Block

编码器和解码器块的设计原理如下：

(1) 遵循深度可分离卷积架构。这是从头开始成功构建轻量级模型的重要关键。深度可分离卷积在使用较少参数的同时，提供了与传统卷积相同的性能，从而降低了计算复杂性，使模型更加紧凑。

(2) 限制使用不必要的操作 op。只需使用普通的 MaxPooling 和 UpSampling 层。不需要诸如转置卷积之类的高参数消耗算子。逐点卷积算子可以同时扮演两个角色：沿着特征图的深度对特征进行编码，同时灵活地改变输入通道的数量。

(3) 每个编码器或解码器块采用一个批量标准化层，并以 GELU 激活功能结束。作者对批处理规范化和层规范化进行了性能比较，但没有太大区别。应用 GELU 是因为与 ReLU 和 ELU 相比，在使用 GELU 时证明了其在准确性方面的改进。

U-Lite 的编码器和解码器结构如 Fig. 4 所示。

Bottleneck Block

为了进一步提高 U-Lite 的性能，作者将 kernel 大小 $n = 3$ 的轴向扩展深度卷积应用于 Bottleneck 块。应用的空洞率为 $d = 1, 2, 3$。作者使用具有大小为 3 的 kernel 的轴向扩张卷积，原因有两个：

(1) 大小为 3 大小的 kernel 更适合底层特征的空间形状，其中这些特征的高度和宽度减少了多次。

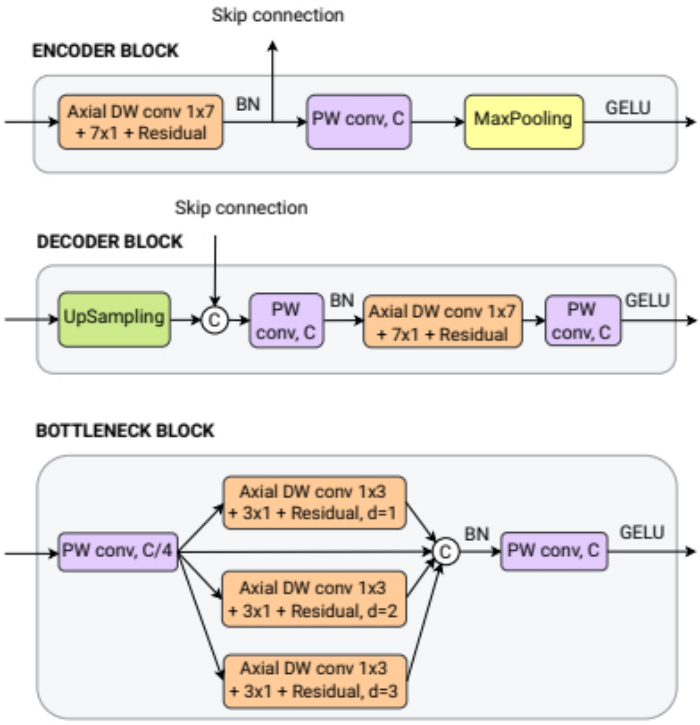

Figure 4: Encoder, decoder, and bottleneck blocks. Designed based on Depthwise Separable Convolution concept. Each block adopts one Batch Normalization layer and ends with a GELU activation function.

(2) 当使用具有不同空洞率的空洞卷积来捕获后面阶段的 High-Level 特征的多空间表示时，它给出了更好的性能。

为了进一步减少可学习参数的数量，在 Bottleneck 块的开头采用了逐点卷积层。这有助于在将最后一层特征提供给轴向扩展深度卷积机制之前缩小其通道尺寸。

3.1.4 Future

-

4 Edge Detection

4.1 (2022.3)Survey of Image Edge Detection

图像边缘检测综述

(Frontiers in Signal Processing 2 区) doi: 10.3389/frsip.2022.826967

4.1.1 Research Background

作者对边缘检测算法进行全面的分析和专题研究。首先，通过对传统边缘检测算法的多层次结构进行分类，介绍了各类算法的原理和方法。其次，重点分析了重点分析了基于深度学习的边缘检测算法，分析了各算法的技术难点、方法优势以及骨干网络选择。然后，通过在 BSDS500 和 NYUD 数据集上的实验，进一

步评估了每种算法的性能。可以看出，目前的边缘检测算法的性能已经接近甚至超越了人类视觉水平。目前，关于图像边缘检测的综合综述文章较少。本文致力于对边缘检测技术进行全面的分析，旨在为相关人员轻松跟进边缘检测的最新发展并做出进一步的改进和创新提供参考和指导。

4.1.2 Contribution

由于图像边缘的重要性，图像边缘检测自提出以来就受到了研究人员的广泛关注，Fig. 5 说明了边缘检测算法的发展。

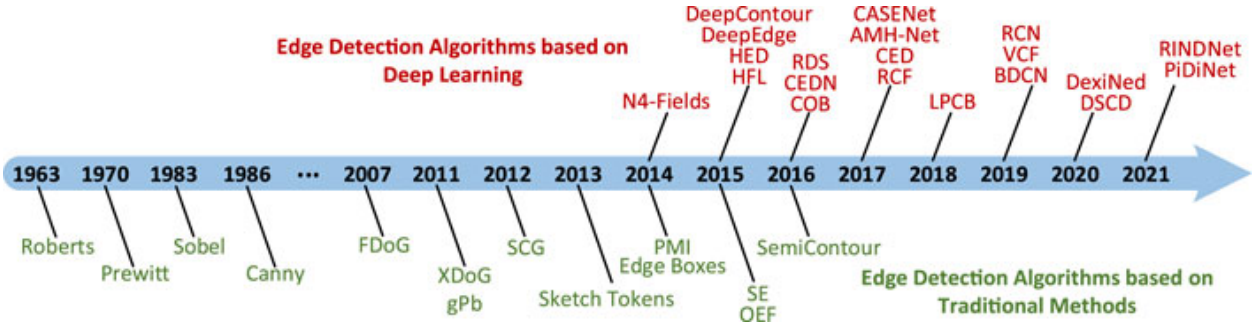

Figure 5: Development of edge detection algorithms based on traditional and deep learning methods.

随着深度学习的不断发展，出现了各种基于 CNN 实现边缘检测的方法。2015 年，Bertasius 等人。改变了传统的自下而上的边缘检测思路，提出了一种自上而下的多尺度发散深度网络 DeepEdge [17] 用于边缘检测。同年，Xie 等人开发了整体嵌套边缘检测算法 HED [18]，解决了基于图像的整体训练和预测以及多尺度多层次特征学习的问题。2017 年，Liu 等人，提出了一种使用更丰富的卷积特征的精确边缘检测器 RCF [19]。2019 年，Deng 等人，提出了一种新颖的端到端边缘检测系统 DSCD [20]，它有效地利用多尺度和多层次特征来产生高质量的物体边缘输出。2021 年 Su 等人, 设计了 PiDiNet [21]，一种简单、轻量级且高效的边缘检测架构。

作者通过在 BSDS500 和 NYUD 数据集上的实验，进一步评价了上述每种算法的性能。

	Canny	FDoG	XDoG	gPb	SCG	Sketch tokens	SE	OEF	SemiContour
ODS↑	0.60	0.63	0.65	0.71	0.72	0.73	0.75	**0.76**	0.74
OIS↑	0.63	0.65	0.66	0.74	0.74	0.75	0.77	**0.79**	0.77

Table 2: Performance comparison of conventional edge detection based algorithms on the BSDS500 dataset.

	Canny	FDoG	XDoG	gPb	SCG	Sketch tokens	SE	OEF	SemiContour
ODS↑	0.47	0.49	0.50	0.53	0.62	0.63	0.65	**0.69**	0.68
OIS↑	0.46	0.50	0.51	0.54	0.63	0.63	0.67	0.69	**0.70**

Table 3: Performance comparison of conventional edge detection based algorithms on the NYUD dataset.

	N4-fields	DeepEdge	HED	HFL	CEDN	RDS	COB	CASENet	RCF
ODS↑	0.47	0.49	0.50	0.53	0.62	0.63	0.65	**0.69**	0.68
OIS↑	0.46	0.50	0.51	0.54	0.63	0.63	0.67	0.69	**0.70**
-	AMH-Net	LPCB	VCF	RCN	BDCN	DexiNed	DSCD	PiDiNet	RINDNet
ODS↑	0.79	0.81	0.81	0.82	0.82	**0.83**	0.82	**0.83**	**0.83**
OIS↑	0.83	0.83	0.82	0.83	0.84	0.84	**0.85**	**0.85**	0.84

Table 4: Performance comparison of deep learning-based edge detection algorithms on the BSDS500 dataset.

	N4-fields	DeepEdge	HED	HFL	CEDN	RDS	COB	CASENet	RCF
ODS↑	0.61	0.68	0.74	0.73	0.75	0.65	0.76	0.77	0.75
OIS↑	0.63	0.69	0.76	0.74	0.76	0.74	0.75	0.76	0.77
-	AMH-Net	LPCB	VCF	RCN	BDCN	DexiNed	DSCD	PiDiNet	RINDNet
ODS↑	0.77	0.76	0.78	0.77	0.77	0.78	**0.80**	0.79	**0.80**
OIS↑	0.78	0.77	0.78	0.79	0.77	0.77	0.79	**0.80**	**0.80**

Table 5: Performance comparison of deep learning-based edge detection algorithms on the NYUD dataset.

4.1.3 Approach

Evaluation Indicators

最佳数据集尺度 (Optimal Dataset Scale, ODS) 和最佳图像尺度 (Optimal Image Scale, OIS) 是评估图像轮廓生成结果时使用最广泛、最具代表性的评价指标，此外还有常用的每秒帧数 (Frames Per Second, FPS) 和精确召回率 (Precision-Recall, PR) 曲线。ODS-F 和 OIS-F 中的 F 值是 Precision(P) 和 Recall(R) 的总和平均值，并被表示为 Eq. 7。

$$\text{F-Score} = \left(1+\beta^2\right) \cdot \frac{Precision \cdot Recall}{\beta^2 \cdot Precision + Recall} \tag{7}$$

通过调整 β 的值，准确度和召回率的显著性程度是可以控制的，如果 $\beta = 1$，表示为 Eq. 8；如果 $\beta = 0$，表示为 Eq. 9；如果 $\beta = \infty$，表示为 Eq. 10。

$$\text{F-Score} = \frac{2Precision \cdot Recall}{Precision + Recall} \tag{8}$$

$$\text{F-Score} = Precision \tag{9}$$

$$\text{F-Score} = Recall \tag{10}$$

最佳数据集尺度 (Optimal Dataset Scale, ODS) 对数据集内所有的图像设置相同的阈值，即固定阈值 β 选择并应用于所有图像，以便最大化整个数据集上的 F 分数。

最佳图像尺度 (Optimal Image Scale, OIS) 对每幅图像上设置不同的阈值 β 选择最大化该图像的 F 分数。

每秒帧数 (Frames Per Second, FPS)，即目标网络每秒可以检测到多少张图像以及图像刷新的频率。用于评价目标检测的速度，时间越短，检测速度越快。

精确召回率 (Precision-Recall, PR) 曲线，PR 曲线是用精度和召回率两个变量绘制的曲线，其中精度为纵坐标，召回率为横坐标，广泛应用于信息提取领域，表示正样本中实际为正的比例。

Backbone Network

主干网络是边缘检测任务的基本特征提取器，强大的主干网络可以提取更丰富的图像特征。目前大多数基于深度学习的边缘检测模型都使用 AlexNet [22]、VGG16 [23] 和 ResNet [24] 作为骨干网络。

Alex 等人在 2012 年的 ImageNet 竞赛中提出了 AlexNet 网络。该网络赢得了当年的 ImageNet LSVRC，准确率远高于第二名，成为深度学习的又一亮点。AlexNet 网络包含 8 个学习层——5 个卷积层和 3 个全连接层，含有 6000 万个参数和 65 万个神经元。引入了 Relu 激活函数、数据增强、Dropout 和级联池化操作，以防止过拟合并提高模型的整体泛化能力。作者发现，模型的深度似乎在神经网络的性能中起着重要作

用，这一发现也启发了后来 VGG 和 ResNet 网络的结构设计。基于网络重构的边缘检测算法 AMH-Net 和 N4-Fields 网络随后使用 AlexNet 网络作为基础网络。

Simonyan 和 Zisserman 在 2014 年设计了深度卷积网络 VGG16，由 13 个卷积层和 3 个全连接层组成，旨在研究大规模图像识别中卷积网络深度的准确性。Simonyan 等人发现，使用非常小的 3×3 卷积滤波器将神经网络深度推进到 16-19 个权重层，可以显著提高 VGG 模型的性能。它采用了更深的网络结构，结合较小的卷积核和池化核，使其在有效控制模型参数大小的同时获得更多的图像特征，避免了大量计算和复杂结构，同时实现了先进的性能。此外，VGG 模型具有良好的泛化能力，可以很好地推广到其他图像处理领域。因此，该模型现在是基于深度学习的图像边缘检测算法中最受欢迎的骨干网络。

ResNet (2016 年)：尽管网络的深度对模型的性能至关重要，但实验上认为深度网络提取更复杂的特征结构。实验发现，随着网络深度的增加，梯度消失或爆炸，导致网络准确率饱和甚至下降。何等人（2016 年）设计了深度残差网络 (Deep Residual Network)。作者识别了深度模型的“退化”问题，并提出了“快捷连接”的解决方案。ResNet 通过引入残差学习，极大地消除了由于过度深度导致的神经网络训练困难问题，使得网络深度首次超过 100 层，甚至可以超过 1000 层。边缘检测模型 COB 和 AMH-Net 的设计基于 ResNet 的一些思想。

4.1.4 Future

-

4.2 (2020)Dense Extreme Inception Network: Towards a Robust CNN Model for Edge Detection

DeixNeD: 面向一个鲁棒的 **CNN** 边缘检测模型

(CVPR 2020) doi: 10.48550/arXiv.1909.01955

4.2.1 Research Background

现在的基于 CNN 方法的边缘检测有很多，像 DeepEdge，HED，RCF，BDCN 等，这些方法的成功主要是由 CNNs 在不同的尺度上应用于一组大的图像，以及训练正则化技术。以前的数据集都或多或少有些毛病，比如边缘信息不完整，使得训练困难等。

4.2.2 Contribution

(1) 作者提出了一种基于深度学习的边缘检测器，该检测器的灵感来自于 HED 和异常网络。该方法生成的薄边缘图对人眼来说是可信的；它可以用于任何边缘检测任务，而无需事先训练或微调过程。

(2) 作者生成了一个带有仔细注释的边缘的大型数据集。该数据集已用于训练所提出的方法以及用于比较的最先进算法。

4.2.3 Approach

Fig. 6 为 DexiNet 的网络架构，主要包含两个模块，一个是 Dexi 网络，一个是上采样模块 (灰色部分)，RBG 图像输入 Dexi 网络进行特征提取 (Dexi 网络能够较好的避免边缘信息的丢失)，然后提取得到的特征图输入上采样模块进行边缘图像的生成，最后再将特征图进行融合，生成最终的边缘信息图像。

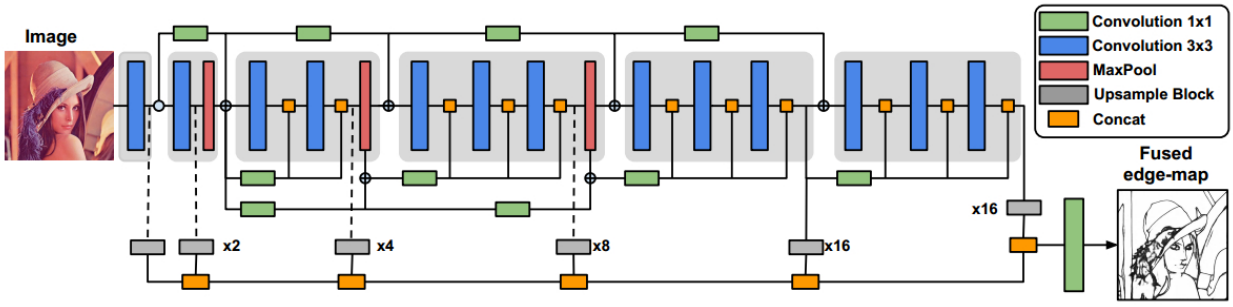

Figure 6: Proposed architecture: Dense Extreme Inception Network, consists of an encoder composed by six main blocks (showed in light gray). The main blocks are connected between them through 1x1 convolutional blocks. Each of the main blocks is composed by sub-blocks that are densely interconnected by the output of the previous main block. The output from each of the main blocks is fed to an upsampling block that produces an intermediate edge-map in order to build a Scale Space Volume, which is used to compose a final fused edge-map.

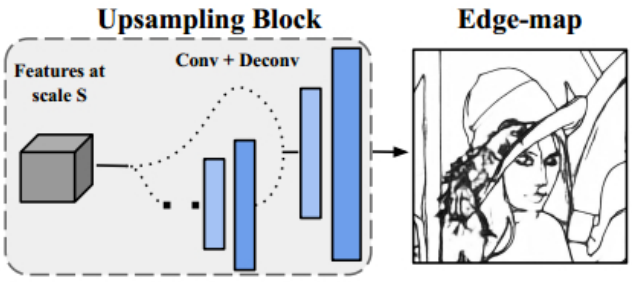

Figure 7: Detail of the upsampling block that receives as input the learned features extracted from each of the main blocks. The features are fed into a stack of learned convolutional and transposed convolutional filters in order to extract an intermediate edge-map.

4.2.4 Future

-

4.3 (2020.10) Deep Structural Contour Detection

深结构轮廓检测

(ACM MM 2020) doi: 10.1145/3394171.3413750

4.3.1 Research Background

4.3.2 Contribution

(1) 作者提出了一种新颖且非常有效的轮廓检测损失函数。所提出的损失函数能够对每对预测与 GT 之间的轮廓结构相似性距离进行惩罚。

(2) 为了更好地区分目标轮廓和背景纹理，作者引入了一种新颖的卷积编码器-解码器网络。在网络中，作者提出了一个捕获高级特征之间密集连接并产生有效语义信息的超级模块。

4.3.3 Approach

作者依照 LPCB, CED, RCF, HED

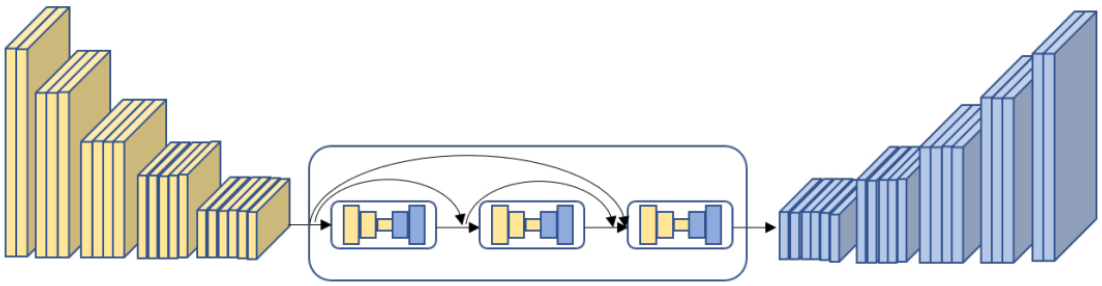

Figure 8: Illustration of the proposed network. The left and the right are the encoder and the decoder, respectively. We adopt the VGG16 as our encoder. The encoder extracts multi-scale, multi-level features and the decoder fuse the features and recover the feature resolution to the original. The middle rectangle is the proposed hyper convolutional module. It adopts three conv blocks and captures dense connection among the hierarchical features. The module can significantly improve model performance. To the best view, we omit the connections between the encoder features and the decoder features.

4.3.4 Future

-

References

[1] T-L Ji, Malur K Sundareshan, and Hans Roehrig. Adaptive image contrast enhancement based on human visual properties. *IEEE transactions on medical imaging*, 13(4):573–586, 1994.

[2] Edwin H Land. The retinex. In *Ciba Foundation Symposium-Colour Vision: Physiology and Experimental Psychology*, pages 217–227. Wiley Online Library, 1965.

[3] Edwin H Land. The retinex theory of color vision. *Scientific american*, 237(6):108–129, 1977.

[4] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. Properties and performance of a center/surround retinex. *IEEE transactions on image processing*, 6(3):451–462, 1997.

[5] Chenglin Yang, Siyuan Qiao, Adam Kortylewski, and Alan Yuille. Locally enhanced self-attention: Combining self-attention and convolution as local and context terms. *arXiv preprint arXiv:2107.05637*, 2021.

[6] Cheng Zhang, Qingsen Yan, Yu Zhu, Xianjun Li, Jinqiu Sun, and Yanning Zhang. Attention-based network for low-light image enhancement. In *2020 IEEE international conference on multimedia and expo (ICME)*, pages 1–6. IEEE, 2020.

[7] Anil K Jain and Farshid Farrokhnia. Unsupervised texture segmentation using gabor filters. *Pattern recognition*, 24(12):1167–1186, 1991.

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

[9] Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7):971–987, 2002.

[10] Binh-Duong Dinh, Thanh-Thu Nguyen, Thi-Thao Tran, and Van-Truong Pham. 1m parameters are enough? a lightweight cnn-based model for medical image segmentation. In *2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 1279–1284. IEEE, 2023.

[11] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *European conference on computer vision*, pages 205–218. Springer, 2022.

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

[13] Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing*, 27(4):2049–2062, 2018.

[14] Haiyang Jiang and Yinqiang Zheng. Learning to see moving objects in the dark. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7324–7333, 2019.

[15] 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*, pages 10012–10022, 2021.

[16] Xiaoyi Dong, Jianmin Bao, Dongdong Chen, Weiming Zhang, Nenghai Yu, Lu Yuan, Dong Chen, and Baining Guo. Cswin transformer: A general vision transformer backbone with cross-shaped windows. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12124–12134, 2022.

[17] Gedas Bertasius, Jianbo Shi, and Lorenzo Torresani. High-for-low and low-for-high: Efficient boundary detection from deep object features and its applications to high-level vision. In *Proceedings of the IEEE international conference on computer vision*, pages 504–512, 2015.

[18] Saining Xie and Zhuowen Tu. Holistically-nested edge detection. In *Proceedings of the IEEE international conference on computer vision*, pages 1395–1403, 2015.

[19] Yun Liu, Ming-Ming Cheng, Xiaowei Hu, Kai Wang, and Xiang Bai. Richer convolutional features for edge detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3000–3009, 2017.

[20] Ruoxi Deng and Shengjun Liu. Deep structural contour detection. In *Proceedings of the 28th ACM international conference on multimedia*, pages 304–312, 2020.

[21] Zhuo Su, Wenzhe Liu, Zitong Yu, Dewen Hu, Qing Liao, Qi Tian, Matti Pietikäinen, and Li Liu. Pixel difference networks for efficient edge detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5117–5127, 2021.

[22] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[23] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[24] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.