无名稿件编号（将由编辑插入）

面向高效超高清图像复原与增强的神经判别提示变换器

Cong Wang[1,2] · **Jinshan Pan**[3] · **Liyan Wang**[4] · **Wei Wang**[1] · **Yang Yang**[5]

arXiv:2603.00853v1 [cs.CV] 1 Mar 2026

收稿日期：日期 / 录用日期：日期

摘要 我们提出了一种简单而有效的UHDPromer，即一种神经判别提示变换器，用于超高清（UHD）图像复原与增强。我们的UHDPromer受到一个有趣观察的启发：高分辨率特征与低分辨率特征之间隐含存在神经差异，而探索这些差异有助于低分辨率特征表示。为此，我们首先引入神经差异先验（NDP）来度量这些差异，然后将NDP集成到所提出的神经判别提示注意力（NDPA）和神经判别提示网络（NDPN）中。所提出的NDPA通过纳入NDP重新表述注意力，以全局感知有用的判别信息；而NDPN则探索一种由NDP引导的连续门控机制，以选择性地允许有益内容通过。为增强复原图像的质量，我们提出一种超分辨率引导的重建方法，该方法通过对低分辨率特征进行超分辨率重建来引导，从而促进最终的超高清图像复原。实验表明，UHD-Promer 在实现最高计算效率的同时，仍在以下任务上保持了最先进的性能：3 超高清图像复原与增强任务，包括低光照图像增强、图像去雾和图像去模糊。源代码和预训练模型将发布于 https://github.com/supersupercong/uhdpromer。

关键词 神经判别先验 · Transformer · 超高清图像复原与增强 · 低光照图像增强 · 图像去雾 · 图像去模糊

Cong Wang 电子邮箱：supercong94@gmail.com
Jinshan Pan 电子邮箱：sdluran@gmail.com
Liyan Wang 电子邮箱：wangliyan@mail.dlut.edu.cn
Wei Wang 电子邮箱：wangwei29@mail.sysu.edu.cn
Yang Yang 电子邮箱：yang.yang4@ucsf.edu

[1] 中国中山大学深圳校区[2] 中国香港香港理工大学[3] 中国南京理工大学[4] 中国大连理工大学[4] 美国加州大学旧金山分校

1 引言

近年来，成像传感器和显示器的进步显著推动了超高清（UHD）成像的发展。然而，超高清图像通常指分辨率为 3，840×2，160 的图像或 4K图像，在低照度、雾霾或快速运动等具有挑战性的条件下拍摄时，往往会出现图像质量退化，损害视觉质量并影响高级视觉任务的性能。本文通过提出一种通用框架，旨在应对这些挑战并恢复退化的超高清图像。

卷积神经网络（卷积神经网络）（Krizhevsky 等，2012年；He 等，2016a,b；Xie 等，2017年；Huang 等，2017年；Liu 等，2022b；Woo 等，2023年）和 Transformer（Dosovitskiy 等，2021年；Liu 等，2021b, 2022a,c）的出现，推动图像复原取得了显著进展（Cai 等，2016年；Zhang 等，2017年、2018年；Li 等，2018年；Zamir 等，2021；Pan 等，2021年、2017年、2023年；Ren 等，2021年；Bai 和 Pan，2024；Wang 等，2022a；Tsai 等，2022年；Pan 等，2022年；Wang 等，2023a、2024a,c）。尽管取得了这些进展，现有方法主要关注通用图像复原任务，但受计算限制，难以处理超高清图像尺寸，也难以有效

提升超高清图像质量。这些局限性限制了此类算法在超高清成像系统中的潜在应用。

近年来，为满足处理退化超高清图像的需求，超高清图像复原方法层出不穷。值得注意的方法包括 Zheng 等人(2021年)；Li 等人(2023年)，这些方法采用复杂的深度神经网络，主要通过对高分辨率特征进行下采样，从低分辨率特征中学习，从而能够处理尺寸更大的超高清图像。然而，这些方法往往忽略了高分辨率特征与低分辨率特征之间神经差异的关键探索，而这些差异隐含着宝贵信息，可指导低分辨率学习，进而提升超高清图像复原效果。忽视这一方面可能会显著降低重建质量。Wang 等人Wang 等人(2024b) 的近期研究表明，高分辨率特征包含更多有用信息，可有效指导低分辨率特征学习。基于这一发现，他们提出了 UHDformer，通过基于相关性匹配的特征变换，将知识从高分辨率特征迁移至低分辨率特征。尽管该方法采用极轻量级模型并取得了出色的性能，但仍存在两个主要局限。首先，相关性计算依赖于高维特征空间中的矩阵相似度运算，这不可避免地增加了计算开销，降低了模型效率。因此，我们需要一种效率更高且性能优异的模型。其次，UHDformer 只关注直接特征变换，却未深入探索高分辨率特征与低分辨率特征之间的关系及层级依赖关系。因此，构建一个更具可解释性和洞察力的模型仍值得探索。

为解决上述挑战，我们提出了一种神经判别提示 Transformer，称为 UHDPromer，以更高效地进行超高清图像复原与增强。UHDPromer 的设计灵感源于一项有趣的观察：高分辨率特征与低分辨率特征之间隐含着神经差异，这些差异能够为低分辨率域提供有用信息，从而改善低分辨率特征表示。具体而言，我们引入神经判别先验（NDP）来衡量神经差异。基于 NDP，我们设计了神经判别提示 Transformer（NDPT），将 NDP 融入 Transformer 设计中，以有效学习更具表征能力的低分辨率特征。NDPT 包含两个关键组件：神经判别提示注意力（NDPA）和神经判别提示网络（NDPN）。NDPA 重新审视了 NDP 与由低分辨率特征得到的查询向量之间的交叉注意力，并将其与从低分辨率特征中获得的其余键向量和值向量进一步组合，形成新的注意力，从而从长程像素依赖关系的角度有效感知更多有用的判别信息。NDPN 探索了一种连续门控机制，利用 NDP 持续引导前馈编码过程。该机制以逐像素门控调制的方式传递更多有益内容，从而增强低分辨率特征的表示能力。

一种自然的最终干净图像重建方式，是融合低分辨率分支的深层特征与高分辨率分支的深层特征。然而，我们注意到，这种方式无法保证取得良好结果。为解决这一问题，我们进一步提出一种超分辨率引导的重建方法：首先对 NDPT 中学习到的低分辨率特征进行超分辨率处理，然后利用这些超分辨率特征引导最终的 UHD 图像重建。通过上述设计，我们的 UHDPromer 在各类 UHD 图像复原与增强任务上实现了最先进的性能，同时兼具高效率。

我们的主要贡献总结如下：

– 我们通过探索高分辨率特征与低分辨率特征之间隐含的神经判别先验，提出了一种简单、有效且性能强大的 UHDPromer，用于高效的 UHD 图像复原。– 我们提出了富有成效的神经判别提示注意力机制和有效的神经判别提示网络，将神经判别先验融入注意力和前馈网络设计中。– 实验表明，在包括低光照增强、去雾和去模糊在内的 3 项 UHD 图像复原与增强任务中，面对不同的训练设置，我们的 UHDPromer 在计算效率高的同时，仍能实现最佳恢复质量。

2 相关工作

在本节中，我们回顾最相关的图像复原方法，包括基于卷积神经网络和基于Transformer的技术。我们还讨论现有的UHD图像复原与增强方法。

2.1 基于CNN的图像复原方法

基于卷积神经网络的架构（Liu 等，2019；Zhao等，2021；Pan等，2018；Yang 等人，2017；Fu等，2017；Cai等，2016；Ren等，2016；Wang等，2020；Cho 等，2021；Zamir 等，2021；Dong 等，2020；Zhang 等，2019；Anwar等，2019；Li 等人，2019b；Tian

等，2020）通过从大规模数据中隐式学习先验，已被证明优于传统复原方法（He 等，2011；Fattal，2014；Pan等，2016，2017；Berman等，2016；Li 等人，2016；Huang等，2013；Chen和Hsu，2013；Luo等，2015）。大多数CNN架构基于残差学习、密集连接、编码器-解码器和注意力模型设计。例如，Cho 等，2021提出了多输入多输出的MIMO-UNet，通过对多尺度输入图像进行编码，并解码出不同尺度的多幅去模糊图像，从而高效融合多尺度特征。为弥补高分辨率特征中缺失的空间信息，Dong 等，2020提出了一种基于U-Net、具有密集特征融合模块的多尺度增强去雾网络。残差学习与跳跃连接已用于图像去模糊（Kupyn等，2018），而密集连接则用于图像去雾（Dong 等，2020）。编码器-解码器（Cho 等，2021）和基于注意力的模型（Liu 等，2019）常用于提取多尺度信息。Liu 等，2019提出了一种名为GridDehazeNet、用于单幅图像去雾的端到端可训练卷积神经网络，该网络采用基于注意力的多尺度估计来提取多尺度信息。随后，Zamir 等，2021提出了一种名为MPRNet的多阶段图像复原方法，通过在每个阶段利用监督式注意力学习子网络，逐步复原清晰图像，取得了显著进展。尽管基于卷积神经网络的方法已取得出色成果，但由于主要学习局部特征，它们仍缺乏建模全局信息的能力。

2.2 基于变换器的图像复原方法

基于变换器的模型（Liang 等人，2021年；Wang 等人，2022c；Zamir 等，2022；Tsai 等人，2022年；Kong 等，2023；Song 等，2023；Chen 等人，2021年）在图像复原方面取得了不错的成果。它们擅长建模长程像素依赖关系，克服了基于CNN的算法仅在局部窗口中进行计算的局限。Liang 等人（2021年）提出了一种强大的基线模型 SwinIR，该模型基于 Swin Transformer，用于图像复原。Zamir 等（2022）还提出了一种高效的 Transformer 模型 Restormer，它采用 U-Net 结构，包含多头注意力和前馈网络，并在多个图像复原任务上取得了最先进的结果。此外，Wang 等人（2022c）提出了分层编码器-解码器网络 Uformer，通过采用局部增强窗口（LeWin）变换器模块和可学习的多尺度复原调制器，利用变换器块进行图像复原。然而，尽管这些方法在一般图像复原任务中表现出色，但由于依赖大量参数——计算量大且 MACs 高，这些基于变换器的方法在超高清成像设备上的广泛应用因此受到限制。

2.3 超高清图像复原与增强

成像系统中处理超高清图像的需求日益增长，推动了大量创新方法的发展，以复原清晰的超高清图像。针对这一需求，研究者提出了众多网络设计，例如用于图像去雾的双边学习（Zheng 等，2021）、用于视频去模糊的多尺度可分离块集成网络（Deng 等人，2021a），以及用于低光照图像增强的傅里叶嵌入网络（Li 等，2023）。受双边学习方法的启发，Zheng 等（2021）创新性地提出了专用于超高清图像去雾的多引导双边学习方法，该方法利用深度卷积神经网络构建仿射双边网格，以保留图像中的细节边缘和纹理。与此同时，Deng 等人（2021a）进一步构建了融合可分离块集成方案的多尺度网络，以获得对超高清去模糊至关重要的宽广感受野。在低光照图像增强领域，Wang 等（2023b）提出了一种配备基于轴的多头自注意力和跨层注意力融合的 Transformer 架构，以降低线性复杂度。受傅里叶域中显著特性的启发，Li 等（2023）主张将傅里叶变换融入级联网络，并通过独特的实现策略有效实现幅度增强和相位增强。Wang 等（2024b）提出了首个用于超高清图像复原问题的 Transformer 架构 UHD-former。作者发现，高分辨率特征能够提供更多有用信息，从而改善低分辨率特征的表示。基于这些发现，作者提出通过建立高分辨率特征到低分辨率特征的变换，并在二者之间进行相关性匹配来实现这一目标。在（Yu 等，2024）中，Yu 等人设计了一种在模型知识指导下的定制化重采样方法，通过补偿信息学习将缩放器与增强器相结合，以解决超高清图像增强问题。Wang 等（2024d）并未仅关注输入图像的内在特征，而是提出了一种双交互先验驱动的超高清图像复原网络，在低分辨率空间中融合法线先验和梯度先验，以动态增强结构与细节，从而引导高分辨率特征复原。

本文通过分析高分辨率与低分辨率表征之间的特征差异，对神经判别先验进行了深入研究。我们的关键洞见是，这些差异编码了

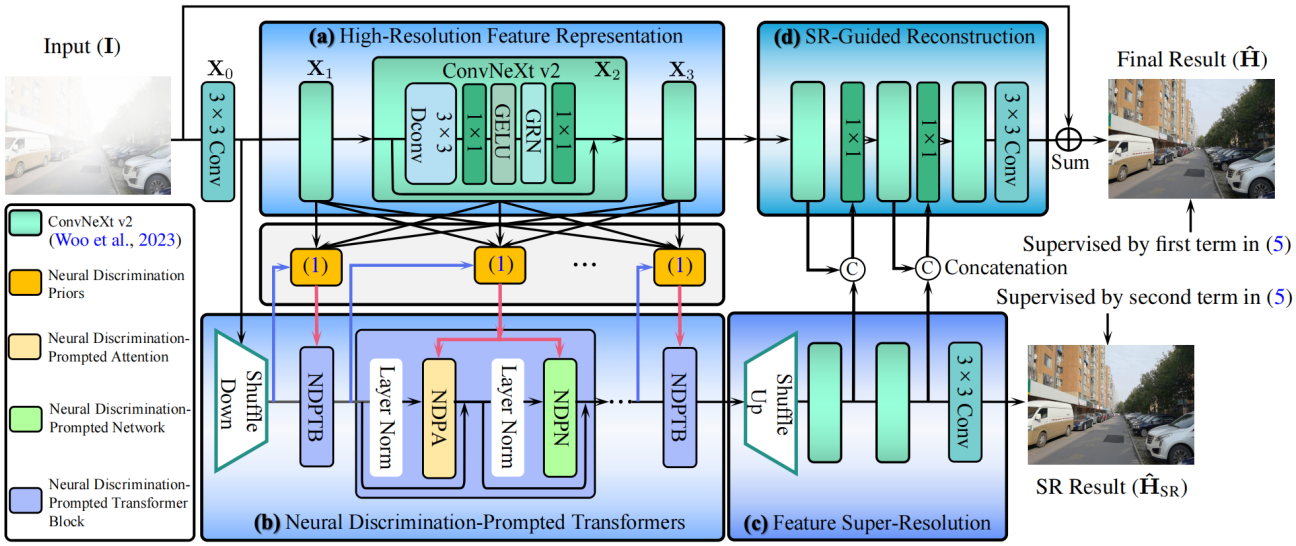


图1 所提出的神经判别提示变换器（UHDPromer）概览。UHDPromer由4个部分组成：**(a)** 高分辨率特征表示（HRFR）；**(b)** 神经判别提示变换器（NDPT）；**(c)**特征超分辨率（FeaSR）；以及**(d)** 超分辨率引导重建（SRG-Recon）。首先，HRFR提取分层多尺度高分辨率特征，这些特征与低分辨率特征共同构成神经判别先验（NDP）。随后，在NDP的引导下，NDPT学习低分辨率特征。接着，FeaSR对NDPT的输出特征进行超分辨率处理。最后，SRG-Recon利用FeaSR中学到的超分辨率特征引导最终图像的重建。

关键的结构信息，可用于提升恢复质量。基于这一发现，我们提出了一种专为超高清图像复原与增强设计的神经判别提示 Transformer 架构。与现有方法相比，所提出的方法在保持计算效率的同时，实现了更优异的复原性能。大量实验验证表明，将神经判别先验融入 Transformer 框架，对超高清图像处理任务十分有效。

3 所提方法

我们的方法受到高分辨率特征与低分辨率特征之间存在神经差异这一现象的启发。这些差异隐含地包含有用的结构，可作为提示来引导低分辨率特征的学习。因此，我们引入神经判别先验（NDP）来度量这些差异。此外，我们设计了一种新型神经判别提示变换器，将NDP融入Transformer设计中，以有效学习低分辨率特征。此外，我们提出了一种超分辨率引导重建方法，以指导最终的超高清图像复原。

3.1 整体流程

图1展示了我们的UHDPromer整体框架，该框架由四个部分组成：(a) 高分辨率特征表示（HRFR）；(b) 神经判别提示Transformer（NDPT）；(c) 特征超分辨率（FeaSR）；以及 (d) 超分辨率引导重建（SRG-Recon）。给定一幅超高清输入图像 $\mathbf{I} \in \mathbb{R}^{H\times W\times 3}$，我们首先应用一个 3×3 卷积来获得低级嵌入 $\mathbf{X}_0 \in \mathbb{R}^{H\times W\times C}$，其中 $H\times W$ 表示空间维度，C 表示通道数。随后，HRFR通过3个ConvNeXt-v2模块（Woo等人，2023年）对浅层特征 $\mathbf{X}_0$ 进行分层编码，以提取多尺度特征 $\{\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3\} \in \mathbb{R}^{H\times W\times C}$。接下来，NDPT接收由$\mathbf{X}_0$,得到的洗牌下采样特征 $\mathbf{X}_{\text{down}} \in \mathbb{R}^{H/s\times W/s\times C}$，随后将$\mathbf{X}_{\text{down}}$ 送入神经判别提示Transformer（NDPT），以学习低分辨率特征。这里，s 表示洗牌下采样因子。同时，NDPT还受到所提出的神经判别先验（NDP）的引导，NDP能够衡量高分辨率特征与低分辨率特征之间有用的判别内容，从而学习更具代表性的特征，以更好地实现超高清图像复原。之后，FeaSR对NDPT的输出特征进行超分辨率处理，并生成超分辨率图像 $\hat{}\mathbf{H}_{\text{SR}} \in \mathbb{R}^{H\times W\times 3}$。最后，SRG-Recon包含三个ConvNeXt-v2模块（Woo

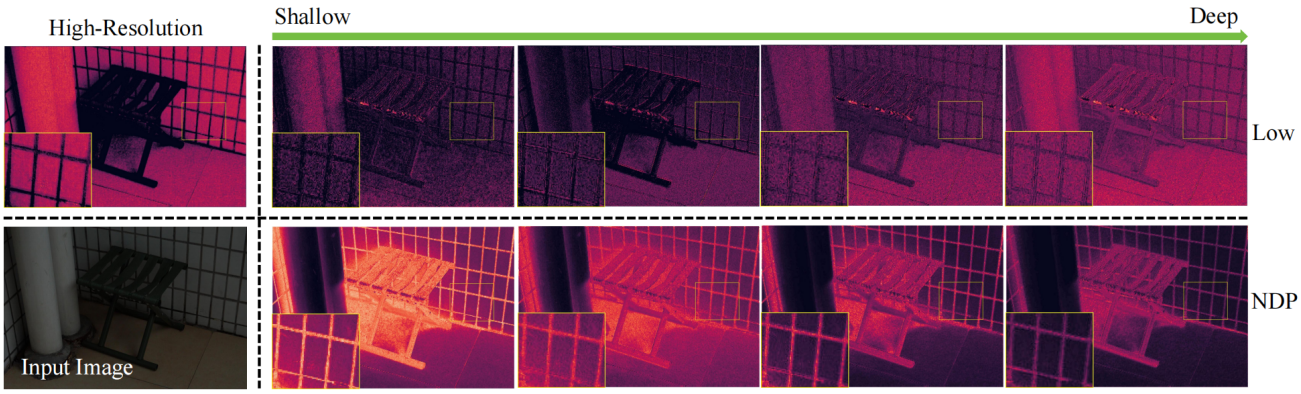


图 2 通过（1）展示了不同层中的神经判别先验（NDP）。我们注意到，与高分辨率特征相比，低分辨率特征存在明显的内容差异。具体而言，低分辨率特征往往携带大量噪声，且某些结构较为模糊。我们的 NDP 能更好地识别这些区域，从而有效判断这些结构位于较浅层还是较深层，并作为判别函数，帮助在低分辨率域中学习更具表征性的特征。

等人，2023年），两个 1×1卷积，以及一个 3×3 卷积，接收来自HRFR的输出特征和来自FeaSR的超分辨率特征，并生成残差图像 $\mathbf{S} \in \mathbb{R}^{H\times W\times 3}$，将超高清输入图像加到该残差图像上^，以获得复原图像：$\mathbf{H} = \mathbf{I} + \mathbf{S}$。

3.2 神经判别先验

由于低分辨率特征是通过对浅层高分辨率嵌入进行下采样重排得到的，因此与输入图像相比，必然会产生一些结构差异。因此，如何校正这些差异并加以利用，从而更好地引导低分辨率特征的学习，是一个值得探索且颇具挑战的课题。本文引入了隐含于高分辨率特征与低分辨率特征之间的神经判别先验，用于衡量上述差异。将这些先验融入低分辨率域的学习过程，可以提升低分辨率特征的表征能力。神经判别先验（NDP）定义如下：

$$NDP_i(x) = 1/\sqrt{e^{\text{abs}|\mathcal{H}_i[\mathbf{X}_1,\mathbf{X}_2,\mathbf{X}_3](x)-\mathbf{Y}_i(x)|}} \tag{1}$$

其中 $\mathcal{H}_i[\cdot,\cdot,\cdot]$ 由拼接操作和步长卷积构成，卷积核大小为 $s\times s$、步长大小为 $s\times s$，而 x 为像素位置；abs $|\cdot|$ 表示绝对值运算。$\mathbf{Y}_i$ 表示第 i个变换器块的低分辨率输入特征，其中 $i = 1、2、\ldots\ldots、L$。公式（1）表明，当NDP_i（x）的值接近1时，位置 x 处的特征显著偏离低分辨率特征，表明其具有更大的判别潜力。因此，我们可以使用NDP_i（x）帮助Transformer更好地区分该特征是否具有判别性，从而使Transformer能够在低分辨率空间中对有用内容投入更多注意力，以获得更好的复原效果。

图 2 展示了高分辨率特征、低分辨率特征和 NDP 的视觉特征图。显然，与低分辨率特征相比，NDP 能更清晰地描述内容，因此有助于引导深度模型更好地学习低分辨率特征，从而实现更佳的复原效果。

3.3 神经判别提示变换器

为了更好地学习低分辨率特征，我们提出了一种简单有效的神经判别提示Transformer块（NDPTB），由开发的NDP引导，并包含神经判别提示注意力（NDPA）和神经判别提示网络（NDPN）。我们的神经判别提示注意力（NDPA）和神经判别提示网络（NDPN）模块受到Restormer中的多深度卷积头转置注意力（MDTA）和门控深度卷积前馈网络（GDFN）架构的启发（Za-mir 等人，2022年）。然而，我们的方法引入了根本性的架构创新。与Restormer的传统注意力机制不同，NDPA通过持续重塑和重组查询-键-值（QKV）表示，动态地将神经判别先验（NDP）融入注意力计算。这使注意力机制能够自适应地聚焦于对高质量重建至关重要的判别性特征

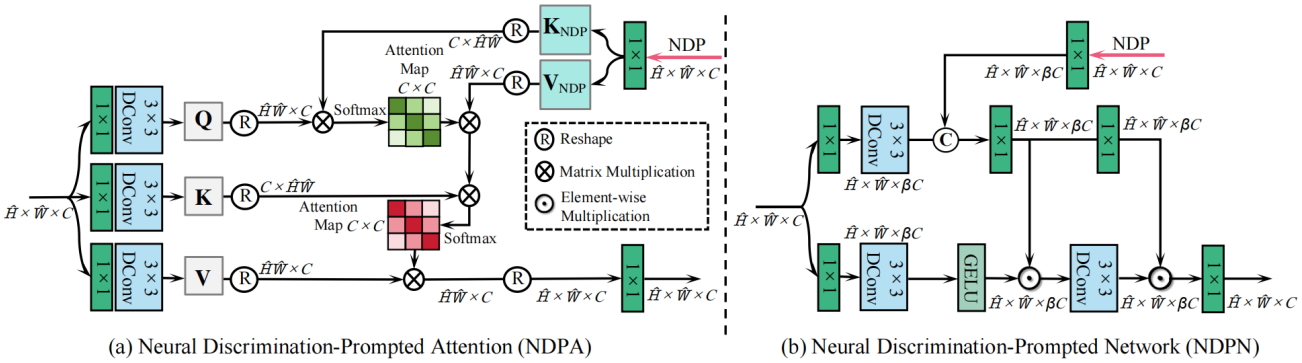


图 3 (a)神经判别提示注意力（**NDPA**）和（**b**）神经判别提示网络（**NDPN**）。NDPA 首先在 NDP 与由低分辨率特征生成的查询向量之间执行交叉注意力，以从全局感知 NDP；随后，通过对低分辨率特征中其余键向量与值向量进行注意力计算，再次执行该过程，以进一步挖掘有用的已学习信息。NDPN 采用由 NDP 引导的连续门控机制，使有用信息得以传递，从而促进最终的图像重建。

。类似地，NDPN通过实施由判别先验引导的持续门控机制，将NDP进一步融入前馈网络。这种设计使前馈网络能够根据特征的判别重要性选择性地处理特征，从而提升整体重建质量。每个NDPTB不仅接收前一个NDPTB的输出，还接收NDP（$\mathbf{Y}_{\text{NDP}}$），以引导低分辨率学习：

$$\begin{aligned}\mathbf{X}^{'} &= NDPA\left(LN(\mathbf{X}^{i}), \mathbf{Y}_{\text{NDP}}\right) + \mathbf{X}^{i},\\ \mathbf{X}^{i+1} &= NDPN\left(LN(\mathbf{X}^{'}), \mathbf{Y}_{\text{NDP}}\right) + \mathbf{X}^{'},\end{aligned} \tag{2}$$

其中$\mathbf{X}^{i}$ 表示 i^{th} NDPTB 的输出特征；LN（·）表示层归一化（Baetal.，2016年）。

3.3.1 神经判别提示注意力

NPDA 如图 3(a) 所示，将 NDP 有效融入注意力机制，从全局视角感知有用的判别信息。不同于现有的注意力设计（Zamir 等人，2022年；Liu 等人，2021b），它们只计算一次注意力，我们提出了一种连续注意力计算方案，通过长程像素依赖建模充分利用 NDP。具体而言，NPDA 首先计算 NDP 特征$\mathbf{Y}_{\text{NDP}}$ 与由低分辨率拆分特征生成的查询**向量**Q^X **之间的交叉注意力**。随后，利用低分辨率拆分特征中剩余的键向量**K** 和值向量**V** ，对注意力图重新计算注意力。该过程利用交叉注意力中聚合的判别信息，有效地引导低分辨率学习。随后可按如下方式计算 NPDA:

$$\begin{aligned}&\mathbf{Q}, \mathbf{K}, \mathbf{V} = \mathcal{S}\left(W_d W_p(\hat{\mathbf{X}})\right),\\ &\mathbf{K}_{\text{NDP}}, \mathbf{V}_{\text{NDP}} = \mathcal{S}\left(W_d W_p(\mathbf{Y}_{\text{NDP}})\right),\\ &\mathbf{X}_{\text{NPDA}} = \mathcal{A}\left(\mathcal{A}\left(\mathbf{Q}, \mathbf{K}_{\text{NDP}}, \mathbf{V}_{\text{NDP}}\right), \mathbf{K}, \mathbf{V}\right),\\ &\text{其中}\ \ \mathcal{A}\left(\hat{}\mathbf{Q}、\hat{}\mathbf{K}、\hat{}\mathbf{V}\right) = \hat{}\mathbf{V} \cdot \text{Softmax}\left(\hat{}\mathbf{K} \cdot \hat{}\mathbf{Q}/\alpha\right);\end{aligned} \tag{3}$$

其中，α是用于控制 ^**K** 与 ^**Q** 点积幅度的可学习缩放参数；$\mathcal{S}$（·）表示拆分操作；W_p（·）表示 1×1 逐点卷积，而 W_d（·）表示 3×3 深度卷积；$\mathbf{X}_{\text{NPDA}}$表示 NPDA 的输出。

3.3.2 神经判别提示网络

我们提出了 NDPN，如图3(b) 所示，以探索由 NDP 引导的连续门控机制，从而允许传递更多有益信息。不同于现有的前馈网络（Zamir 等人，2022年；Liang 等人，2021年；Chen etal.，2021年；Wang 等人，2022c），这些网络只考虑单一的门控方案或点式感知；我们提出一种连续门控机制，它不仅以像素级调制的方式将 NDP 充分融入前馈网络设计，还探索高阶门控，以提供更具判别性的信息，并更好地指导低分辨率学习。具体而言，我们首先使用 1×1卷积，以及 3×3 深度卷积来处理输入的低分辨率特征，并将其拆分为两部分：$\mathbf{Z}_1$ 和 $\mathbf{Z}_2$。然后，我们将 NDP（$\mathbf{Y}_{\text{NDP}}$）与 $\mathbf{Z}_1$ 通过拼接操作进行融合

并通过 1×1 卷积生成融合特征$\mathbf{X}_{\text{fusion}}$。接下来，对$\mathbf{Z}_2$，应用 GELU，再由$\mathbf{X}_{\text{fusion}}$ 对其进行门控，得到门控特征$\mathbf{X}^1_{\text{gate}}$。随后，将门控特征$\mathbf{X}^1_{\text{gate}}$ 输入 3×3深度卷积，并进一步由经过 1×1 卷积处理的融合特征进行门控。最后，对门控特征进行 1×1 卷积，得到最终输出特征。

NDPN 的计算过程可进一步表示为：

$$\begin{aligned}
&\mathbf{Z}_1, \mathbf{Z}_2 = \mathcal{S}\left(W_d W_p(\hat{\mathbf{X}})\right), \\
&\mathbf{X}_{\text{fusion}} = W_p\big(\mathcal{C}\left[\mathbf{Z}_1, \mathbf{Y}_{\text{NDP}}\right]\big), \\
&\mathbf{X}^1_{\text{gate}} = \mathbf{X}_{\text{fusion}} \odot \sigma(\mathbf{Z}_2), \\
&\mathbf{X}_{\text{NDPN}} = W_p\Big(W_d\left(\mathbf{X}^1_{\text{gate}}\right) \odot W_p\left(\mathbf{X}_{\text{fusion}}\right)\Big),
\end{aligned} \tag{4}$$

其中 $\mathcal{C}[\cdot,\cdot]$ 表示拼接操作； $\sigma(\cdot)$表示 GELU 函数；$\odot$ 表示逐元素乘法；$\mathbf{X}_{\text{NDPN}}$ 表示 NDPN 的输出。遵循（Zamir 等人，2022年），我们采用扩展因子 β 来增大中间特征的通道维度，从而提升模型容量。

3.4 训练目标

由于我们的 UHDPromer 包含一个 SR 引导重建分支和一个超分辨率分支，因此有必要设计损失函数，分别约束这两个分支的训练^，以获得更好的结果。具体而言，令$\mathbf{H}^{和}\mathbf{H}_{\text{SR}}$ 分别表示 SR 引导重建分支和超分辨率分支的输出，$\mathbf{H}$ 表示真实值，我们使用以下损失函数来约束网络训练：

$$\mathcal{L} = \phi(\hat{\mathbf{H}}, \mathbf{H}) + \alpha\phi(\hat{\mathbf{H}}_{\text{SR}}, \mathbf{H}), \tag{5}$$

其中 $\phi(\cdot,\cdot)$ 表示（Sun 等人，2022年；Cho 等人，2021年；Wang 等人，2024b）中使用的基于空间域和频域的损失函数。权重 α 经经验设定为0。5。

4 实验

我们针对3项超高清图像复原与增强任务评估了我们的 UHDPromer：**(a)**超高清低光图像增强，**(b)**超高清图像去雾，以及**(c)**超高清图像去模糊。我们还进行了广泛的消融实验，以分析所提出的UHDPromer的有效性，并讨论其局限性。

表 1 UHD图像复原基准的数据集统计。

数据集	训练	测试	分辨率
UHD-LL	2,000	150	3,840 × 2,160
UHD-Haze	2,290	230	3,840 × 2,160
UHD-Blur	1,964	300	3,840 × 2,160

4.1 实验设置

4.1.1 实现

这些是15个NDPTB，即 $L = 15$，其中每个NDPTB的注意力头数量为8。我们将通道数 C 设为16。我们使用 AdamW优化器（Loshchilov and Hutter，2019年）进行模型训练，初始学习率 $5e^{-4}$ 逐渐降至 $1e^{-7}$，并采用余弦退火（Loshchilovand Hutter，2017年）。训练图块大小设为 512×512。洗牌下采样因子设为8，即$s = 8$，而步长卷积中的卷积核大小与步长大小也均设为8。

4.1.2 基准数据集与评估设置

遵循 UHD- former（Wang 等人，2024b）制定的实验方案，我们在三个 UHD 数据集上评估我们的方法：用于超高清低光照图像增强的 UHD-LL（Li 等人，2023年）、用于超高清图像去雾的 UHD-Haze，以及用于超高清图像去模糊的 UHD-Blur。UHD-LL（Li 等人，2023年）由真实场景中采集的低光照/正常光照配对图像组成，噪声伪影极少。对于 UHD-Haze 和 UHD-Blur，我们采用（Wang 等人，2024b）提供的重组版本，这些版本是根据原始数据集（Zheng 等人，2021年；Deng 等人，2021b）重建的，原因是原始版本存在文件访问问题。这些重组后的数据集保留了原始数据集的基本特征，同时确保了数据完整性和可访问性。UHD图像复原基准的数据集统计信息汇总于表1。遵循（Wang 等人，2024b），在评估复原性能时，有两种设置：

设置 ***1***：我们利用通用图像复原数据集训练深度模型，然后在 UHD-LL、UHD-Haze 和 UHD-Blur 中的 UHD测试图像上进行测试。在本研究中，我们采用广泛使用的 LOL（Wei 等人，2018年）、SOTS- ITS（Li 等人，2019a）和 GoPro（Nah 等人，2017年）作为通用低光照图像增强、去雾和去模糊的基准数据集。

设置 ***2***：我们在超高清图像复原基准数据集上训练深度模型，即 UHD-LL、UHD-Haze 和 UHD-Blur 训练集，然后在相应的 UHD测试数据集上进行测试。

表 2 超高清低光照图像增强。我们的**UHDPromer**在不同训练设置下取得了最先进的性能。最佳和次佳结果分别以**粗体**和下划线标出。

方法	发表场所	峰值信噪比 ↑	结构相似性 ↑	学习感知图像块相似度 ↓
设置 *1*：在 LOL 上训练				
SwinIR（Liang 等人，2021年）	ICCVW'21	17.900	0.7379	0.5217
Restormer（Zamir 等人，2022年）	CVPR'22	19.728	0.7703	0.4566
Uformer（Wang 等人，2022b）	CVPR'22	18.168	0.7201	0.5593
LLFlow（Wang 等人，2022b）	AAAI'22	19.596	0.7333	0.4606
LLformer（Wang 等人，2023b）	AAAI'23	21.440	0.7763	0.4528
UHDFour（Li 等人，2023年）	ICLR'23	14.771	0.3760	0.7608
LMAR（Yu 等人，2024）	CVPR'24	22.430	**0.8628**	0.4211
UHDformer（Wang 等人，2024b）	AAAI'24	**22.615**	0.7754	0.4241
UHDPromer（我们的方法）	-	21.714	0.7807	**0.4176**
设置 *2*：在 UHD-LL 上训练				
SwinIR（Liang 等人，2021年）	ICCVW'21	21.165	0.8450	0.3995
Restormer（Zamir 等人，2022年）	CVPR'22	21.536	0.8437	0.3608
Uformer（Wang 等人，2022c）	CVPR'22	21.303	0.8233	0.4013
LLformer（Wang 等人，2023b）	AAAI'23	24.065	0.8580	0.3516
UHDFour（Li 等人，2023年）	ICLR'23	26.226	0.9000	0.2194
LMAR（Yu 等人，2024）	CVPR'24	26.821	0.8840	0.3076
UHDformer（Wang 等人，2024b）	AAAI'24	27.113	0.9271	0.2240
UHDPromer（我们的方法）	-	**27.159**	**0.9285**	**0.2118**

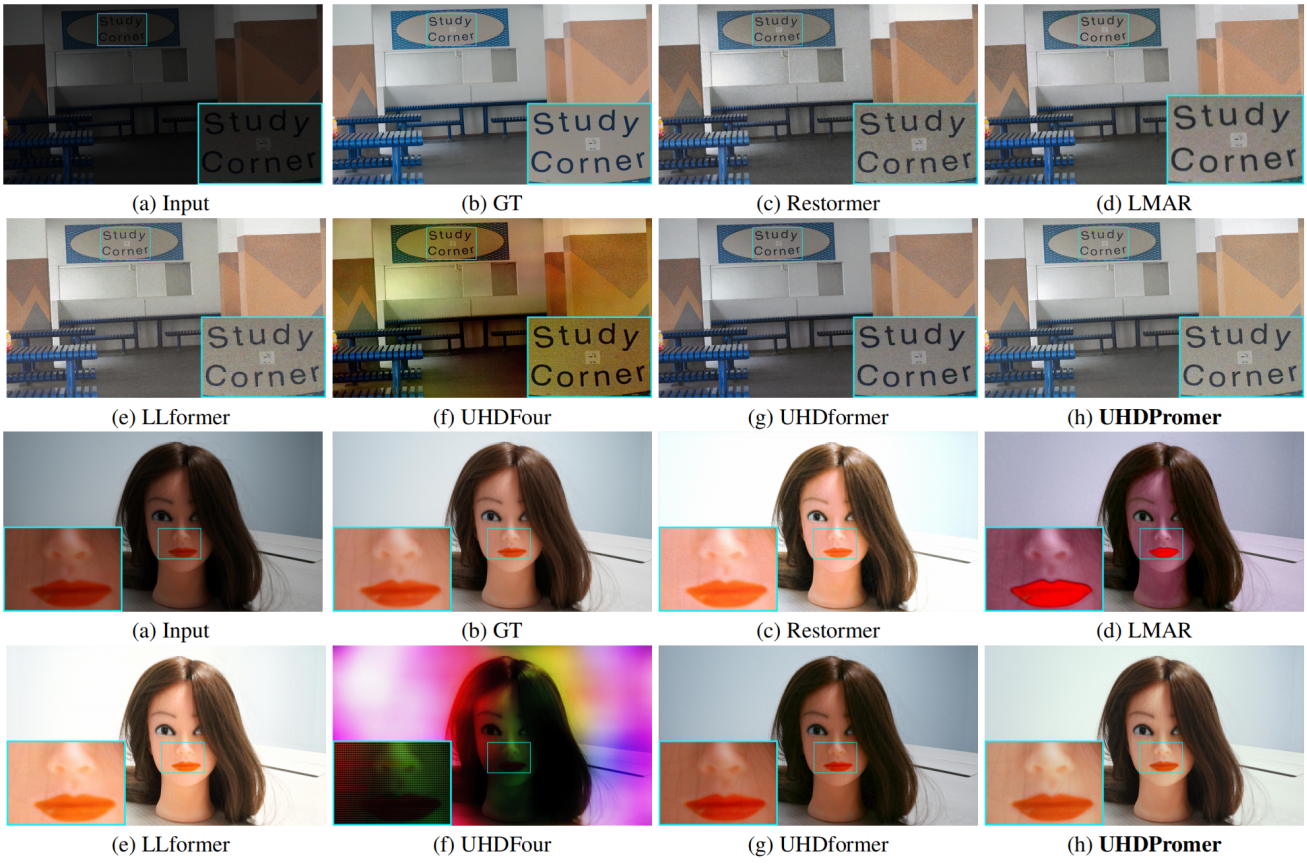


图4 UHD-LL 上的 UHD低光照增强（Li 等人，2023年）在设置 *1*。**UHDPromer**能够生成更清晰且色彩鲜艳的结果。

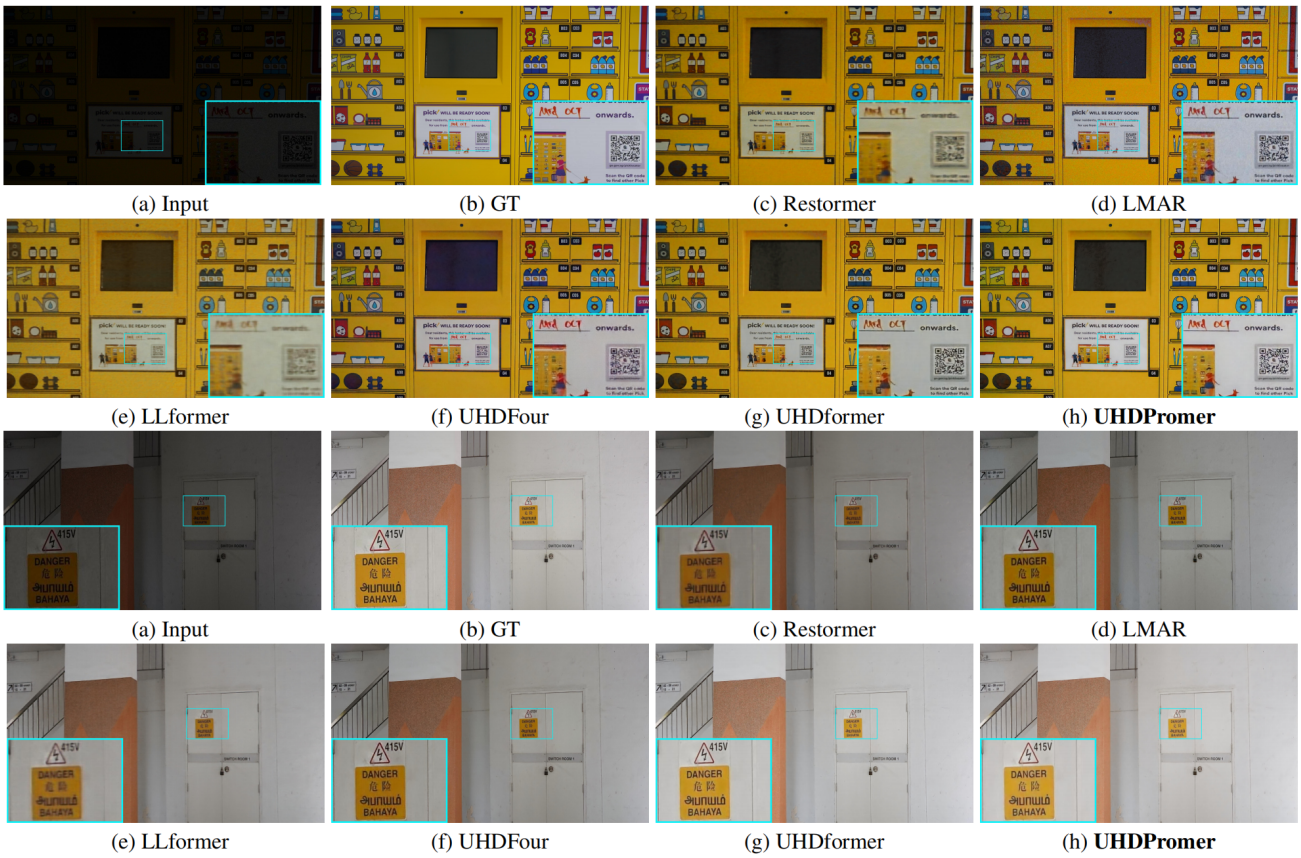


图5 UHD-LL 上的 UHD低光照增强（Li 等人，2023年）在 设置 *2*下。 **UHDPromer** 能够生成更清晰、色彩更自然的结果。

4.1.3 评估指标

遵循 UHDFour（Li 等人，2023年），我们采用常用的 IQA PyTorch 工具箱 [1] 来计算峰值信噪比（Huynh- Thu 和 Ghanbari，2008）和结构相似性（Wang 等人，2004）得分。我们还通过报告可学习参数（参数量）来评估模型大小。由于一些方法无法直接处理全分辨率超高清图像，我们必须采用额外的方法进行实验。根据 UHDFour（Li 等人，2023年），将输入调整至模型可处理的最大尺寸，比将输入分成四个图像块再拼接结果效果更好。

4.2 主要结果

我们对 UHD 低光照图像增强、UHD 图像去雾和 UHD 图像去模糊任务中的超高清图像复原与增强结果进行了定性和定量评估。

4.2.1 低光照图像增强结果

我们评估了不同训练设置下的超高清低光照图像增强效果。表2 显示，在这两种训练设置下，我们的 UHD-Promer 均优于当前最先进的方法。与近期的最先进方法 UHDFour[2] (Li 等人，2023年) 相比，我们的 UHDPromer 分别取得了 6.943 dB 和 0.933 dB 的峰值信噪比提升，并且在不同训练设置下始终表现更优。值得注意的是，设置 1 下的 UHDFour 无法处理超高清图像，而我们的 UHDformer 仍能保持出色的增强性能。此外，尽管一些现有方法（例如 UHDformer（Wang 等人，2024b））无需其他处理步骤即可处理全分辨率 UHD 图像，我们的 UHDPromer 仍优于它们。我们还注意到，专为图像增强设计的现有方法 LMAR Yu 等人(2024) 在设置 1 下的结构相似性方面显著优于所有方法，但在设置 2 下质量较差。

[1] https://github.com/chaofengc/iqa-pytorch

[2] 请注意，为了与我们的方法进行公平比较，我们使用公开发布的代码，采用 8 的洗牌下采样因子，在 LOL（Li 等人，2018年）数据集上重新训练 UHD-Four（Wei 等人，2023年）。

表 3 超高清图像去雾。我们的**UHDPromer**优于最先进的方法。

方法	发表场所	峰值信噪比 ↑	结构相似性 ↑	学习感知图像块相似度 ↓
设置 *1*：在 SOTS-ITS 上训练				
GridNet（Liu 等人，2019年）	ICCV'19	14.783	0.8466	0.2518
MSBDN（Dong 等人，2020年）	CVPR'20	15.043	0.8570	0.2709
UHD（Wang 等人，2021年）	ICCV'21	11.708	0.6569	0.5948
Restormer（Zamir 等人，2022年）	CVPR'22	13.875	0.6405	0.4477
Uformer（Wang 等人，2022c）	CVPR'22	15.264	0.6724	0.4288
DehazeFormer（Song等人，2023年）	TIP'23	14.721	0.6641	0.4100
LMAR（Yu 等人，2024）	CVPR'24	16.896	0.7506	0.3741
UHDformer（Wang 等人，2024b）	AAAI'24	15.325	0.8560	0.2294
UHDPromer（我们的方法）	-	**16.927**	**0.8666**	**0.2034**
设置 *2*：在 UHD-Haze 上训练				
UHD（Zheng 等人，2021年）	ICCV'21	18.043	0.8113	0.3593
Restormer（Zamir 等人，2022年）	CVPR'22	12.718	0.6930	0.4560
Uformer（Wang 等人，2022年）	CVPR'22	19.828	0.7374	0.4220
DehazeFormer（Song等人，2023年）	TIP'23	14.986	0.7202	0.3998
LMAR（Yu 等人，2024）	CVPR'24	20.728	0.9192	0.1789
UHDformer（Wang 等人，2024b）	AAAI'24	22.586	0.9427	0.1188
UHDPromer（我们的方法）	-	**22.725**	**0.9432**	**0.1134**



图6 UHD-Haze 上的超高清图像去雾（Wang 等人，2024b）在设置 *1*下。**UHDPromer** 能够生成干净得多的结果。

这可能表明，LMAR 在不同训练数据下的鲁棒性不足。图 4 和图 5 分别展示了超高清图像上的视觉对比结果。不同设置下的 LL，其中 UHDPromer 能够生成更清晰、色彩更自然的结果。

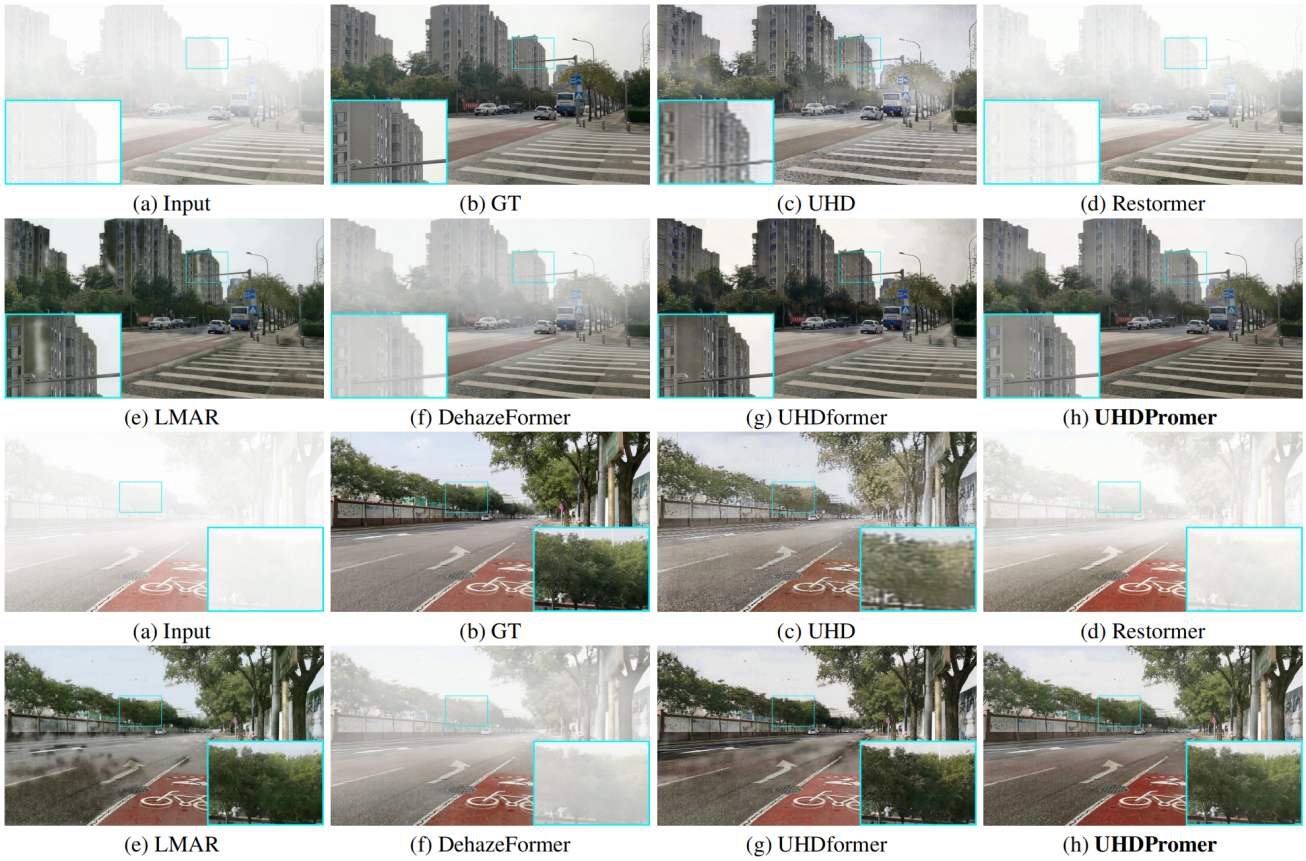


图 7 设置 2 下 UHD-Haze 上的超高清图像去雾（Wang 等人，2024b）设置 *2*。**UHDPromer** 可以生成清晰得多的去雾结果。

4.2.2 图像去雾结果

表 3 总结了不同训练设置下的定量去雾结果。与近期工作 DehazeFormer （Song 等人，2023 年）和 UHDformer（Wang 等人，2024b）相比，我们的 UHDPromer 分别取得了7.739dB 和0.139dB 的峰值信噪比增益（设置 2）。我们注意到，尽管一些基于卷积神经网络的方法，例如 GridNet（Liu 等人，2019年）、UHD（Zheng 等人，2021年）和 MSBDN（Dong 等人，2020年），能够直接处理超高清图像，但它们在设置 1 下处理超高清图像的效果较差，而我们的 UHDformer 仍保持最先进的性能。此外，我们注意到 Restormer（Zamir 等人，2022年）在设置 2 下的指标低于设置 1 下的交叉验证。这可能表明 Restormer 在处理超高清雾霾方面并不有效且不适合。相比之下，我们的 UHDPromer 更适合处理超高清图像。图 6和图7展示了不同训练场景下的视觉对比，表明我们的 UHDPromer 能够产生更清晰的结果。相比之下，其他方法始终会导致雾霾残留过多或产生明显伪影。

4.2.3 图像去模糊结果

我们采用两种建议的训练设置来评估超高清图像去模糊性能。表4表明，我们的UHDPromer在不同训练设置下显著超越最先进方法。与FFTFormer（Kong 等，2023）以及最新的最先进去模糊方法UHDformer（Wang 等，2024b）相比，我们的UHDPromer在设置 2下分别取得了4.118dB和0.706dB的峰值信噪比增益。我们还注意到，一些基于CNN的方法，即DMPHN（Zhang 等，2019）、MIMO-Unet++（Cho 等，2021）和MPRNet（Zamir 等，2021），能够处理全分辨率超高清图像，但它们大约多消耗 100× 和 30×的训练参数，同时在设置 1下峰值信噪比至少下降0.774dB和1.092dB。图8和图9分别提供了超高清图像去模糊的可视化示例，其中我们的 UHDPromer能够产生更清晰的去模糊结果。

表4 图像去模糊。UHDformer优于最先进方法。

方法	发表场所	峰值信噪比 ↑	结构相似性 ↑	学习感知图像块相似度 ↓
设置 *1*：在 GoPro 上训练				
DMPHN（Zhang等，2019年）	CVPR'19	26.490	0.7985	0.2696
MIMO-UNet++（Cho 等人，2021年）	ICCV'21	24.290	0.7354	0.3723
MPRNet（Zamir 等人，2021年）	CVPR'21	24.571	0.7426	0.3597
Restormer（Zamir 等人，2022年）	CVPR'22	24.872	0.7484	0.3550
Uformer（Wang 等人，2022c）	CVPR'22	24.382	0.7209	0.3901
Stripformer(Tsai 等人, 2022年)	ECCV'22	24.915	0.7463	0.3299
FFTformer (Kong 等人, 2023年)	CVPR'23	24.625	0.7396	0.3428
LMAR(Yu 等人,2024)	CVPR'24	23.040	0.7894	0.3513
UHDformer(Wang 等人，2024b)	AAAI'24	27.436	0.8231	0.2774
UHDPromer (我们的方法)	-	**27.582**	**0.8263**	**0.2554**
设置 *2*：在 UHD-Blur 上训练				
MIMO-Unet++(Cho 等人，2021年)	ICCV'21	25.025	0.7517	0.3874
Restormer (Zamir 等人，2022年)	CVPR'22	25.210	0.7522	0.3695
Uformer (Wang 等人，2022c)	CVPR'22	25.267	0.7515	0.3851
Stripformer(Tsai 等人，2022年)	ECCV'22	25.052	0.7501	0.3740
FFTformer（Kong 等人，2023年）	CVPR'23	25.409	0.7571	0.3708
LMAR（Yu 等人，2024）	CVPR'24	26.472	0.7989	0.2470
UHDformer（Wang 等人，2024b）	AAAI'24	28.821	0.8440	0.2350
UHDPromer (我们的)	-	**29.527**	**0.8584**	**0.2163**

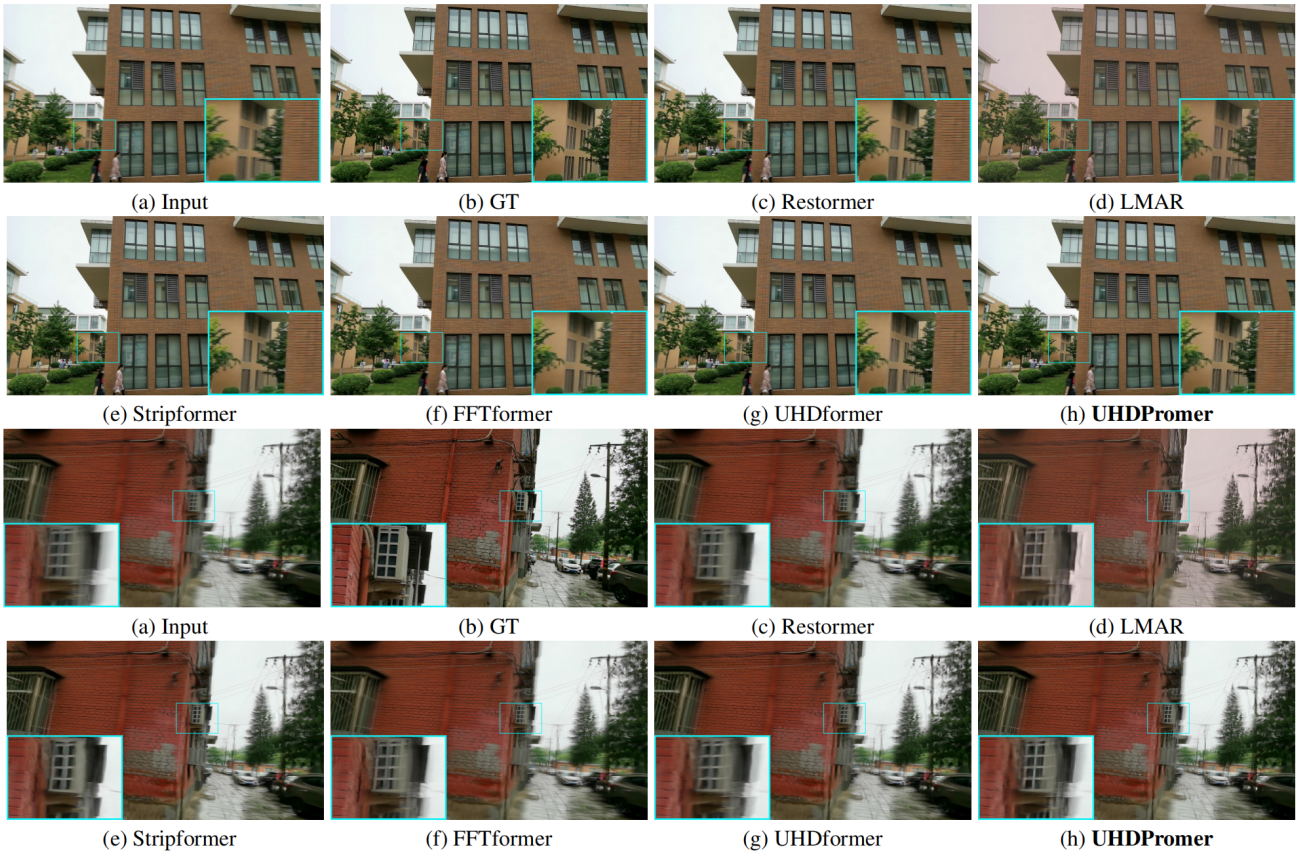


图 8 在 UHD-Blur 上的超高清图像去模糊 (Wang 等人，2024b) 在 设置 *1* 下。 UHDformer 能够生成结构更精细的结果。

4.3 消融实验

我们使用UHD-LL（Li 等人，2023年）开展消融研究，以分析所提出的组件并演示

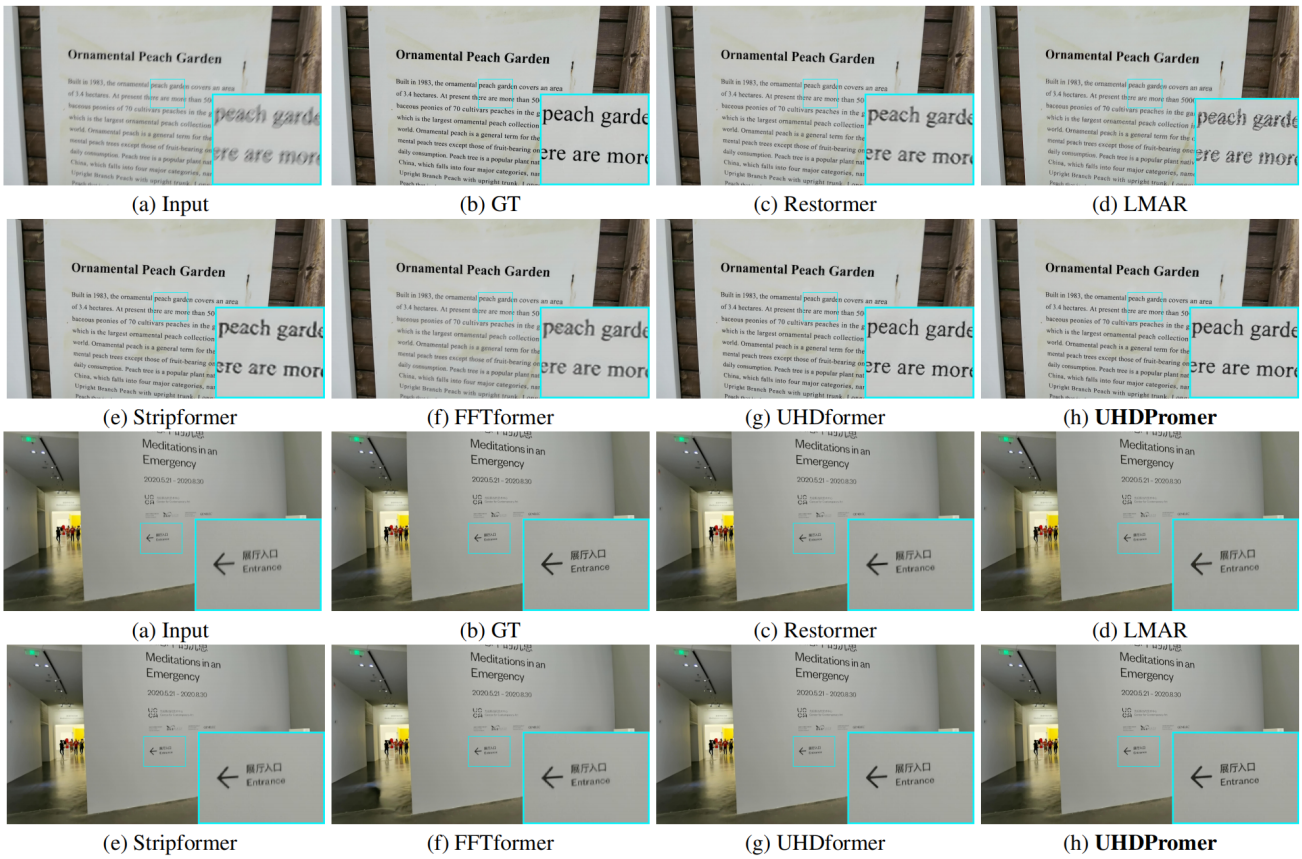


图 9 在 UHD-Blur （Wang 等人，2024b）上进行超高清图像去模糊设置 2**下，**UHDformer 能够生成字符清晰得多的结果。

展示所提出方法的有效性。除了报告最终复原性能（**主分支**）外，我们还提供超分辨率分支（**超分辨率分支**）的峰值信噪比/结构相似性分数供参考。在消融实验中，我们也报告不同模型的参数量供参考。

4.3.1 神经判别提示变换器的效果

我们 UHDPromer 的一个关键之处在于引入了由神经差异先验（NDP）引导的神经差异提示 Transformer（NDPT）。因此，分析 NDPT 的作用至关重要。如表 5 所示，在 NDPA 或 NDPN 中禁用 NDP，或同时在二者中禁用 NDP，都会导致性能明显下降（表 5(f) 与表 5(a)、(b) 和 (c) 的对比）。对于那些想了解使用直接特征（即步长卷积的输出）有何影响的读者，我们用这些直接特征替代 NDP。我们注意到，将 NDP 引入 NDPN 后，性能会下降。这可能是因为将 NDP 集成到 NDPA 中，会使 Transformer 关注判别性内容。然而，单独使用 NDPN 可能无法关注更具判别性的特征，因此 NDPN 无法有效处理中间特征，导致性能下降。

表5 表明，采用我们改进的神经判别先验能够得到更优的模型，相较于使用直接特征的模型，峰值信噪比提升了1.021dB（表 5（f）与表5（d））。为进一步阐明 NDP 在注意力和前馈网络机制中的有效性，我们在每个 Transformer 块之前引入 NDP，而非仅将其集成到注意力和前馈网络设计中。这种做法带来了约 1dB 的峰值信噪比提升（表 5（f）与表5（e））。图 10 展示了视觉示例，其中使用 NDP 引导 Transformer 的学习，有助于生成视觉效果出色且色彩更加自然的结果。

4.3.2 对超分辨率引导重建的影响

UHDPromer的一项关键创新是整合了超分辨率引导重建。为评估其影响，我们还进行了两项补充实验。在

表 5 神经判别提示 Transformer 的消融实验。在注意力和前馈网络中均引入 NDP 会产生积极效果。

ID	实验	超分辨率分支		主分支		参数 ↓
		峰值信噪比 ↑	结构相似性 ↑	峰值信噪比 ↑	结构相似性 ↑	
(a)	NDPA 和 NDPN 中不含 NDP	26.487	0.8974	26.811	0.9282	0.5322M
(b)	NDPA 中不使用 NDP	25.962	0.8952	26.183	0.9252	0.7348M
(c)	NDPN 中不使用 NDP	26.745	0.8984	27.026	0.9283	0.5404M
(d)	NDP→ 直接特征	25.999	**0.9021**	26.138	0.9272	0.7430M
(e)	在 NDPTB 之前使用 NDP	25.790	0.8974	26.161	0.9263	0.5322M
(f)	**完整模型** (我们的)	**26.859**	0.8985	**27.159**	**0.9285**	0.7430M

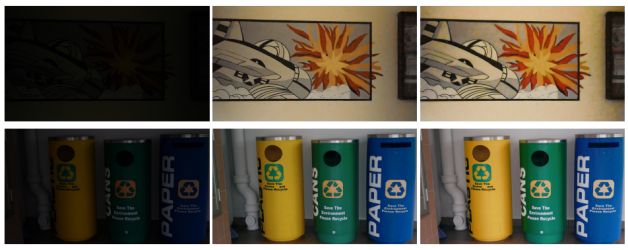
表6 超分辨率引导重建的影响。引入的超分辨率引导重建显著提升了结构相似性方面的恢复质量。

ID	实验	超分辨率分支		主分支		参数 ↓
		峰值信噪比 ↑	结构相似性 ↑	峰值信噪比 ↑	结构相似性 ↑	
(a)	级联重建	-	-	26.375	0.9043	0.7414M
(b)	不含（5）中的第二项	-	-	27.125	0.9051	0.7425M
(c)	含超分辨率分支（我们的方法）	**26.859**	**0.8985**	**27.159**	**0.9285**	0.7430M



(a) Input (b) w/o NDP **(c) Ours**

图 10 神经判别先验的视觉效果。使用 NDP 能够帮助生成视觉上更悦目、具有更多自然色彩的结果。



(a) Input (b) Cascaded **(c) Ours**

图 11 超分辨率引导重建的视觉效果。所提出的超分辨率引导重建能够改善视觉结果，使色彩更加鲜艳。

第一项实验中，我们省略超分辨率分支，将混洗上采样特征直接传入主分支，并使用5个ConvNeXt-v2模块（Woo等人，2023年）（其数量等于主分支和超分辨率分支中ConvNeXt-v2模块的总数）来重建最终结果。

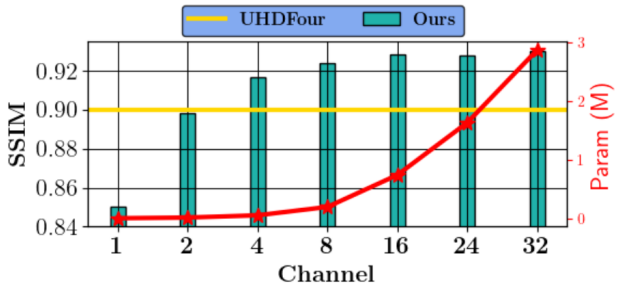


图 12 通道数的影响。我们的模型仅使用 2 通道和 0.0168M 参数，即可达到与 UHDFour（Li 等人，2023年）相当的性能，同时模型规模大约小 ***1,000*** 倍（UHDFour 具有 17.5M 参数，如表7所示）。

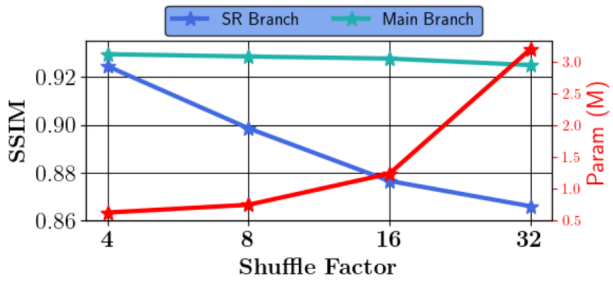


图 13 洗牌下采样因子的影响。值得注意的是，尽管使用洗牌下采样因子为 4 的模型相比因子为 8 的基线模型获得了更大的感受野，但结构相似性并无显著提升。

这种方法称为级联重建，可确保比较公平。在第二个实验中，

表7 参数量比较。与先进方法相比，我们的 UHDPromer 显著减少了参数量，但略高于 UHDformer。

方法	LMAR (Yu 等人，2024)	UHD (Zheng 等人，2021年)	UHDFour (Li 等人，2023年)	FFTformer (Kong 等人，2023年)	DehazeFormer (Song等人，2023年)	UHDformer (Wang 等人，2024b)	**UHDPromer** (本文方法)
参数量（M）	1.314	34.5	17.5	16.6	2.5	**0.3393**	0.7430
缩减	43.5%	97.8%	95.8%	95.5%	70.3%	119.0%	-

表 8 浮点运算次数对比（1024×1024 像素）。我们的 UHDPromer 显著降低了浮点运算次数所体现的计算复杂度，表明该方法具有较高的效率。

方法	LMAR (Yu 等人，2024)	UHD (Zheng 等人，2021年)	UHDFour (Li 等人，2023年)	FFTformer (Kong 等人，2023年)	DehazeFormer (Song等人，2023年)	UHDformer (Wang 等人，2024b)	**UHDPromer** (我们的方法)
浮点运算次数（G）	208.34	113.46	75.63	2107.94	375.40	51.63	**32.56**
缩减	84.4%	71.3%	56.9%	98.5%	91.3%	36.9%	-

表 9 运行时间（RT）（1024 × 1024 像素）在相同的图形处理器上。我们的**UHDPromer** 运行速度快于基于 Transformer 的方法以及基于 CNN 的方法（Yu 等，2024）。

方法	Restormer (Zamir 等人，2022年)	LMAR (Yu 等人, 2024)	DehazeFormer (Song等人,2023年)	FFTformer (Kong 等人, 2023年)	UHDformer (Wang 等人, 2024b)	**UHDPromer** (我们的)
运行时间（秒）	1.86	3.40	0.45	1.27	0.16	**0.12**

未使用超分辨率损失，即式（5）中的第二项。表 6显示，我们的超分辨率引导重建在参数数量相近的情况下，相较于采用级联重建的模型，峰值信噪比提升了 0.784 dB。值得注意的是，省略超分辨率损失（即式（5）中的第二项）会导致性能下降。图 11展示了视觉效果；我们提出的超分辨率引导重建生成的图像色彩更加鲜艳。

4.3.3 通道数的影响

我们还考察了通道数对性能的影响，如图 12 所示。此外，我们将 UHDFour（Li 等人，2023年）的性能作为参考，以确立 UHDPromer 稳健的基线。图 12 表明，随着通道数增加至 16，结构相似性指标逐步提升。通道数超过 16 后，性能没有显著提升。值得注意的是，即使仅使用 2 个通道，我们的模型也能达到与 UHDFour（Li 等人，2023年）相当的性能，而模型大小约为 UHDFour 的 ***1,000*** 分之一，展现了 UHDPromer 在超高清图像增强方面强大的基线性能。

4.3.4 洗牌下采样因子的影响

在实验中，我们采用 8 的洗牌下采样因子来降低分辨率，以便在低分辨率空间中进行学习。此外，我们还分析了其他不同洗牌下采样因子的影响，如图 13 所示。值得注意的是，虽然采用 4 的洗牌下采样因子的模型相比采用 8 的基线模型拥有更大的感受野，但结构相似性并未显著提升。相比之下，我们采用 8 的洗牌下采样因子的模型，其结构相似性表现优于采用更大洗牌下采样因子的模型。

4.3.5 神经差异先验的展示

从图14中，我们展示了第一层和最后一层的神经差异先验（NDP）。显然可以观察到，我们的 NDP 能有效衡量结构位于较浅层还是较深层，这表明 NDP 是一种有效的先验，可作为判别函数，帮助在低分辨率域中学习更具表征性的特征。

4.3.6 计算复杂度

表 7展示了若干先进方法的参数量对比。与一些先进方法相比，我们的 UHDPromer 显著减少了参数量，至少降低了 70.3%，但参数量略高于 UHDformer。表 8 展示了浮点运算次数的对比。与基于 Transformer 的方法（例如 Restormer（Zamir 等人，2022年）、FFTformer（Kong 等人，2023年）和 DehazeFormer（Song等人，2023年）相比，我们的 UHDPromer 至少减少了 91.3% 的浮点运算次数。另一方面，与 UHD（Zheng 等人，2021年）、UHDFour（Li 等人，2023年）、LMAR Yu 等人 (2024) 和 UHDformer（Wang 等人，2024b）相比，这些方法均专为超高清图像设计，而我们的 UHDPromer 至少减少了 36.9%



(a) Input (b) NDP at first layer (c) NDP at last layer (d) Input (e) NDP at first layer (f) NDP at last layer

图 14 首层和末层的神经判别先验示意图（NDP）。我们的 NDP 能更好地衡量首层和末层的显著结构，可作为判别函数，帮助 Transformer 的学习，从而促进图像复原。

浮点运算次数。表9展示了运行时间（RT）对比，我们的 UHDPromer 运行速度显著快于基于 Transformer 的方法。值得注意的是，与基于卷积神经网络的方法 LMAR（Yu 等人，2024）相比，我们的 UHDPromer 在模型大小、计算复杂度和运行时间方面均显著优于该方法，进一步体现了我们方法的有效性和高效性。

4.4 通用图像复原与增强结果

我们的 UHDPromer 专为超高清（UHD）图像设计，其大部分运算都在低分辨率空间中执行。或许有人会质疑，UHDPromer 是否也适用于通用图像复原与增强任务。为此，我们将 UHDPromer 与标准图像复原和增强方法进行了比较分析，测试涵盖多个通用图像复原与增强基准，包括 LOL 上的低光照图像增强（Wei 等人，2018年）、SOTS-ITS 上的图像去雾（Li 等人，2019a）以及 GoPro 上的图像去模糊（Nah 等人，2017年）。

4.4.1 低光照图像增强结果

表 10 展示了使用 LOL 数据集进行低光照图像增强的对比结果，参考文献为（Wei 等人，2018年）。尽管我们的 UHDPromer 专为超高清（UHD）图像设计，主要在低分辨率空间中运行，但如广受认可的 LOL 基准（Wei 等人，2018年）所示，我们的 UHDPromer 在保持卓越计算效率的同时展现出了具有竞争力的性能，相较于现有方法更具优势。根据峰值信噪比指标，我们的 UHDPromer 也略优于 UHDformer（Wang 等人，2024b），但性能低于针对常规图像分辨率设计的专用低光照增强方法。例如，UHDPromer 的性能低于 LLFlow（Wang 等人，2022b），后者包含 17.4M 个参数，而 UHDPromer 仅需 0.743M 个参数——约比 LLFlow 少 23.7 倍。同样，尽管我们的方法取得了与 LLFormer（Wang 等人，2023b）相当的结构相似性性能，但其参数量减少了 94.4%。值得注意的是，LLFormer 在不进行额外预处理的情况下无法直接处理超高清图像，而我们的 UHDPromer 能够原生处理超高清输入，并且仅产生极小的计算开销。这一性能差距表明，尽管我们的方法在计算效率和超高清处理能力方面表现出色，但在基准测试所采用的常规图像尺寸下，其恢复质量与面向特定任务的架构相比仍有提升空间。

4.4.2 图像去雾结果

表。11 展示了我们在图像去雾方面的对比研究，具体针对 SOTS-Indoor 数据集（Li 等人，2019a）。尽管 UHDPromer 专为 UHD 图像设计，且主要侧重于低分辨率空间操作，但与专门针对图像去雾设计的通用方法相比，其性能较差。值得注意的是，UHDPromer 仅需0。7430M 参数量，显著少于其他最先进的方法。在图像去雾的复原性能方面，UHDPromer 在 SOTS-Indoor 数据集（Li 等人，2019a）上的表现并不理想，用于图像——

LOL 上的低光照图像增强（Wei 等人，2018年）。尽管我们的 **UHDPromer**专为 UHD 图像设计，且大多数操作都在低分辨率空间中进行，但在通用图像增强基准（Wei等人，2018年）上，我们的 UHDPromer 仍取得了最佳性能。“UHDFour-2/8”指下采样因子为 2/8 的 UHDFour 变体。相反，“UHDFour-8”表示采用下采样因子 8 的 UHDFour 版本，专为增强超高清低光照图像而设计。

方法	发表场所	峰值信噪比 ↑	结构相似性 ↑	参数 ↓
Retinex-Net（Guo等人，2020年）	BMVC'18	16.77	0.54	0.45M
Zero-DCE（Guo等人，2020年）	CVPR'20	16.79	0.67	79.416K
AGLLNet (吕等，2021年)	IJCV'21	17.52	0.77	3.438K
Zhao等 (Zhao等，2021年)	ICCV'21	21.67	0.87	11.560M
RUAS（Liu 等人，2021a）	CVPR'21	16.44	0.70	3K
SCI（马等人，2022年）	CVPR'22	14.78	0.62	0.258K
URetinex-Net（吴等人，2022年）	CVPR'22	19.84	0.87	1.5M
LLFlow（Wang 等人，2023b）	AAAI'22	28.99	0.94	17.4M
LLFormer（Wang 等人，2023b）	AAAI'23	27.22	0.90	13.2M
UHDFour-2（Li 等人，2023年）	ICLR'23	23.09	0.87	28.518M
UHDFour-8（Li 等人，2023年）	ICLR'23	26.65	0.87	17.537M
UHDformer (Wang 等人, 2024b)	AAAI'24	25.36	0.91	0.3393M
UHDPromer	-	26.64	0.91	0.7430M

表 11 在 SOTS-Indoor 上的比较（Li 等人，2019年a）。尽管我们的**UHDPromer**是为 UHD 图像设计的，且大多数操作都在低分辨率空间中执行，但我们的方法在结构相似性方面与通用图像去雾方法相当，而峰值信噪比则较低。

方法	发表场所	峰值信噪比 ↑	结构相似性 ↑	参数 ↓
GridNet（Liu 等人，2019年）	ICCV'19	33.8244	0.9924	0.96M
MSBDN（Dong 等人，2020年）	CVPR'20	34.4219	0.9899	31.4M
UHD (Zheng 等人,2021年)	CVPR'21	18.7488	0.8486	34.5M
PFDN (Dong 和 Pan, 2020年)	ECCV'20	31.2308	0.9853	12M
D4 (Yang 等人,2022年)	CVPR'22	20.5449	0.9236	22.9M
DeHamer（Guo等人，2022年）	CVPR'22	38.2672	0.9936	132.4M
DehazeFormer（Song等人，2023年）	TIP'23	39.9100	0.9960	25.5M
UHDformer（Wang 等人，2024b）	AAAI'24	35.7413	0.9920	0.3393M
UHDPromer	-	32.4778	0.9892	0.7430M

表12 GoPro 上的比较 (Nah 等人，2017年)。我们的 UHDPromer 在处理通用图像去模糊基准（如 GoPro（Nah 等人，2017年）方面效果较差。请注意，由于我们使用常用的 IQA PyTorch 工具箱（https://github.com/chaofengc/IQA-PyTorch）重新计算了本文所有基准的结果，因此峰值信噪比和结构相似性可能与现有论文中的数值不同。

方法	发表场所	峰值信噪比 ↑	结构相似性 ↑	参数 ↓
SRN（Tao等，2018年）	CVPR'18	31.751	0.9144	6.8M
DMPHN（Zhang等，2019年）	CVPR'19	31.734	0.9130	21.7M
MIMO-Unet++(Cho 等人，2021年)	ICCV'21	33.277	0.9348	16.1M
MPRNet(Zamir 等人，2021年)	CVPR'21	34.195	0.9454	20.1M
Restormer（Zamir 等人，2022年）	CVPR'22	34.483	0.9488	26.1M
Uformer（Wang 等人，2022c）	CVPR'22	34.668	0.9502	20.6M
Stripformer(Tsai 等人, 2022年)	ECCV'22	34.678	0.9506	19.7M
FFTformer (Kong 等人, 2023年)	CVPR'23	35.798	0.9594	16.6M
UHDformer (Wang 等人，2024b)	AAAI'24	29.407	0.8660	0.3393M
UHDPromer	-	29.487	0.8694	0.7430M

age 与 UHDformer (Wang 等人，2024b) 相比的去雾效果。这种性能模式进一步证实，我们的方法架构设计虽然针对超高清处理效率进行了优化，但可能无法充分捕捉普通分辨率图像在各种复原任务中呈现的复杂退化模式。

4.4.3 动态图像去模糊结果

尽管我们的 UHDPromer 擅长处理超高清（UHD）图像，但它也存在局限性。值得注意的是，UHDPromer 在通用图像去模糊任务中表现不佳，例如处理 GoPro数据集中的图像（Nah

等人，2017年），详见表 12。这一局限性可归因于几个因素：首先，通用图像去模糊任务通常需要较大的模型，才能有效处理复杂的模糊。然而，UHDPromer 的参数数量仅为 0.7430M，相对较小，可能不足以充分处理复杂的运动模糊。其次，要取得更好的效果，有效的去模糊通常要求模型在更大的空间特征上执行大量操作。相比之下，UHDPromer 专为超高清图像设计，主要处理 8×下采样特征。这一设计选择显著影响了它在更通用场景中的去模糊性能。此外，在动态去模糊基准测试（Wang 等人，2024b）中，我们的 UHDPromer 略优于 UHDformer（Gong 等人，2017年）。这些比较进一步凸显了 UHDPromer 的局限性：它不适合处理通用尺寸的图像。我们的 UHDPromer 仅在超高清图像尺寸下表现良好，这限制了它在更多场景中的进一步应用。

5 结论

我们提出了一种简单而有效的UHDPromer，将所开发的、蕴含于高分辨率特征与低分辨率特征之间的神经判别先验融入Transformer设计中，用于超高清图像复原与增强。为了更好地学习低分辨率特征，我们提出了一种简洁而有效的神经判别提示变换器块，该模块由所开发的NDP引导，包含神经判别提示注意力（NDPA）和神经判别提示网络（NDPN）。NDPA将NDP有效融入注意力机制，以从全局视角感知有用的判别信息。NDPN用于探索由NDP引导的连续门控机制，从而允许传递更多有益信息。为了更好地重建结果，我们提出了一种超分辨率引导的重建方法。大量实验表明，在包括超高清低光照图像增强、超高清图像去雾和超高清图像去模糊在内的3项超高清图像复原与增强任务中，在不同训练设置下，我们的UHDPromer均优于现有最先进方法。额外实验还表明，与通用图像复原方法相比，我们提出的UHDPromer不擅长处理常规图像尺寸。这一缺陷限制了我们的方法同时处理常规图像尺寸和超高清图像尺寸的能力，进而限制了其更广泛的应用。

未来，我们将重点研究如何在保持一般图像尺寸复原性能的同时，处理 UHD 图像尺寸的潜在方法。

数据可用性声明。本文中使用的所有数据均已公开。源代码和预训练模型也将在以下网址公开：https://github.com/supersupercong/uhdpromer。

参考文献

Anwar S、Khan S、Barnes N（2019）深入探索超分辨率：一项综述。ACM Computing Surveys Ba JL、Kiros JR、Hinton GE（2016）层归一化。arXiv:160706450 Bai H、Pan J（2024）自监督深度盲视频超分辨率。IEEE模式分析与机器智能汇刊 Berman D、Treibitz T、Avidan S（2016）非局部图像去雾。载于：CVPR，第1674–1682页 Cai B、Xu X、Jia K、Qing C、Tao D（2016）Dehazenet：用于单幅图像去雾的端到端系统。TIP 25(11):5187–5198 Chen H、Wang Y、Guo T、Xu C、Deng Y、Liu Z、Ma S、Xu C、Xu C、Gao W（2021）预训练图像处理Transformer。载于：CVPR Chen Y、Hsu C（2013）一种用于时空相关雨纹的广义低秩外观模型。载于：ICCV，第1968–1975页 Cho SJ、Ji SW、Hong JP、Jung SW、Ko SJ（2021）重新思考单幅图像去模糊中的由粗到精方法。载于：ICCV Deng S、Ren W、Yan Y、Wang T、Song F、Cao X（2021a）用于超高清视频去模糊的多尺度可分离网络。载于：ICCV，第14010–14019页 Deng S、Ren W、Yan Y、Wang T、Song F、Cao X（2021b）用于超高清视频去模糊的多尺度可分离网络。载于：ICCV，第14030–14039页 Dong H、Pan J、Xiang L、Hu Z、Zhang X、Wang F、Yang M（2020）采用密集特征融合的多尺度增强去雾网络。载于：CVPR，第2154–2164页 Dong J、Pan J（2020）基于物理的特征去雾网络。载于：ECCV，第188–204页 Dosovitskiy A、Beyer L、Kolesnikov A、Weissenborn D、Zhai X、Unterthiner T、Dehghani M、Minderer M、Heigold G、Gelly S 等（2021）一张图像胜过16×16个词：用于大规模图像识别的Transformer。载于：ICLR Fattal R（2014）利用颜色线去雾。ACM TOG 34(1):13:1–13:14 Fu X、Huang J、Zeng D、Huang Y、Ding X、Paisley J（2017）通过深度细节网络去除单幅图像中的雨纹。载于：CVPR Gong D、Yang J、Liu L、Zhang Y、Reid I、Shen C、Van Den Hengel A、Shi Q（2017）从运动模糊到运动流：一种去除异质运动模糊的深度学习解决方案。载于：CVPR Guo C、Li C、Guo J、Loy CC、Hou J、Kwong S、Cong R（2020）用于低光照图像增强的零参考深度曲线估计。载于：CVPR，第1777–1786页

Guo CL、Yan Q、Anwar S、Cong R、Ren W、Li C（2022）采用传输感知三维位置嵌入的图像去雾变换器。载于：CVPR，第5812–5820页 He K、Sun J、Tang X（2011）使用暗通道先验进行单幅图像去雾。IEEE模式分析与机器智能汇刊 33(12):2341–2353 He K、Zhang X、Ren S、Sun J（2016a）用于图像识别的深度残差学习。载于：CVPR，第770–778页 He K、Zhang X、Ren S、Sun J（2016b）深度残差网络中的恒等映射。载于：ECCV，第630–645页 Huang G、Liu Z、van der Maaten L、Weinberger KQ（2017）密集连接卷积网络。载于：CVPR，第2261–2269页 Huang S、Cheng F、Chiu Y（2013）利用加权分布自适应伽马校正进行高效对比度增强。TIP 22(3):1032–1041 Huynh-Thu Q、Ghanbari M（2008）峰值信噪比在图像／视频质量评估中的适用范围。Electronics Letters 44(13):800–801 Kong L、Dong J、Ge J、Li M、Pan J（2023）用于高质量图像去模糊的高效频域Transformer。载于：CVPR，第5886–5895页 Krizhevsky A、Sutskever I、Hinton GE（2012）使用深度卷积神经网络进行ImageNet分类。载于：神经信息处理系统大会 Kupyn O、Budzan V、Mykhailych M、Mishkin D、Matas J（2018）DeblurGAN：使用条件对抗网络进行盲运动去模糊。载于：CVPR，第8183–8192页 Li B、Ren W、Fu D、Tao D、Feng D、Zeng W、Wang Z（2019a）单幅图像去雾及相关任务的基准评测。TIP 28(1):492–505 Li C、Guo CL、Zhou M、Liang Z、Zhou S、Feng R、Loy CC（2023）通过嵌入傅里叶实现超高清低光照图像增强。载于：ICLR Li R、Pan J、Li Z、Tang J（2018）通过条件生成对抗网络实现单幅图像去雾。载于：CVPR，第8202–8211页 Li S、Araujo IB、Ren W、Wang Z、Tokuda EK、Junior RH、Cesar-Junior R、Zhang J、Guo X、Cao X（2019b）单幅图像去雨：综合基准分析。载于：CVPR Li Y、Tan RT、Guo X、Lu J、Brown MS（2016）利用层先验去除雨纹。载于：CVPR，第2736–2744页 Liang J、Cao J、Sun G、Zhang K、Van Gool L、Timofte R（2021）SwinIR：使用Swin Transformer进行图像复原。载于：国际计算机视觉会议研讨会 Liu R、Ma L、Zhang J、Fan X、Luo Z（2021a）受Retinex启发的展开方法，结合协同先验架构搜索实现低光照图像增强。载于：CVPR，第10561–10570页 Liu X、Ma Y、Shi Z、Chen J（2019）GridDehazeNet：用于图像去雾的基于注意力的多尺度网络。载于：ICCV，第7313–7322页 Liu Z、Lin Y、Cao Y、Hu H、Wei Y、Zhang Z、Lin S、Guo B（2021b）Swin Transformer：使用移位窗口的分层视觉Transformer。arXiv：210314030 Liu Z、Hu H、Lin Y、Yao Z、Xie Z、Wei Y、Ning J、Cao Y、Zhang Z、Dong L、Wei F、Guo B（2022a）Swin Transformer v2：扩展容量与分辨率。载于：CVPR，第12009–12019页 Liu Z、Mao H、Wu CY、Feichtenhofer C、Darrell T、Xie S（2022b）面向2020年代的卷积网络。CVPR Liu Z、Ning J、Cao Y、Wei Y、Zhang Z、Lin S、Hu H（2022c）视频Swin Transformer。载于：CVPR，第3202–3211页 Loshchilov I、Hutter F（2017）SGDR：带热重启的随机梯度下降。载于：ICLR Loshchilov I、Hutter F（2019）解耦权重衰减正则化。载于：ICLR Luo Y、Xu Y、Ji H（2015）通过判别式稀疏编码去除单幅图像中的雨。载于：ICCV，第3397–3405页 吕 F、Li Y、Lu F（2021）利用大规模低光仿真数据集的注意力引导低光照图像增强。IJCV 129(7)：2175–2193 马 L、Ma T、Liu R、Fan X、Luo Z（2022）迈向快速、灵活且稳健的低光照图像增强。载于：CVPR，第5637–5646页 Nah S、Hyun Kim T、Mu Lee K（2017）用于动态场景去模糊的深度多尺度卷积神经网络。载于：CVPR Pan J、Sun D、Pfister H、Yang MH（2016）使用暗通道先验进行盲图像去模糊。载于：CVPR Pan J、Hu Z、Su Z、Yang MH（2017）用于文本图像去模糊及其他任务的 l_0 -正则化强度与梯度先验。IEEE模式分析与机器智能汇刊 39(2)：342–355 Pan J、Liu S、Sun D、Zhang J、Liu Y、Ren JSJ、Li Z、Tang J、Lu H、Tai Y、Yang M（2018）学习用于底层视觉的双卷积神经网络。载于：CVPR，第3070–3079页 Pan J、Dong J、Liu Y、Zhang J、Ren JSJ、Tang J、Tai Y、Yang M（2021）用于图像复原及其他任务的基于物理的生成对抗模型。IEEE模式分析与机器智能汇刊 Pan J、Sun D、Zhang J、Tang J、Yang J、Tai YW、Yang MH（2022）用于底层视觉的双卷积神经网络。IJCV 130(6)：1440–1458 Pan J、Xu B、Bai H、Tang J、Yang MH（2023）利用时间锐度先验与非局部时空相似性的级联深度视频去模糊。IEEE模式分析与机器智能汇刊 Ren W、Liu S、Zhang H、Pan J、Cao X、Yang M（2016）通过多尺度卷积神经网络进行单幅图像去雾。载于：ECCV，第154–169页 Ren W、Zhang J、Pan J、Liu S、Ren JS、Du J、Cao X、Yang MH（2021）通过空间变化的循环神经网络对动态场景进行去模糊。IEEE模式分析与机器智能汇刊 44(8)：3974–3987 Song Y、He Z、Qian H、Du X（2023）用于单幅图像去雾的视觉Transformer。TIP 32：1927–1941

Sun L、Pan J、Tang J（2022）Shufflemixer：用于图像超分辨率的高效卷积网络。NeurIPS 35：17314–17326 Tao X、Gao H、Shen X、Wang J、Jia J（2018）用于深度图像去模糊的尺度循环网络。载于：CVPR Tian C、Fei L、Zheng W、Xu Y、Zuo W、Lin CW（2020）图像去噪中的深度学习：综述。神经网络 Tsai FJ、Peng YT、Lin YY、Tsai CC、Lin CW（2022）Stripformer：用于快速图像去模糊的条带变换器。载于：ECCV Wang C、Xing X、Wu Y、Su Z、Chen J（2020）DCSFN：用于单幅图像去雨的深度跨尺度融合网络。载于：ACM MM，第1643–1651页 Wang C、Pan J、Wu X（2022a）用于单幅图像去雨的在线更新高阶协作网络。载于：AAAI，第2406–2413页 Wang C、Pan J、Wang W、Dong J、Wang M、Ju Y、Chen J（2023a）Promptrestorer：一种具有退化感知的提示图像复原方法。载于：NeurIPS Wang C、Pan J、Lin W、DongJ、Wang W、Wu X（2024a）Self-promer：具有深度一致性的自提示去雾变换器。载于：AAAI，第5327–5335页 Wang C、Pan J、Wang W、Fu G、Liang S、Wang M、Wu X、Liu J（2024b）用于UHD图像复原的相关性匹配变换Transformer。载于：AAAI，第5336–5344页 Wang C、Wang W、Yu C、Mu J（2024c）利用图神经网络探索单幅图像去雨中的内部与外部相似性。载于：IJCAI，第1371–1379页 Wang L、Wang C、Pan J、Zhou W、Sun X、Wang W、Su Z（2024d）超高清复原：新基准与一种双重交互先验驱动方案。arXiv预印本，arXiv：240613607 Wang T、Zhang K、Shen T、Luo W、Stenger B、Lu T（2023b）超高清低照度图像增强：基准与基于变换器的方法。载于：AAAI Wang Y、Wan R、Yang W、Li H、Chau LP、Kot AC（2022b）利用归一化流进行低光照图像增强。载于：AAAI Wang Z、Bovik AC、Sheikh HR、Simoncelli EP（2004）图像质量评价：从误差可见性到结构相似性。TIP 13(4)：600–612 Wang Z、Cun X、Bao J、Liu J（2022c）Uformer：用于图像复原的通用U形变换器。载于：CVPR Wei C、Wang W、Yang W、Liu J（2018）用于低光照增强的深度Retinex分解。载于：BMVC，第155页 Woo S、Debnath S、Hu R、Chen X、Liu Z、Kweon IS、Xie S（2023）Convnext v2：利用掩码自编码器协同设计并扩展卷积网络。载于：CVPR，第16133–16142页 Wu W、Weng J、Zhang P、Wang X、Yang W、Jiang J（2022）URetinex-Net：用于低光照图像增强的基于Retinex的深度展开网络。载于：CVPR，第5901–5910页 Xie S, Girshick R, Doll´ar P, Tu Z, He K（2017年）用于深度神经网络的聚合残差变换。在：CVPR，第1492–1500页 Yang W, Tan RT, Feng J, Liu J, Guo Z, Yan S（2017年）基于单幅图像的深度联合雨滴检测与去除。在：CVPR，第1685–1694页 Yang Y, Wang C, Liu R, Zhang L, Guo X, Tao D（2022年）通过密度与深度分解实现自增强非配对图像去雾。在：CVPR，第2037–2046页 Yu W, Huang J, Li B, Zheng K, Zhu Q, Zhou M, Zhao F（2024）利用模型感知引导增强超高清图像的重采样操作。在：CVPR，第25722–25731页 Zamir SW, Arora A, Khan S, Hayat M, Khan FS, Yang MH, Shao L（2021年）多阶段渐进式图像复原。在：CVPR Zamir SW, Arora A, Khan S, Hayat M, Khan FS, Yang MH（2022年）Restormer：用于高分辨率图像复原的高效Transformer。在：CVPR，第5718–5729页 Zhang H, Dai Y, Li H, Koniusz P（2019年）用于图像去模糊的深度堆叠分层多块网络。在：CVPR Zhang K, Zuo W, Chen Y, Meng D, Zhang L（2017年）超越高斯去噪器：用于图像去噪的深度卷积神经网络残差学习。TIP Zhang Y, Li K, Li K, Wang L, Zhong B, Fu Y（2018年）使用极深残差通道注意力网络进行图像超分辨率。在：ECCV Zhao L, Lu S, Chen T, Yang Z, Shamir A（2021年）利用循环注意力学习实现欠曝图像增强的深度对称网络。在：ICCV，第12055–12064页 Zheng Z, Ren W, Cao X, Hu X, Wang T, Song F, Jia X（2021年）通过多引导双边学习实现超高清图像去雾。在：CVPR，第16185–16194页