用于全景图像超分辨率的多级失真感知可变形网络

Cuixin Yang[a], Rongkang Dong[a], Kin-ManLam[a], YuhangZhang[b] 和GuopingQiu[c]

[a]香港理工大学电机及电子工程学系，中国香港[b]广州大学计算机科学与网络工程学院，中国广州[c]诺丁汉大学计算机科学学院，英国诺丁汉

文章信息

关键词：全景图像
可变形网络 失真 超分辨率

摘 要

随着增强现实和虚拟现实应用日益普及，全景图像（ODIs）的图像处理受到越来越多的关注。全景图像超分辨率（ODISR）是一种提升全景图像视觉质量的有前景技术。在进行超分辨率处理之前，通常会使用等距圆柱投影（ERP），将全景图像从球面投影到平面上。这种投影会在ERP图像中引入随纬度变化的几何畸变：赤道附近的畸变最小，而越靠近极点，畸变越严重，图像内容会被拉伸到更大的区域。然而，现有的全向图像超分辨率方法的采样范围和特征提取能力有限，难以捕捉大范围区域内的失真模式。为解决这一问题，我们提出了一种用于全景图像超分辨率的新型多层级畸变感知可变形网络（MDDN），旨在扩大采样范围和感受野。具体而言，MDDN中的特征提取器由三个并行分支组成：一个可变形注意力机制（作为膨胀率=1 路径），以及两个膨胀率分别为2和3的空洞可变形卷积。该架构扩大了采样范围，使其能够覆盖更广区域内的更多失真模式，从而生成密集且全面的特征，有效捕捉ERP图像中的几何畸变。通过多层级特征融合模块，对这些可变形特征提取器提取的表征进行自适应融合。此外，为降低计算成本，我们对空洞可变形卷积采用了低秩分解策略。在公开数据集上进行的大量实验表明，MDDN的表现优于最先进的方法，凸显了其在全景图像超分辨率方面的有效性和优越性。

arXiv:2512.17343v1 [cs.CV] 19 Dec 2025

1. 引言

全景图像（ODI），也称为 360°图像或全景图像，是增强现实（AR）和虚拟现实（VR）应用的基础[33]。随着 AR/VR 应用日益普及，ODI 的处理变得愈发重要，因为 ODI 的质量会显著影响用户在 AR/VR 环境中的沉浸式体验。ODI 通常使用鱼眼镜头拍摄，将场景投射到两个半球上。尽管高分辨率全景图像对于呈现逼真、细致的视觉效果至关重要，但光学成像系统成本高昂，加上存储和带宽方面的限制，往往使 ODI 的分辨率受到限制，显著降低沉浸式应用的视觉质量。

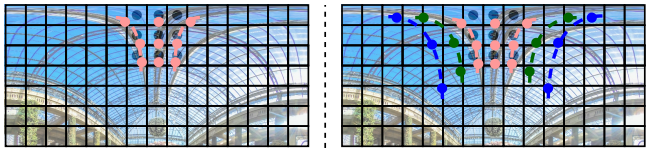


图 1：以往单层（左）与本文提出的多层畸变感知可变形特征提取器（右）的采样示意图。规则网格的中心表示固定规则形状卷积核的采样位置。由于 ERP 图像存在几何失真，图像内容（尤其是高纬度地区的内容）会被大幅拉伸。以往方法采用的单层可变形特征提取器采样范围有限，难以捕捉多样的模式。相比之下，本文提出的多层可变形特征提取器具有更大的采样范围，能够更好地适应失真，并更有效地提取高度失真区域。

图像超分辨率（SR） [53, 44, 13, 43, 42]是解决这一问题的一项有前景的技术。近年来，研究人员专注于将超分辨率方法应用于获取高分辨率全景图像 [16, 51, 48, 49]。为了便于存储和传输，ODI 通常会被投影为等距圆柱投影（ERP）图像，因为其其简单性以及与现有图像处理流程的兼容性。因此，全景图像超分辨率（ODISR）通常在 ERP图像上进行。

在等距柱状投影过程中，整个球面场景会被映射到一个二维平面上，将球面坐标（即经度和纬度）投影到矩形网格中。这种方法可在单张连续的平面ERP图像中保留球面全景图像的 360° × 180° 视场（FoV）。然而，这种投影会在ERP图像中引入显著且不均匀的几何畸变，且这种畸变会随纬度

cuixin.yang@connect.polyu.hk (C. Yang); rongkang97.dong@connect.polyu.hk (R. Dong); enkmlam@polyu.edu.hk (K. Lam); yuhang.zhang@gzhu.edu.cn (Y. Zhang); guoping.qiu@nottingham.ac.yk (G. Qiu) ORCID(s):

而变化。具体而言，极地区域的水平拉伸最为严重，而赤道处的失真最小，甚至微不足道。直接将二维平面图像超分辨率方法应用于 [36, 11, 23] ERP图像会忽略几何畸变，导致超分辨率性能不佳。同时，如图1所示，固定且形状规则的卷积核难以适应失真，因此也不适用于全景图像超分辨率重建。

当前全向图像超分辨率（ODISR）方法 [16, 50, 51, 49] 将失真视为重要先验，并设计网络以利用几何畸变。OSRT [51] 和 GDGT-OSR [49] 采用可变形注意力 Transformer [46]，利用不规则形状卷积核提取特征。与固定形状卷积核相比，这类卷积核更能适应失真。尽管不规则形状卷积核具有强大的几何变换能力，但其采样范围只有单级，如图左侧所示。1，因此 OSRT [51] 和 GDGT-OSR [49] 仍难以适应随纬度变化的失真。例如，高纬度地区的场景会大幅拉伸，沿水平方向扩展。在这种情况下，以往全向图像超分辨率方法 [51, 49] 所用的单级采样范围可变形特征提取器，捕获扩展失真场景的采样能力有限。因此，单级采样范围的可变形特征提取器可能只能覆盖规则卷积核位置周围少量稀疏的失真点，而忽略对有效图像重建至关重要的大量严重失真上下文信息。如图 1所示，屋顶的钢结构被显著拉伸，分布范围更广；然而，单级失真感知可变形特征提取器的采样范围仍局限于固定卷积核，无法充分捕捉更广的失真区域。为解决这一问题，我们提出多级失真感知可变形网络（MDDN），其中包含多级失真感知可变形特征提取器（MDDE），以扩大采样范围并捕捉更多相似的失真模式。图 1展示了所提出的 MDDE 如何扩大采样范围，以涵盖扩展至更广区域的模式。具体而言，MDDE 由失真感知可变形注意力机制和具有不同膨胀率的膨胀失真感知可变形卷积组成。膨胀率为 1 的可变形注意力机制充当第 1 级特征提取器。膨胀率为 2 和 3 的膨胀失真感知可变形卷积分别充当第 2 级和第 3 级特征提取器。这里，第i 级对应于在 Transformer 或卷积层中使用膨胀率 i，它决定了卷积核的采样范围。级别越高，特征提取器的采样范围越大，反之亦然。第 1 级特征提取器捕捉基础特征，而第 2 级和第 3 级特征提取器则利用更大采样范围和感受野中的特征对其进行补充。借助这种多级结构，可以捕捉到更多自相似模式和纹理，从而有助于提升超分辨率性能。从全局—局部视角来看，失真感知可变形注意力机制利用其建模长程依赖的能力，为全景图像形成全局表示。膨胀可变形卷积采用具有不同扩展采样范围的膨胀卷积核，提取密集且互补的局部表示。此外，这些表示在多层特征融合（MFF）模块中自适应协同。为提高模型效率，我们采用低秩分解策略对可变形卷积中的权重进行分解，以降低计算成本。

本文的主要贡献总结如下:

- 我们提出了一种用于全景图像超分辨率的新型多级失真感知可变形网络（MDDN），该网络能够以更大的采样范围和感受野提取特征，从而更好地适应大范围分布的失真模式。
- 所提出的多层级畸变感知可变形特征提取器（MDDE）由失真感知可变形注意力机制和具有不同膨胀率的膨胀畸变感知可变形卷积组成。这些组件构成了具有不同采样范围的多层级表示，通过融合全局与局部特征，有效应对ERP图像的几何畸变。随后，在所提出的多层特征融合（MFF）模块中对这些多层级表示进行自适应融合。
- 为降低计算成本，我们采用低秩分解策略对空洞可变形卷积中的权重进行分解。
- 大量实验结果表明，MDDN在定量和定性评估方面均优于最先进水平（SOTA）的二维平面图像超分辨率方法和全向图像超分辨率方法，证明了所提方法的有效性和优越性。

2. 相关工作

2.1. 图像超分辨率

图像超分辨率 [35, 31, 25]旨在从低分辨率（LR）输入中重建高分辨率（HR）图像，这是计算机视觉中的一项经典低层任务。早期的超分辨率方法依赖插值 [54, 34, 20] 或浅层卷积神经网络（卷积神经网络） [17]。自2017年以来，Transformer [38, 27] 在包括低层计算机视觉任务在内的许多计算机视觉任务中取得了出色的表现 [18, 22]，[6, 41, 45, 2, 37, 24]。Transformer 凭借建模长程依赖的能力，主导了超分辨率领域 [29, 52, 12, 8, 7, 14]。基于分层 Swin Transformer [32], 的 SwinIR [29],采用移位窗口自注意力以提高计算效率，现已成为一种流行的

图像复原架构。然而，SwinIR 的固定大小窗口限制了跨窗口交互，阻碍了全局依赖建模。Restormer [52] 采用多尺度 Transformer，兼顾局部与全局特征提取。CAT [12] 引入矩形窗口自注意力，利用水平和垂直矩形窗口增强全局上下文。HAT [8, 7] 结合通道注意力、自注意力和重叠交叉注意力，在超分辨率任务中实现了更优的感知质量。Swin2SR [14],作为 SwinIR 的演进版本，引入频域注意力以增强纹理恢复。

尽管取得了这些进展，二维平面图像超分辨率方法并未考虑全向图像中存在的各种几何畸变。因此，我们提出了一种专为全向图像超分辨率设计的新型多层级畸变感知可变形网络。

2.2. 全向图像超分辨率

全向图像超分辨率将超分辨率扩展到360° 图像，这类图像通常采用等距圆柱投影或立方体贴图格式表示。等距圆柱投影图像具有非均匀像素密度和几何畸变，尤其是在极点处，因此与平面图像超分辨率相比面临独特挑战。早期的全向图像超分辨率方法 [19, 1] 通过直接微调基于卷积神经网络的超分辨率模型 [19] 或引入畸变感知损失 [1]对其进行改造。然而，这些改造通常难以处理全向图像的球面特性。

渐进式金字塔网络 LAU-Net [16], 旨在突出不同纬度之间非均匀的像素密度。具体而言，LAU-Net 使用不同的网络，对不同纬度带中的像素进行分层超分辨率处理。然而，这种方法计算开销较大，需要进行多层网络训练，导致输出结果中不同纬度带之间存在不一致。为应对投影相关挑战，SphereSR [50] 利用局部隐式图像函数（LIIF） [9] 获得连续球面图像表示。尽管 SphereSR 能够灵活处理各种投影类型的全景图像，但对于每种不同的投影类型，都需要单独训练一个网络。TCCL-Net [5] 提出了一种用于全景图像超分辨率的 Transformer 与卷积协同学习网络，可有效建模短程与长程依赖。该网络还从左右视图探索全景图像超分辨率，将双目视觉信息与全景特性相结合[4]。DiffOSR [30] 提出了一种条件扩散概率模型，并利用几何畸变以及离散小波变换得到的高频子带作为先验条件，以优化去噪过程。BPOSR[39]采用两种全向投影，即等距圆柱投影和立方体贴图，以促进这两种投影之间的交互。OSRT [51]采用可变形注意力机制来利用畸变信息。OmniSSR [28]提出了一种零样本方法，利用扩散模型中的图像先验来实现保真度和真实感。GDGT-OSR [49]通过结合多种具有不同窗口形状的注意力机制来扩大注意力区域。

尽管取得了这些进展，现有方法在处理存在非均匀畸变的等距圆柱投影图像时，仍难以进行特征提取。不同于以往的全景图像超分辨率方法 [51,49],，我们提出的方法采用多个具有多层采样范围的特征提取器，以分层采样的方式获取特征表示，从而有效捕捉等距圆柱投影图像中的畸变内容，并使模型能够有效适应畸变。此外，所提出的方法结合了局部和全局特征提取器，为等距圆柱投影图像构建更全面的表示。

2.3. 可变形网络

可变形网络 [15, 46] 已成为计算机视觉领域一种强大的方法，尤其适用于涉及几何变化和不规则数据分布的任务，例如全向图像超分辨率。与依赖固定网格感受野的标准卷积神经网络不同，可变形网络会根据输入内容，自适应地调整卷积核或注意力机制的采样位置。这种灵活性使其非常适合处理全景图像固有的非均匀像素密度和几何畸变，尤其是等距圆柱投影图像。可变形卷积网络（DCN）[15] 为标准卷积引入可学习偏移量，使卷积核能够根据输入中的几何变换动态调整。EDVR [40] 利用可变形对齐模块处理大幅运动并对齐视频帧。然而，由于卷积网络本身的局限，基于DCN的方法难以捕捉长距离依赖关系，因此在大规模或全局特征提取方面效果较差。为解决这一问题，可变形注意力变换器（DAT） [46] 采用内容自适应注意力模块，根据图像特征动态调整注意力区域。在全向图像超分辨率中，由于全向图像具有球面几何特性，可变形网络尤为重要。OSRT [51] 和GDGT-OSR[49] 均利用可变形注意力机制来适应等距圆柱投影图像中固有的几何畸变。

在本研究中，为了更好地适应几何畸变和非均匀像素分布，我们将失真感知可变形卷积与失真感知可变形注意力 Transformer 相结合，充分挖掘局部和全局特征，从而提升全景图像超分辨率的性能。

3. 方法

3.1. O 中失真的基础知识 DIs

图 2（a）描绘了球体与投影平面之间的几何变换。$d\theta d\varphi$表示一个微元，其面积为 $\delta S(\theta,\varphi)$，位于以(θ,φ)为中心的球体上；而 $dxdy$ 是其对应的微元，其面积为 $\delta P(x,y)$，位于以 (x,y)为中心的投影平面上。球面上的每个点 (θ,φ) 都对应于投影平面上的一个特定点 (x,y)，如下所示

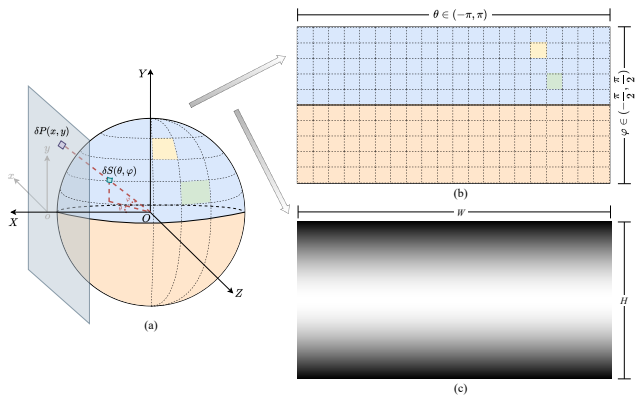


图 2： (a) 球体与投影平面之间的关系。(b) 等距圆柱投影的可视化。全向图像通过等距圆柱投影映射到二维平面上。(c) 失真图。颜色越深表示失真越大，颜色越浅表示失真越小。等距圆柱投影会导致等距圆柱投影图像产生几何畸变。失真程度随纬度升高而加剧。以赤道为对称轴，北半球和南半球的几何畸变呈对称分布。

：

$$\begin{aligned} x &= f(\theta, \varphi) \\ y &= g(\theta, \varphi), \end{aligned} \tag{1}$$

其中 $\theta \in (-\pi, \pi)$ 和 $\varphi \in (-\frac{\pi}{2}, \frac{\pi}{2})$。利用这一关系，可通过雅可比矩阵实现 $dxdy$ 和 $d\theta d\varphi$ 之间的几何变换。具体而言，雅可比行列式 $J(\theta, \varphi)$ 定义如下：

$$J(\theta, \varphi) = \frac{\partial(x, y)}{\partial(\theta, \varphi)} = \begin{vmatrix} \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial \varphi} \\ \frac{\partial y}{\partial \theta} & \frac{\partial y}{\partial \varphi} \end{vmatrix}. \tag{2}$$

球面上微元 $d\theta d\varphi$ 的面积，即 $\delta S(\theta, \varphi)$，可表示为 $\cos(\varphi)|d\theta d\varphi|$；投影平面上微元 $dxdy$ 的面积，即 $\delta P(x, y)$，等于 $|dxdy|$。球面与投影平面之间的拉伸比 ρ 定义如下：

$$\rho(x, y) = \frac{\delta S(\theta, \varphi)}{\delta P(x, y)} = \frac{\cos(\varphi)|d\theta d\varphi|}{|dxdy|} = \frac{\cos(\varphi)}{|J(\theta, \varphi)|}. \tag{3}$$

对于图2(b)所示的等距圆柱投影，(x, y) 与 (θ, φ) 之间的关系如下：

$$\begin{aligned} x &= \theta \\ y &= \varphi. \end{aligned} \tag{4}$$

此时，等距圆柱投影 $J_{ERP}(\theta, \varphi)$ 的雅可比行列式等于 1。因此，等距圆柱投影 ρ_{ERP} 的拉伸比如下简化：

$$\rho_{ERP}(x, y) = \frac{\cos(\varphi)}{|J_{ERP}(\theta, \varphi)|} = \cos(\varphi) = \cos(y), \tag{5}$$

其中 $y \in (-\frac{\pi}{2}, \frac{\pi}{2})$。

该投影会导致 ERP图像出现几何畸变。对于一幅低分辨率图像 $I \in \mathbb{R}^{C \times H \times W}$，其失真图$D \in \mathbb{R}^{1 \times H \times W}$如图 2(c) 所示，定义如下：

$$D = \cos \frac{(h + 0.5 - H/2)\pi}{H}, \tag{6}$$

其中， h表示点位置的当前高度。可以推断，失真随纬度增加而加剧，并且在两个半球呈对称分布，沿纬度方向变化。

3.2. 框架概述

所提出的 MDDN 架构概览如图3所示。我们提出的网络由三个主要部分组成：浅层特征提取、深层特征提取和图像重建。深层特征提取部分由若干多层失真感知可变形块（MDDB）、一个卷积层和一个分解式失真感知可变形卷积（D3C）组成。图3下半部分展示了 MDDB 的结构，该结构包含若干多层失真感知可变形层（MDDL）、一个卷积层和一个 D3C 层。为捕获密集且全面的表征，从而有效建模失真模式，每个 MDDL 都将提取的特征$f_{i,\ k}$（其中i 表示第 i个 MDDB， k表示第 k个 MDDL）输入多层失真感知可变形提取器（MDDE）。MDDE 由两个膨胀率分别为 2 和 3 的分解式膨胀失真感知可变形卷积（D4C），以及一个失真感知可变形交叉注意力（DDCA）机制组成。为确保这三个分支更好地协同工作，我们提出了多级特征融合（MFF）模块，以自适应地融合相应的输出特征。 $L1$ 损失是训练损失，用于计算最终超分辨率结果 I^{SR} 与其对应真实值 I^{GT}之间的逐像素差异。

3.3. 多层失真感知可变形提取器

为密集捕获相似模式并适应失真内容，我们提出了多级失真感知可变形提取器（MDDE），其卷积核具有不同的采样范围。具体而言，每个 MDDE 包含两个具有不同膨胀率（$d = 2$ 和 $d = 3$）的分解式膨胀失真感知可变形卷积（D4C），以及一个失真感知可变形交叉注意力（DDCA）模块。DDCA 模块可视为第 1 级提取器（$d = 1$），而 D4C 则分别充当第 2 级（$d = 2$）和第 3 级（$d = 3$）提取器。借助多级提取器，MDDE 通过具有多层采样范围的不同卷积核，有效适应各种几何失真，从而扩大感受野。这使其能够捕获更多相似且相关的特征，进而改善当前图像块的重建效果。

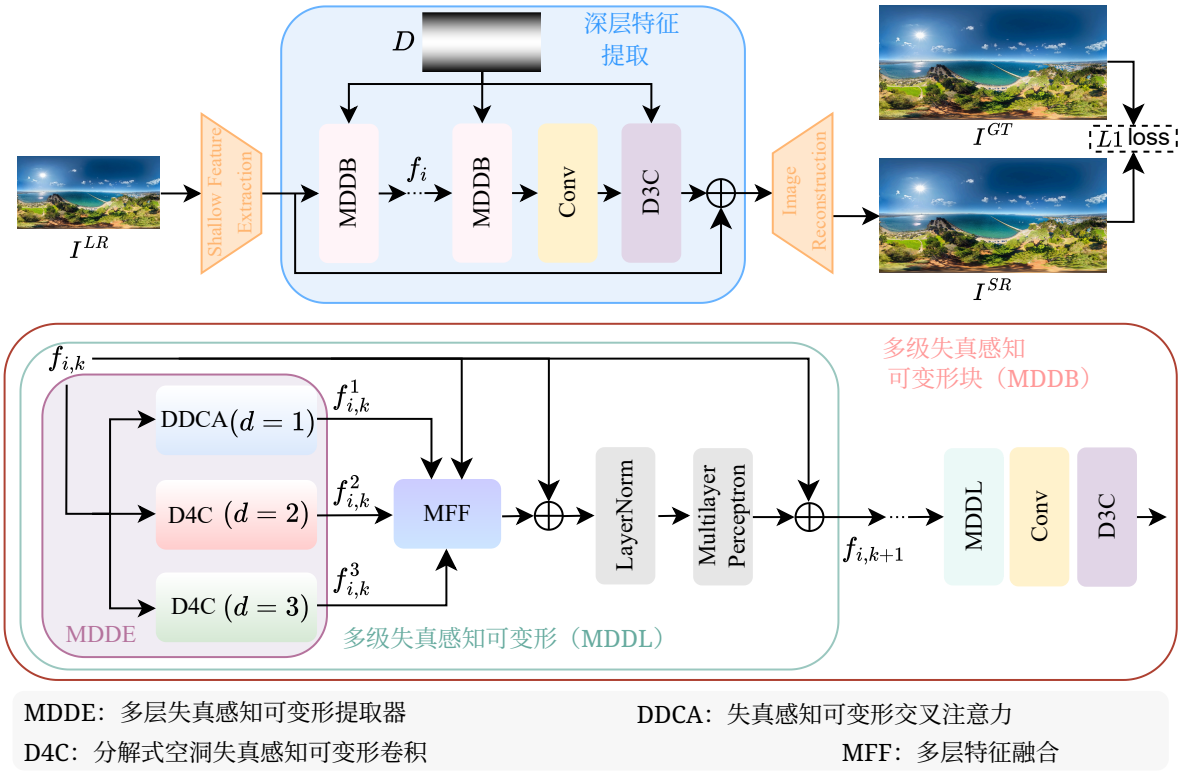


图3： 所提出网络的架构概览。 D 表示失真图。 d 表示膨胀率。

3.3.1. 失真感知可变形交叉注意力

作为第 1 级特征提取器，DDCA 机制凭借其强大的变换能力，在所提出的 MDDE 中发挥着关键作用（参见第4.3.1节）。DDCA 机制及其相关的偏移网络如图4的蓝色框所示。失真图 D 首先输入偏移网络，由其生成相应的偏移量。随后，根据这些偏移量，通过双线性插值对特征 f 进行变形。变形后的特征f' 经过投影形成键特征和值特征，而原始特征 f 则经过投影形成查询特征。该过程可表述如下：

$$\Delta p^1 = \mathcal{N}^1(D), f' = \mathcal{BI}(f, p^1 + \Delta p^1), \tag{7}$$

$$f^1 = \mathcal{CA}(W_q f, W_k f', W_v f'), \tag{8}$$

其中， $\mathcal{N}^1$ 表示用于 DDCA 的偏移网络。继承 OSRT, [51],偏移网络由若干卷积层和激活层组成，其输入通道数和输出通道数标注于各层下方。 $\mathcal{BI}$ 表示双线性插值操作， $\mathcal{CA}$ 表示交叉注意力机制。 p^1 表示规则网格上的位置， Δp^1 表示相应生成的偏移量。变形后的特征 f' 与原始特征 f在交叉注意力机制中进行交互，使 DDCA 能够为输入生成全局表示。

尽管 DDCA 能够有效变换畸变特征，但依赖采样范围仅为单级的采样核会限制其对严重畸变区域中特征的准确建模能力。为解决这一局限，我们引入了 D4C 层，通过使用具有更宽采样范围的扩张卷积核提取局部畸变内容，增强网络的多层特征表示，从而补充 DDCA 提供的全局表示。

3.3.2. 分解式空洞失真感知可变形卷积

在大视场图像中，常常可以观察到宽范围的重复和相似图案，例如地面。由于几何畸变，这些图案在等距圆柱投影（ERP）图像中会进一步扩展。此外，失真程度会随纬度而变化。因此，单层特征提取器捕获 ERP 图像中相似图案的能力有限。为了使特征表示更好地适应 ERP 图像中的几何畸变，我们提出将 DDCA 与具有不同膨胀率的 D4C 层相结合。如图 4所示，红色和绿色方框分别表示具有$d = 2$ 和 $d = 3$的 D4C 层及其对应的偏移网络。通过采用不同的膨胀率，膨胀卷积核会以不同程度扩大其采样范围，如图 4所示，从而为输入形成不同的局部表示，补充 DDCA 提供的全局表示。此外，偏移网络根据几何畸变生成自适应偏移量，以调整卷积核的采样位置，从而适应不断变化的失真。该过程可表示为

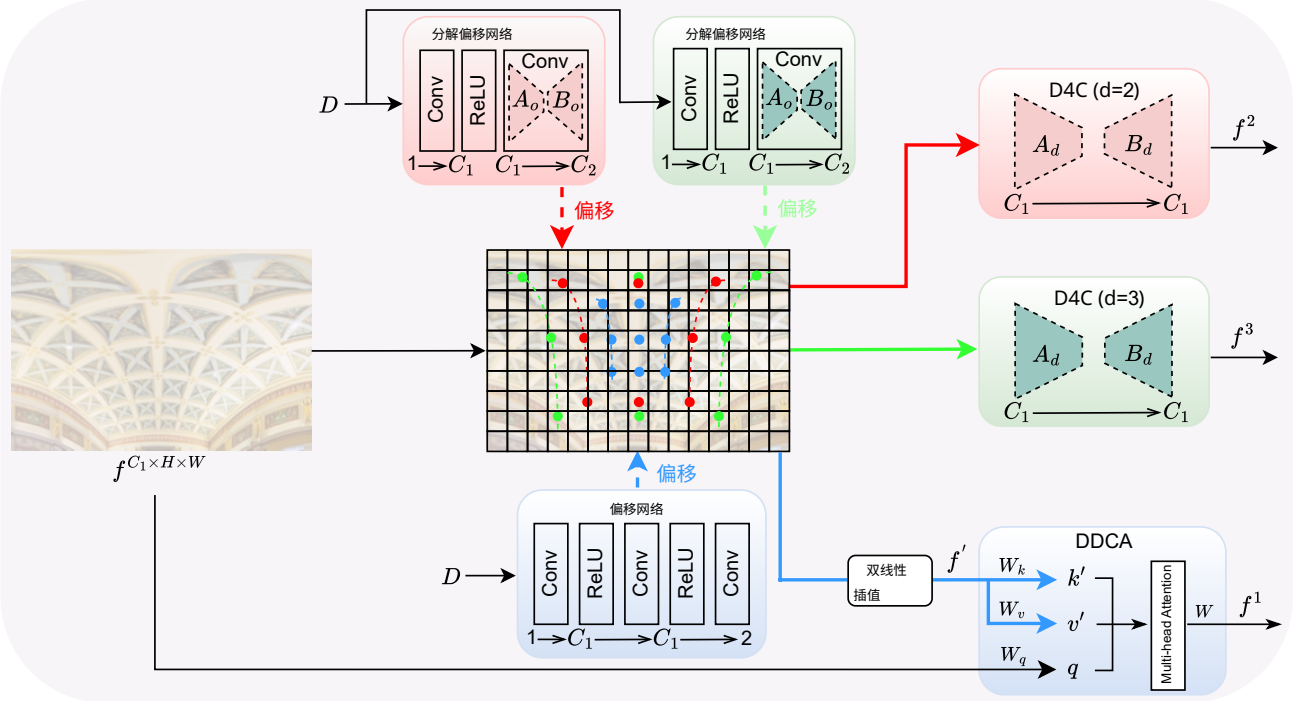


图 4： 多层失真感知可变形提取器（MDDE）的结构。它由失真感知可变形交叉注意力（DDCA）机制和两个分解式空洞失真感知可变形卷积（D4C）层组成，分别具有 $d = 2$ 和 $d = 3$，以及对应的偏移网络。各层的输入通道数和输出通道数标示在层下方。

如下：

$$\Delta p^2 = ONet2(D),\ f^2 = D4C^2(f, \Delta p^2), \tag{9}$$

$$\Delta p^3 = ONet3(D),\ f^3 = D4C^3(f, \Delta p^3), \tag{10}$$

其中，$ONet2$表示具有 $d = 2$的 D4C 层的偏移网络，记为 $D4C^2$，即二级提取器。同理，$ONet3$ 表示具有 $d = 3$的 D4C 层的偏移网络，记为 $D4C^3$，即三级提取器。此外，Δp^2 和 Δp^3 分别表示二级和三级提取器生成的偏移量。

如图4所示，$ONet2$ 和 $ONet3$ 具有相同的结构，由两个卷积层和一个激活层组成。输入通道数和输出通道数标注在相应层的下方。具体而言，失真图 $D \in \mathbb{R}^{1\times H\times W}$ 只有一个通道，因此，第一个卷积层有一个输入通道和C_1 个输出通道。第二个卷积层有 C_1个输入通道和 C_2 个输出通道，其中 $C_1 \gg 1$ 且$C_2 \gg 1$。第二个卷积层的权重矩阵为 $W \in \mathbb{R}^{C_1\times C_2\times N\times N}$，其中 N 表示卷积核大小。考虑到第二个卷积层需要大量训练参数，我们采用低秩分解策略，将其权重分解为两个低秩矩阵，以减少参数数量。分解过程的详细信息将在下一节介绍。

总体而言，在 MDDE 中，DDCA 和 D4C 层利用不同卷积核提取特征，并采用多级采样范围，从而更好地适应随纬度变化的几何失真，并更全面地表征失真的 ERP 图像。

3.3.3. 低秩分解

为减少训练参数的数量，我们不仅在偏移网络中采用低秩分解，也在失真感知可变形卷积中采用低秩分解，包括 D3C 和 D4C。我们主要将一个规模较大且复杂度较高的权重分解为两个低秩矩阵。具体而言，卷积权重 $W \in \mathbb{R}^{C_i\times C_o\times N\times N}$（其中 C_i 和 C_o 分别表示输入通道数和输出通道数，N 表示卷积核大小）首先被重塑为一个二维权重矩阵 $W' \in \mathbb{R}^{C_i\times C_o N^2}$，随后按如下方式分解为两个低秩权重矩阵：

$$W' = AB^T, \tag{11}$$

其中 $A \in \mathbb{R}^{C_i\times r}$ 和 $B \in \mathbb{R}^{C_o N^2\times r}$的秩为 $r \ll$最小值(C_i, C_o)。W' 随后被重塑为其原始形状，并用于对特征进行卷积。在图 4中，A_o 和 B_o表示分解后的偏移网络的低秩权重矩阵，而 A_d 和 B_d 表示D4C的低秩权重矩阵。通过应用低秩分解，可以减少大量训练参数，从而显著节省内存占用和计算成本。

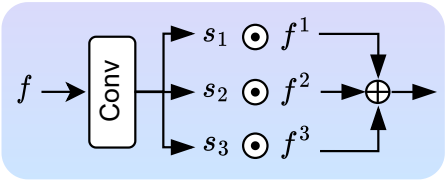


图5：多层特征融合。

3.4. 多层特征融合

在ERP图像中，失真会因区域不同而发生变化。具体而言，较高纬度处的场景会经历更严重的拉伸，因此需要更大的采样范围；而在较低纬度处，失真有所减小，只需适中的采样范围。这种几何畸变使得不同纬度的区域需要采用不同的处理方式。为了有效适应ERP的几何特性，我们在多层级特征融合（MFF）模块中自适应地融合多层级表示。具体而言，我们生成逐像素空间注意力，以调整各特征提取器的贡献。如图 5所示，特征 f 首先经过卷积，为三个特征提取器生成空间注意力。随后，将空间注意力拆分为三组，分别记为 s_1、 s_2和 s_3，并分别与三个特征提取器的输出相乘。对于特征图中的每个像素，不同特征提取器的贡献各不相同。为不同提取器的输出分配不同权重，可以有效适应整幅图像中失真的变化。随后，通过逐像素求和聚合三个加权输出，从而获得融合特征。

4. 实验

4.1. 实验设置

4.1.1. 数据集

在我们的实验中，我们使用 ODI-SR [16] 和 Flickr360[3] 数据集进行训练。ODI-SR 数据集和 Flickr360 数据集分别包含 1,200 张和 3,000 张高分辨率训练图像。测试数据集包括 ODI-SR 测试数据集 [16],、SUN360全景数据集 [47], 和 Flick360-val 数据集[3]。这三个测试数据集分别包含 100 张、100 张和 50 张高分辨率图像。高分辨率图像的分辨率为1024 × 2048 像素。遵循 OSRT [51], ，我们采用鱼眼下采样来获得高分辨率-低分辨率图像对。我们使用 ×4、 ×8和 ×16下采样因子进行实验。

4.1.2. 实现细节

训练过程中，批量大小设置为 4，图像块大小为 256×256 像素。网络训练 500,000 次迭代，初始学习率为 2×10^{-4}，并在第 $250,000^{th}$、$400,000^{th}$、$450,000^{th}$ 和 $475,000^{th}$ 次迭代时减半。采用 $L1$ 损失作为训练损失，用于计算输出超分辨率结果中每个像素与对应真实值之间的平均绝对误差。使用 Adam优化器 [26] ，参数为 $\beta_1 = 0.9$ 和 $\beta_2 = 0.99$。所提出网络的深层特征提取模块由六个 MDDB 组成，每个 MDDB 包含六个 MDDL。隐藏特征维度为 156。实验使用 PyTorch 框架，在 NVIDIA GeForceRTX 4090 GPU 上进行。

4.2. 与最先进方法的比较*4.2.1.* 定量结果

为评估所提方法的性能，我们将其与最先进水平（SOTA）的二维平面图像超分辨率方法和全向图像超分辨率方法进行比较。选取用于比较的二维平面图像超分辨率方法包括 SwinIR[29],HAT [8], 和 RGT [10],, 全向图像超分辨率方法包括 OSRT [51], BPOSR [39], 和 GDGT-OSR [49]。为确保公平比较，所有方法均在相同条件下训练。我们采用轻量版 GDGT-OSR，其参数数量与所提模型相当。表1 展示了三种测试数据集（即 ODI-SR、SUN 360 全景和 Flickr360 验证集）在 ×4 、 ×8 和 ×16 缩放因子下的比较结果。由于无法有效变换几何畸变，在平面图像上表现良好的二维平面图像超分辨率方法（如 HAT [8] 和 RGT [10],）应用于全景图像时性能显著下降，尤其是在缩放因子较大时。相比之下，所提出的 MDDN 优于所有二维平面图像超分辨率方法和全向图像超分辨率模型，展现出处理全景图像的卓越性能。与 GDGT-OSR [49], （一种全向图像超分辨率方法）相比，在 ×4缩放因子下，MDDN 在 Flickr360 验证集上的峰值信噪比和加权球面峰值信噪比分别提升 0.20 分贝和 0.16 分贝。这些结果表明，所提出的 MDDN 能够有效利用多层特征表示，从而在全景图像超分辨率方面取得显著性能。

4.2.2. 定性结果

不同超分辨率方法在×4 缩放因子下的定性评估如图6所示。对应图像块下方标注了峰值信噪比。第一个样本中，高纬度区域的天花板失真严重，重复图案较为稀疏。得益于强大的失真变换能力和较大的采样范围，所提出的 MDDN 在定量指标和视觉效果方面均优于其他对比方法。第二个样本中，高亮图像块中的纹理更为复杂。二维平面图像超分辨率方法和全景图像超分辨率方法都难以恢复精细细节，边缘几乎都被平滑处理（红色箭头所示）；而所提出的方法能够恢复精细边缘，结果更接近真实值。第三个样本由木地板和摄影师的条纹衬衫组成，其中包含许多复杂的纹理，难以恢复，尤其是条纹衬衫。如图6所示，其他最先进水平的超分辨率方法，如 SwinIR [29],HAT [8] 和 RGT [10],

表 1ODI-SR 上不同方法的定量比较 [16], SUN 360 全景 [47] 和 Flickr360 验证集 [3] 数据集。放大倍数为 4、8 和 16。最佳结果以粗体标出。

	ODI-SR [16]											
	×4				×8				×16			
	PSNR	SSIM	加权球面峰值信噪比	加权球面结构相似性	PSNR	SSIM	加权球面峰值信噪比	加权球面结构相似性	PSNR	SSIM	加权球面峰值信噪比	加权球面结构相似性
双三次插值	25.59	0.7118	24.95	0.6923	23.75	0.6442	23.08	0.6123	22.13	0.6116	21.44	0.5732
SwinIR [29]	27.21	0.7683	26.53	0.7542	25.12	0.6873	24.41	0.6616	23.35	0.6395	22.60	0.6052
HAT [8, 7]	27.29	0.7711	26.60	0.7567	25.01	0.6791	24.27	0.6509	23.34	0.6354	22.58	0.5997
RGT [10]	27.25	0.7708	26.55	0.7566	25.14	0.6876	24.42	0.6621	23.32	0.6380	22.57	0.6037
OSRT [51]	27.28	0.7711	26.58	0.7559	25.20	0.6901	24.46	0.6635	23.46	0.6410	22.68	0.6063
BPOSR [39]	26.10	0.7231	25.43	0.7038	24.15	0.6524	23.42	0.6213	22.54	0.6168	21.77	0.5791
GDGT-OSR [49]	27.29	0.7706	26.59	0.7557	25.22	0.6901	24.48	0.6636	23.45	0.6408	22.68	0.6063
MDDN（我们的方法）	**27.39**	**0.7751**	**26.68**	**0.7595**	**25.27**	**0.6926**	**24.52**	**0.6659**	**23.52**	**0.6423**	**22.73**	**0.6077**
	SUN 360 全景 [47]											
	×4				×8				×16			
	PSNR	SSIM	加权球面峰值信噪比	加权球面结构相似性	PSNR	SSIM	加权球面峰值信噪比	加权球面结构相似性	PSNR	SSIM	加权球面峰值信噪比	加权球面结构相似性
双三次插值	25.29	0.6993	24.90	0.7083	23.23	0.6290	22.72	0.6241	21.52	0.5978	20.92	0.5850
SwinIR [29]	27.54	0.7739	27.47	0.7933	25.01	0.6879	24.61	0.6947	22.87	0.6288	22.26	0.6223
HAT [8, 7]	27.67	0.7777	27.60	0.7970	24.80	0.6743	24.34	0.6765	22.80	0.6237	22.21	0.6154
RGT [10]	27.61	0.7770	27.54	0.7965	25.04	0.6886	24.64	0.6955	22.82	0.6276	22.24	0.6210
OSRT [51]	27.65	0.7773	27.55	0.7956	25.16	0.6919	24.70	0.6978	23.04	0.6316	22.40	0.6253
BPOSR [39]	25.99	0.7201	25.71	0.7329	23.67	0.6429	23.15	0.6403	21.94	0.6050	21.30	0.5934
GDGT-OSR [49]	27.68	0.7770	27.58	0.7956	25.19	0.6920	24.74	0.6982	23.04	0.6316	22.41	0.6252
MDDN（我们的方法）	**27.81**	**0.7823**	**27.71**	**0.8002**	**25.27**	**0.6952**	**24.80**	**0.7012**	**23.12**	**0.6336**	**22.48**	**0.6274**
	Flickr360 验证集 [3]											
	×4				×8				×16			
	PSNR	SSIM	加权球面峰值信噪比	加权球面结构相似性	PSNR	SSIM	加权球面峰值信噪比	加权球面结构相似性	PSNR	SSIM	加权球面峰值信噪比	加权球面结构相似性
双三次插值	27.26	0.7579	26.90	0.7496	24.59	0.6697	24.22	0.6497	22.45	0.6280	22.11	0.6022
SwinIR [29]	30.11	0.8300	29.62	0.8235	26.79	0.7265	26.25	0.7088	24.23	0.6618	23.69	0.6370
HAT [8, 7]	30.27	0.8333	29.76	0.8264	26.57	0.7165	26.00	0.6960	24.18	0.6577	23.63	0.6318
RGT [10]	30.18	0.8326	29.67	0.8260	26.79	0.7265	26.25	0.7089	24.17	0.6600	23.66	0.6355
OSRT [51]	30.24	0.8331	29.70	0.8253	26.94	0.7305	26.35	0.7113	24.40	0.6641	23.81	0.6385
BPOSR [39]	28.17	0.7746	27.70	0.7644	25.26	0.6812	24.74	0.6597	23.09	0.6355	22.60	0.6094
GDGT-OSR [49]	30.24	0.8323	29.71	0.8249	26.94	0.7300	26.36	0.7111	24.41	0.6640	23.83	0.6385
MDDN（本文方法）	**30.44**	**0.8380**	**29.87**	**0.8295**	**27.05**	**0.7339**	**26.45**	**0.7144**	**24.48**	**0.6662**	**23.88**	**0.6403**

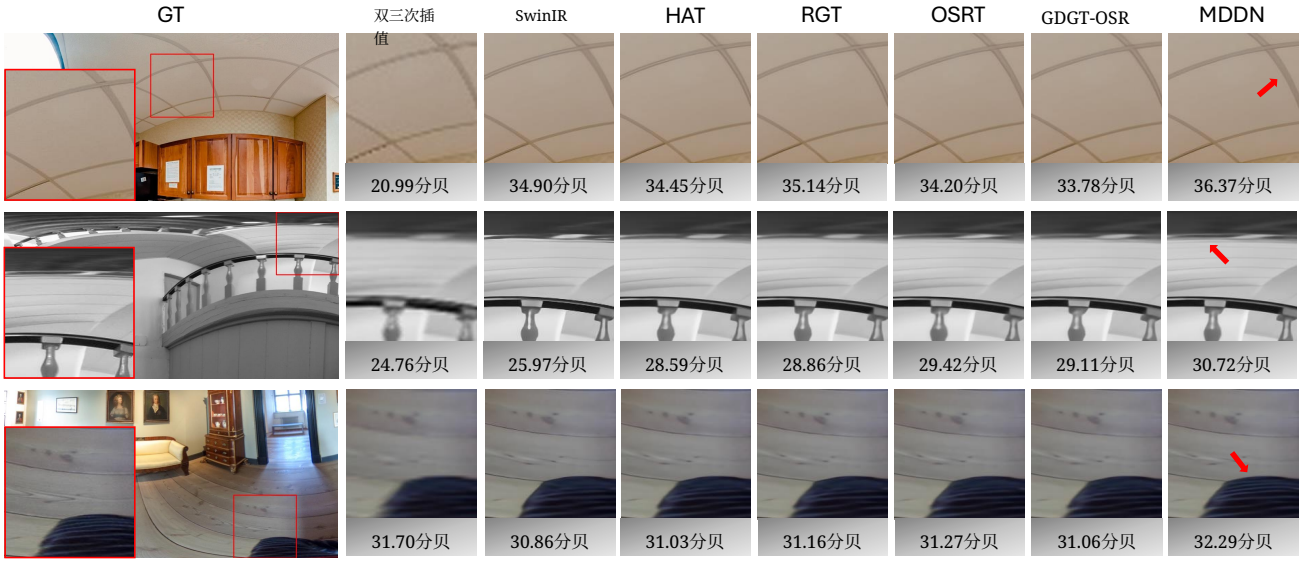


图 6: 不同方法的视觉对比。放大因子为 4。峰值信噪比的计算依据是各方法生成的图像块与真实图像块之间的差异，相应数值显示在对应图像块下方。

会将部分条纹平滑处理。尽管 OSRT [51]和 GDGT-OSR [49] 专为处理全景图像而设计，但由于感受野有限且特征

对全景图像的特征提取能力。得益于全局和局部特征提取的结合，我们提出的方法不仅能够很好地还原地面，还能准确保留条纹中的精细纹理，从而获得

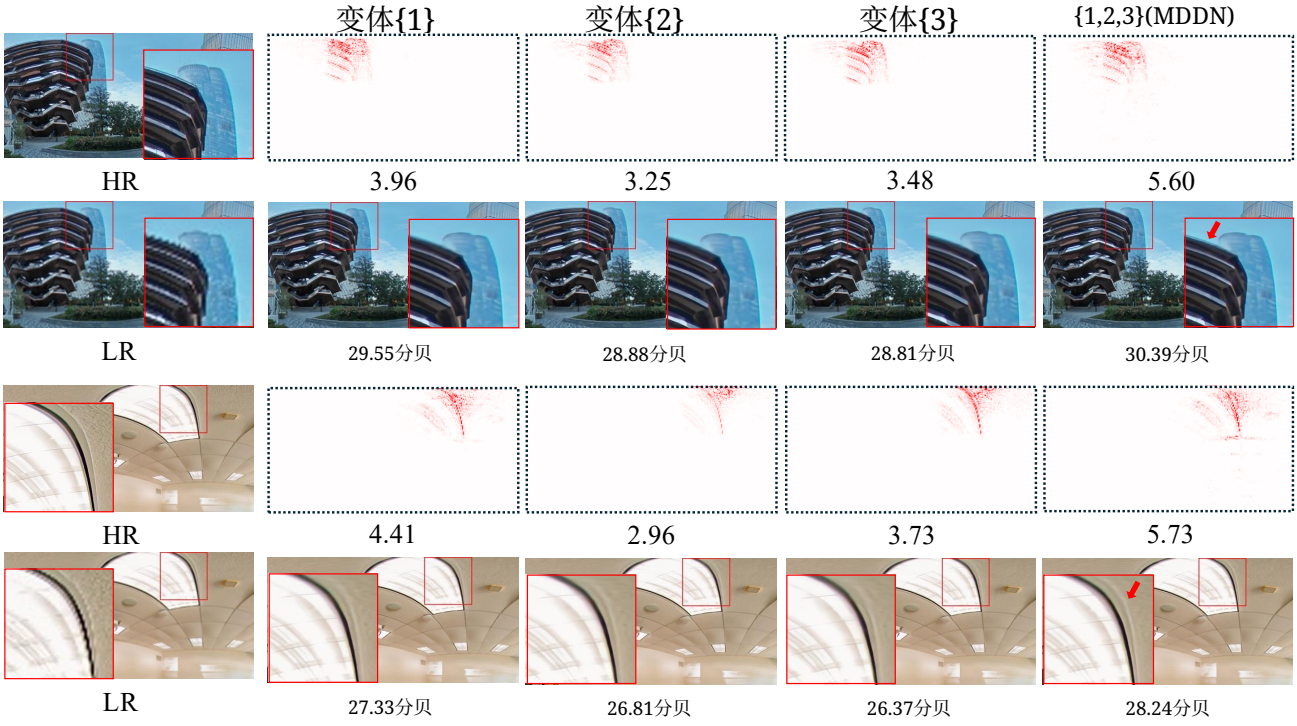


图7： 所提出的 MDDN 与其他单级变体之间的 LAM 和视觉结果对比。每个 LAM 下方显示扩散指数（DI）；DI 值越高，表示参与重建红框标出的图像块的像素范围越广。

表2采用不同特征提取器组合的变体结果。缩放因子为4。在 Flickr360 验证集上评估性能。

变体	PSNR	SSIM	加权球面峰值信噪比	加权球面结构相似性	参数数量（M）
{1}	30.25	0.8337	29.72	0.8259	11,73
{2}	30.10	0.8297	29.57	0.8219	7.82
{3}	30.10	0.8296	29.57	0.8219	7.82
{1, 2}	30.27	0.8339	29.73	0.8259	12.41
{1, 3}	30.27	0.8338	29.74	0.8260	12.41
{2, 3}	30.21	0.8327	29.65	0.8241	8.50
{1, 2, 3}	30.44	0.8380	29.87	0.8295	13.04
{1, 2, 4}	30.44	0.8381	29.87	0.8296	13.04
{1, 3, 4}	30.44	0.8379	29.86	0.8294	13.04
{1, 2, 5}	30.44	0.8380	29.87	0.8295	13.04
{1, 2, 3, 4}	30.45	0.8382	29.88	0.8297	13.67

较高的峰值信噪比。需要注意的是，双三次插值生成的图像块的峰值信噪比高于其他超分辨率方法生成的图像块，MDDN 除外。尽管双三次插值生成的图像块在量化指标上表现优异，但其视觉效果最为平滑，几乎缺失了所有细节。一些过度平滑的结果可能在峰值信噪比和结构相似性指数等像素级指标上取得良好表现，这也是这些指标固有的缺陷。

4.3. 消融实验

4.3.1. 多层提取器的组合

表2 展示了采用不同特征提取器组合的变体之间的定量比较。我们用记号 {1,..., n} 表示采用不同层级特征提取器的变体。例如，变体 {1, 2} 在 MDDE 中使用四级1 ($d = 1$) 和五级2 ($d = 2$)特征提取器，而 {1, 2, 5} 则由四级1 ($d = 1$)、二级($d = 2$)和五级5 ($d = 5$) 特征提取器组成。从表 2可以看出，采用单级特征提取器的变体，即 {1}、{2} 和 {3}，性能明显不如多级变体。其中，变体 {1} 优于 {2} 和 {3}，凸显了一级特征提取器所提供的全局表示至关重要。为了进一步直观展示 MDDN 与单级变体之间的差异，我们采用局部归因图（LAM）这一归因工具，以可视化超分辨率重建中各像素的重要性。图 7 比较了 MDDN 与单级变体的 LAM 和视觉结果。LAM 突出了红框所圈定图像块的重建过程中各特定像素的重要性。每张 LAM 下方显示的扩散指数（DI）量化了所涉及像素的范围，DI 值越高表示像素覆盖范围越广。如图 7所示，与单级变体相比，所提出的 MDDN 的 LAM 显示出更大的采样范围和更多参与重建的像素，这体现在更高的 DI 值上。视觉结果也表明，MDDN 输出中的边缘更加锐利

表3低秩分解的消融研究。

	PSNR	SSIM	加权球面峰值信噪比	加权球面结构相似性	参数数量(M)	乘加运算次数(G)
Less Dim	27.32	0.7724	26.61	0.7569	13.05	45.34
Full Dim	**27.40**	**0.7751**	**26.69**	**0.7597**	30.62	56.50
MDDN	27.39	**0.7751**	26.68	0.7595	13.04	45.41

表4秩 (*r*)的消融研究。缩放因子为4。在 ODI-SR 数据集上进行评估。

r	PSNR	SSIM	加权球面峰值信噪比	加权球面结构相似性	参数数量(M)	乘加运算次数(G)
4	27.32	0.7721	26.61	0.7569	12.47	44.90
8	27.39	0.7751	**26.68**	0.7595	13.04	45.41
12	**27.40**	**0.7752**	**26.68**	**0.7596**	13.62	45.92
16	**27.40**	**0.7752**	**26.68**	**0.7596**	14.19	46.43
20	**27.40**	**0.7752**	**26.68**	**0.7596**	14.64	46.94

表5不同分支输出融合方法的结果。缩放因子为 4。性能在 Flicker360 验证集上进行评估。

	PSNR	SSIM	加权球面峰值信噪比	加权球面结构相似性
加法	30.38	0.8364	29.82	0.8283
MFF	**30.44**	**0.8380**	**29.87**	**0.8295**

清晰。与单级变体的输出相比，这表明，更大的相关像素和自相似像素采样范围有助于提升超分辨率重建效果。MDDN 中的全局和局部特征提取器扩大了模型的采样范围，从而提高重建质量。此外，如表 2中的变体 {1, 2}、{1, 3} 和 {2, 3} 所示，为单级变体增加一个特征提取器能持续提升性能，体现了特征提取器之间的协同作用与互补性。包含全部三种类型提取器（即一级、二级和三级特征提取器）的模型取得了显著的性能提升。这表明全局与局部特征提取器能够有效互补，且各自都发挥着关键作用。我们进一步研究了较高层级特征提取器（如四级和五级）在 {1, 2, 4}、{1, 3, 4} 和 {1, 2, 5} 等组合中的影响。尽管这些变体包含采样范围更大的高层级提取器，其性能仍与 {1, 2, 3} 相当，甚至更差。这表明，过大的采样范围可能会纳入无关像素，容易导致重建质量下降。同样，增加第四个特征提取器（例如 {1, 2, 3, 4}）会增加参数，却未带来明显的性能提升。基于上述分析，我们采用变体 {1, 2, 3} 作为模型中多级特征提取器的默认配置。

4.3.2. 低秩分解的影响

为降低计算成本，我们采用低秩分解，将大型卷积核权重矩阵分解为两个低秩矩阵，从而提高模型效率。为评估其有效性，我们进行了实验，将 MDDN 与未采用低秩分解策略的变体进行比较。第一个变体 Less Dim 不进行低秩分解，而是直接降低偏移网络和 D4C 层中的通道维度，使其参数总量与 MDDN 相当。如表 3 所示，尽管参数数量相近，Less Dim 变体的性能却较差。这表明，与简单的降维相比，低秩分解能够提取更有效的表示，凸显了我们低秩分解策略的优越性。我们还引入了一个全维度变体 Full Dim，该变体保留原始特征维度 C_1，但不采用低秩分解。与 MDDN 相比，该变体的参数数量超过其两倍，计算复杂度也显著更高。然而，其性能仅与 MDDN 相当，并无明显提升。这表明全维度特征图中存在大量冗余。这些实验结果表明，采用低秩分解对于提升模型性能、降低复杂度既有必要，也有益。

4.3.3. 秩的影响

我们探究了 D4C 层和偏移网络中低秩分解所采用的不同秩值的影响。如表 4所示，秩从 4 增至 8 时，性能有所提升。当秩进一步从 8 增至 20 时，模型性能提升变得有限，而参数数量和模型复杂度却显著增加，导致训练内存和计算成本上升。在极端情况下，如表 3中的“Full Dim”所示，完全禁用低秩分解会使模型的参数数量和计算复杂度大幅增加，却未带来显著的性能提升。因此，为平衡模型复杂度和性能，我们在实验中采用 $r = 8$ 作为默认秩值。

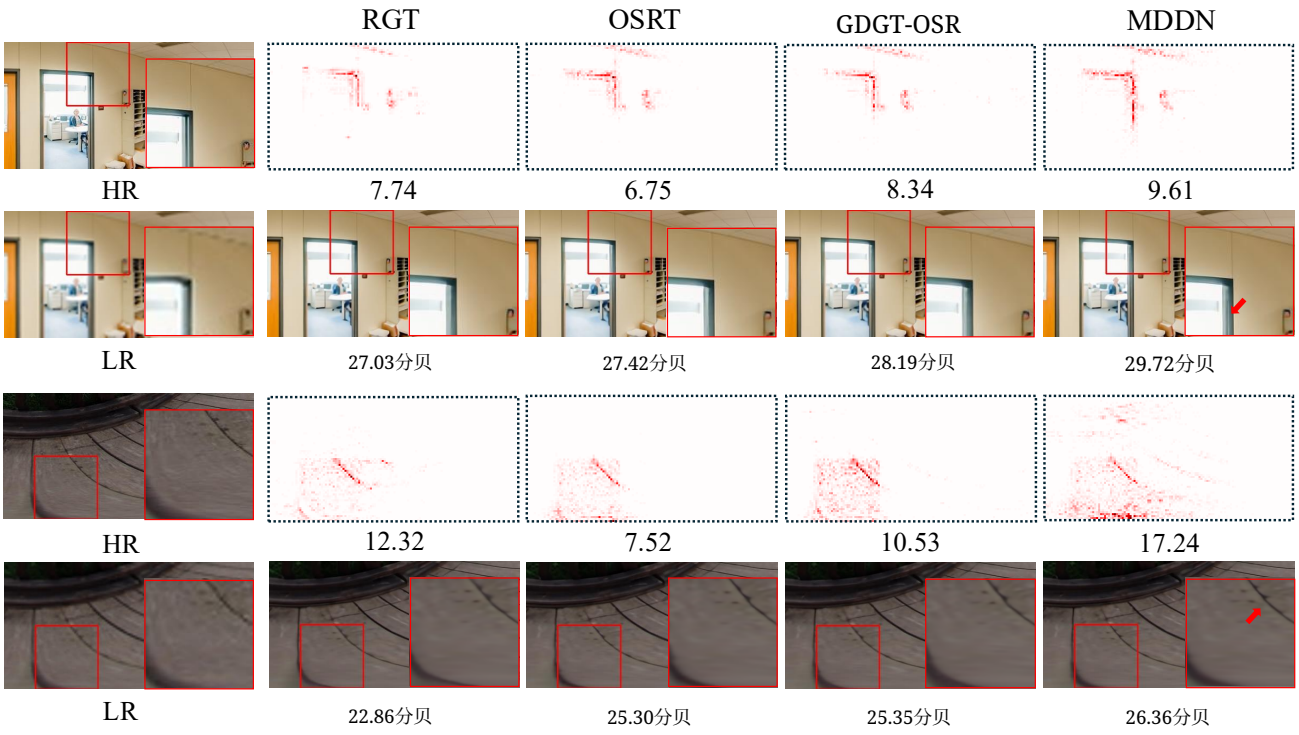


图8： 不同方法的局部归因图（LAM）和视觉结果对比。每个 LAM 下方显示扩散指数（DI），DI 值越高，表示参与重建红框中图像块的像素范围越广。

4.3.4. 多级特征融合的影响

多级提取器输出的三组特征在 MFF模块中进行自适应聚合。我们在表 5中比较了不同的融合方法。“加法”方法表示将提取器输出的三组特征直接逐像素相加进行融合。如表5所示，直接相加融合特征会导致所有评估指标的性能下降。这可能是由 ERP图像的几何属性所致，其畸变程度随纬度而变化。不同区域需要不同的特征提取范围；畸变严重的区域由于场景极度扩张，需要更大的提取范围。采用自适应特征融合策略，能够在不同畸变程度下优化聚合特征表示，使模型从中受益。

4.4. 对涉及区域的探索

为了分析与所提出的方法相比，其他超分辨率方法涉及的像素范围，我们考察了它们的 LAM [21] 和图 8 中的视觉结果。LAM 突显了相应像素在重建图像块时的重要性。我们还在每个 LAM 下方给出了 DI 值，用于量化像素参与的程度。DI 值越高，表示像素覆盖范围越广。如图 8所示，在两个样例中，所提出的方法都实现了最大的贡献区域和最高的 DI 值。这表明

表6不同方法的模型复杂度对比。

模型	参数数量（M）	#乘加运算次数（G）	PSNR	[illegible]
SwinIR [29]	12.05	60.59	26.79	26.25
HAT [8, 7]	20.92	96.05	26.57	26.00
RGT [10]	13.51	62.17	26.79	26.25
OSRT [51]	12.08	54.50	26.94	26.35
BPOSR [39]	2.19	40.33	25.26	24.74
GDGT-OSR [49]	13.45	56.69	26.94	26.36
MDDN	13.19	55.43	27.05	26.45

采样范围更大。得益于多层特征提取器的设计，该方法能够捕获更多自相似纹理，这对于准确重建全景图像超分辨率结果至关重要。例如，在第二个样例中，其他方法（如 RGT [10], OSRT[51],和 GDGT-OSR [49], ）的 LAM 贡献区域几乎都局限在红框内。这一限制降低了定位更多相似区域以改进重建的能力。相比之下，所提出的方法凭借融合全局和局部特征提取的多层特征提取器，将贡献区域扩展到红框之外，从而纳入更多相似区域。通过利用自相似纹理中的更多像素，所提出的 MDDN 带来了视觉效果最出色的超分辨率结果，并取得了最高的峰值信噪比。

4.5. 模型复杂度

表6 展示了不同方法在参数数量和乘加运算次数方面的模型复杂度。

在 Flickr360 验证集上，采用 ×8 缩放因子进行性能评估，乘加运算次数按输入尺寸为 64×64 像素计算。尽管 HAT 等二维平面图像超分辨率方法 [8] 和 RGT [10], 的参数数量和乘加运算次数高于 OSRT 等全向图像超分辨率方法 [51], GDGT-OSR [49], 以及我们提出的 MDDN，但其性能仍不及专为全景图像超分辨率设计的方法。这也说明，处理全景图像需要采用专门设计的全景图像超分辨率方法。GDGT-OSR 轻量版 [49]的参数数量和乘加运算次数均高于 MDDN。然而，MDDN 的模型复杂度更低，却仍优于最先进水平的 GDGT-OSR。这证明了所提出的 MDDN 在模型复杂度方面的有效性和优越性。

4.6. 局限性

我们的方法采用 Transformer架构作为骨干网络，擅长捕捉长程依赖关系并对数据中的复杂模式进行建模。然而，一个显著的局限性在于其缺乏对显式语义信息的利用。Transformer模型主要关注低层和中层特征提取，往往忽略更高层次的语义上下文，而这些信息有助于增强其对输入内容的理解。近期基于扩散的超分辨率模型强调了语义信息对于提升重建性能的重要性。这些模型能够有效整合类别标签和文本描述等语义线索，以引导重建过程。相比之下，我们的方法仅依赖基于图像的特征，这限制了其在语义引导至关重要的场景中的灵活性和表达能力。

5. 结论

本文提出了多层级畸变感知可变形网络（MDDN），一种用于全向图像超分辨率（ODISR）中有效提取多层级表征的新方法。通过将畸变感知可变形注意力机制与扩张畸变感知可变形卷积相结合，MDDN配备了多样化的特征提取器，这些提取器具有多层级采样范围，能够适应失真的ERP图像，并利用互补的全局和局部特征。这些多层特征提取器在多层特征融合模块中进行自适应协作，从而提升重建性能。此外，本文采用低秩分解策略，以降低计算成本并优化效率。在公开数据集上的大量实验表明，MDDN优于二维平面超分辨率方法和现有的全向图像超分辨率方法，在定量结果和定性结果方面均取得了更优表现，从而验证了所提方法的有效性和优越性。

参考文献

[1] Baniya, A.A., Lee, T.K., Eklund, P.W., Aryal, S., 2023. 使用深度学习进行全景视频超分辨率。IEEE Transactions 《多媒体》26, 540–554。[2] Cai, Y., Bian, H., Lin, J., Wang, H., Timofte, R., Zhang, Y., 2023。Retinexformer：基于 Retinex 的单阶段低光照图像增强Transformer，载于：IEEE/CVF国际计算机视觉会议论文集，第12504–12513页。 [3] Cao, M., Mou, C., Yu, F., Wang, X., Zheng, Y., Zhang, J., Dong, C., Li, G., Shan, Y., Timofte, R., 等，2023。Ntire 2023年360度全向图像与视频超分辨率挑战赛：数据集、方法与结果，载于：IEEE/CVF计算机视觉与模式识别会议论文集，第1731–1745页。 [4] Chai, X., Shao, F., Chen, H., Mu, B., Ho, Y.S., 2023年a。通过动态卷积与跨视图Transformer实现立体全景显示系统的超分辨率重建。IEEE仪器与测量汇刊72, 1–12。[5] Chai, X., Shao, F., Jiang, Q., Ying, H., 2023年b。TCCL-Net：用于全向图像超分辨率的Transformer-卷积协同学习网络。Knowledge-Based Systems 274, 110625。 [6] Chen, H., Wang, Y., Guo, T., Xu, C., Deng, Y., Liu, Z., Ma, S., Xu, C., Xu, C., Gao, W., 2021年a。预训练图像处理Transformer，载于：IEEE/CVF计算机视觉与模式识别会议论文集，第12299–12310页。 [7] Chen, X., Wang, X., Zhang, W., Kong, X., Qiao, Y., Zhou, J., Dong, C., 2023年a。HAT：用于图像复原的混合注意力Transformer。arXiv预印本 arXiv:2309.05239。 [8] Chen, X., Wang, X., Zhou, J., Qiao, Y., Dong, C., 2023年b。激活图像超分辨率Transformer中的更多像素，载于：IEEE/CVF计算机视觉与模式识别会议论文集，第22367–22377页。 [9] Chen, Y., Liu, S., Wang, X., 2021年b。利用局部隐式图像函数学习连续图像表示，载于：IEEE/CVF计算机视觉与模式识别会议论文集，第8628–8638页。 [10] Chen, Z., Zhang, Y., Gu, J., Kong, L., Yang, X., 2024 年。用于图像超分辨率的递归泛化Transformer，载于：国际学习表征会议论文集。 [11] Chen, Z., Zhang, Y., Gu, J., Kong, L., Yang, X., Yu, F., 2023年c。用于图像超分辨率的双重聚合Transformer，载于：IEEE/CVF计算机视觉与模式识别会议论文集，第12312–12321页。 [12] Chen, Z., Zhang, Y., Gu, J., Kong, L., Yuan, X., 等，2022年。用于图像复原的交叉聚合Transformer，载于：神经信息处理系统进展，第25478–25490页。 [13] Chi, Y., Gu, J., Zhang, J., Yang, W., Tian, Y., 2024年。Egovsr：迈向高质量第一人称视频超分辨率。IEEE视频技术电路与系统汇刊。 [14] Conde, M.V., Choi, U.J., Burchi, M., Timofte, R., 2022年。Swin2SR：用于压缩图像超分辨率与复原的SwinV2 Transformer，载于：欧洲计算机视觉会议，Springer，第669–687页。 [15] Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., Wei, Y., 2017年。可变形卷积网络，载于：IEEE国际计算机视觉会议论文集，第764–773页。 [16] Deng, X., Wang, H., Xu, M., Guo, Y., Song, Y., Yang, L., 2021年。LAU-Net：用于全向图像超分辨率的纬度自适应上采样网络，载于：IEEE/CVF计算机视觉与模式识别会议论文集，第9189–9198页。 [17] Dong, C., Loy, C.C., He, K., Tang, X., 2015年。使用深度卷积网络进行图像超分辨率。IEEE模式分析与机器智能汇刊38, 295–307。 [18] Dong, R., Yang, C., Lam, K.M., 2024 年。利用掩码图像进行文本引导的视觉提示调优以实现面部表情识别，载于：2024年亚太信号与信息处理协会年度峰会与会议（APSIPA ASC），IEEE，第1–6页。 [19] Fakour-Sevom, V., Guldogan, E., Kämäräinen, J.K., 2018年。使用深度卷积网络进行360度全景超分辨率，载于：国际计算机视觉理论与应用会议（VISAPP），第2页。

[20] Gilman, A., Bailey, D.G., Marsland, S.R., 2008年。用于图像超分辨率的插值模型，载于：第四届IEEE国际电子设计、测试与应用研讨会（delta 2008），IEEE，第55–60页。[21] Gu, J., Dong, C., 2021年。利用局部归因图解释超分辨率网络，载于：IEEE/CVF计算机视觉与模式识别会议论文集，第9199–9208页。 [22] Gu, Y., Chen, S., Sun, X., Ji, J., Zhou, Y., Ji, R., 2025年。通过双向交叉注意力与注意力恢复实现光学遥感图像显著目标检测。《模式识别》164, 111478。 [23] Guo, Y., Wu, X., Shu, X., 2022年。图像超分辨率双参考深度学习的数据采集与准备。IEEE图像处理汇刊31, 4393–4404。 [24] He, C., Fang, C., Zhang, Y., Ye, T., Li, K., Tang, L., Guo, Z., Li, X., Farsiu, S., 2025年。Reti-diff：利用基于 Retinex 的潜在扩散模型进行光照退化图像复原，载于：国际学习表征会议论文集。 [25] Huang, D., Lin, M., Gan, X., Dai, L., Zeng, H., 2025年。具有自适应令牌选择与特征增强的轻量级图像超分辨率网络。Knowledge-Based Systems，115055。[26] Kingma, D.P., Ba, J., 2015年。Adam：一种随机优化方法，载于：国际学习表征会议论文集。 [27] Kitaev, N., Kaiser, Ł., Levskaya, A., 2020年。Reformer：高效的Transformer，载于：国际学习表征会议论文集。[28] Li, R., Sheng, X., Li, W., Zhang, J., 2024 年。OmniSSR：使用 Stable Diffusion 模型实现零样本全向图像超分辨率，载于：欧洲计算机视觉会议，Springer，第198–216页。 [29] Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R., 2021年。SwinIR：使用Swin Transformer进行图像复原，载于：IEEE/CVF国际计算机视觉会议论文集，第1833–1844页。[30] Liu, L., Luo, T., Jiang, G., Chen, Y., Xu, H., Hu, R., He, Z., 2025年a。DiffOSR：用于全向图像超分辨率的纬度感知条件扩散概率模型。Knowledge-Based Systems 315, 113244。[31] Liu, X., Wang, X., Wang, Z., He, X., Ren, C., 2025年b。基于混合概率方案合成退化流程的真实场景盲图像超分辨率。Knowledge-Based Systems，114643。 [32] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021年。Swin Transformer：使用移位窗口的分层视觉 Transformer，载于：IEEE/CVF计算机视觉与模式识别会议论文集，第10012–10022页。 [33] Nishiyama, A., Ikehata, S., Aizawa, K., 2021年。通过失真感知网络和失真透视图像实现360度单幅图像超分辨率，载于：IEEE国际图像处理会议（ICIP），IEEE，第1829–1833页。[34] Qiu, G., 1999年。用于高效无损编码的渐进预测图像金字塔。IEEE图像处理汇刊8, 109–115。 [35] Qiu, G., 2000年。用于改进图像空间放大的分辨率间查找表。《视觉通信与图像表示期刊》11, 360–373。 [36] Tai, Y., Yang, J., Liu, X., 2017年。通过深度递归残差网络实现图像超分辨率，载于：IEEE计算机视觉与模式识别会议论文集，第3147–3155页。 [37] Tian, X., Liao, X., Liu, X., Li, M., Ren, C., 2025年。用于一体化图像复原的退化感知特征扰动，载于：IEEE/CVF计算机视觉与模式识别会议论文集，第28165–28175页。[38] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017年。你所需的一切就是注意力，载于：神经信息处理系统进展。 [39] Wang, J., Cui, Y., Li, Y., Ren, W., Cao, X., 2024年。通过双投影融合实现全向图像超分辨率，载于：AAAI 人工智能会议论文集，第5454–5462页。[40] Wang, X., Chan, K.C., Yu, K., Dong, C., Change Loy, C., 2019年。EDVR：采用增强型可变形卷积网络的视频复原，载于：IEEE/CVF计算机视觉与模式识别会议研讨会论文集，第0–0页。[41] Wang, Z., Cun, X., Bao, J., Zhou, W., Liu, J., Li, H., 2022年。Uformer：用于图像复原的通用U形Transformer，载于：IEEE/CVF计算机视觉与模式识别会议论文集，第17683–17693页。[42] Wei, J., Yang, G., Wei, W., Liu, A., Chen, X., 2025年。基于动态隐式网络的多对比度磁共振成像任意尺度超分辨率。IEEE视频技术电路与系统汇刊。[43] Wu, X., Lopez-Tapia, S., Wang, X., Molina, R., Katsaggelos, A.K., 2024年。带有基于RRED的感知约束的实时轻量级视频超分辨率。IEEE视频技术电路与系统汇刊。[44] Wu, Z., Liu, W., Li, J., Xu, C., Huang, D., 2023年。SFHN：用于图像超分辨率的空域-频域混合网络。IEEE视频技术电路与系统汇刊33，6459 6473–。[45] Xia, B., Zhang, Y., Wang, S., Wang, Y., Wu, X., Tian, Y., Yang, W., Van Gool, L., 2023年。DiffIR：用于图像复原的高效扩散模型，载于：IEEE/CVF国际计算机视觉会议论文集，第13095–13105页。[46] Xia, Z., Pan, X., Song, S., Li, L.E., Huang, G., 2022年。具有可变形注意力的视觉Transformer，载于：IEEE/CVF计算机视觉与模式识别会议论文集，第4794–4803页。[47] Xiao, J., Ehinger, K.A., Oliva, A., Torralba, A., 2012年。利用全景地点表示识别场景视点，载于：IEEE计算机视觉与模式识别会议论文集，IEEE。第2695–2702页。[48] Yang, C., Dong, R., Lam, K.M., 2024年。面向真实世界全向图像超分辨率的高效适配，载于：2024年亚太信号与信息处理协会年度峰会与会议（APSIPA ASC），IEEE。第1–6页。[49] Yang, C., Dong, R., Xiao, J., Zhang, C., La m, K.M., Zhou, F., Qiu, G., 2025年。面向全向图像超分辨率的几何畸变引导Transformer。IEEE视频技术电路与系统汇刊。[50] Yoon, Y., Chung, I., Wang, L., Yoon, K.J., 2022年。SphereSR：通过连续球面图像表示实现任意投影下的360度图像超分辨率，载于：IEEE/CVF计算机视觉与模式识别会议论文集，第5677–5686页。[51] Yu, F., Wang, X., Cao, M., Li, G., Shan, Y., Dong, C., 2023年。OSRT：基于失真感知Transformer的全向图像超分辨率，载于：IEEE/CVF计算机视觉与模式识别会议论文集，第13283–13292页。[52] Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., 2022年。Restormer：用于高分辨率图像复原的高效Transformer，载于：IEEE/CVF计算机视觉与模式识别会议论文集，第5728–5739页。[53] Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y., 2018年。用于图像超分辨率的残差密集网络，载于：IEEE计算机视觉与模式识别会议论文集，第2472–2481页。[54] Zhou, F., Yang, W., Liao, Q., 2012年。基于插值和多曲面拟合的图像超分辨率。IEEE图像处理汇刊 21，3312–3318。