arXiv:2303.10961v1 [eess.IV] 20 Mar 2023

LFACon：将角度注意力引入光场空间中的无参考质量评估

Qiang Qu、Xiaoming Chen、Yuk Ying Chung，*IEEE* 会员，以及 Weidong Cai，*IEEE* 会员

摘要—光场成像能够捕捉光线的强度信息和方向信息。它天然支持六自由度观看体验，并能提升虚拟现实中的用户参与度。与二维图像评估相比，光场图像质量评估（LFIQA）不仅需要考虑空间域中的图像质量，还需要考虑角度域中的质量一致性。然而，目前缺乏能够有效反映光场图像（LFI）角度一致性、进而反映其角度质量的度量指标。此外，现有的光场图像质量评估指标因光场图像数据量过大而面临较高的计算成本。本文提出了“角度注意力”这一新概念，将多头自注意力机制引入光场图像的角度域。该机制能够更好地反映光场图像质量。具体而言，我们提出了三种新型注意力核：角度自注意力、角度网格注意力和角度中心注意力。这些注意力核能够实现角度自注意力、全局或有选择地提取多角度特征，并降低特征提取的计算成本。通过有效整合所提出的注意力核，我们进一步提出了光场注意力卷积神经网络（LFACon），作为一种光场图像质量评估指标。实验结果表明，提出的 LFACon 指标显著优于最先进的光场图像质量评估指标。对于大多数失真类型，LFACon 均以更低的复杂度和更短的计算时间取得最佳性能。

索引词—无参考图像质量评价、体验质量、光场成像、沉浸式媒体、注意力机制、深度学习。

1 引言

光场成像是虚拟现实中实现六自由度沉浸式用户体验的基石，可同时捕获多个子视图（或子孔径图像）[4,14,20]。与传统二维成像不同，光场成像能够记录现实世界中光线的强度和方向信息[11,39]。因此，光场图像（LFI）不仅具有空间域（继承自二维图像），还具有独特的角度域。LFI 中的每个点都可以表示为一个四维函数$L(u,v,x,y)$，其中（u,v）表示角度坐标，（x,y）表示空间坐标。凭借这种独特的成像原理，光场图像能够自然支持六自由度观看体验和用户交互。然而，在压缩和重建等光场图像处理过程中，光场图像容易受到各种失真的影响，从而降低用户感知的光场图像质量。因此，在实际应用中，开发光场图像质量评估（LFIQA）指标至关重要，以便定量评估用户感知的光场图像质量，并监测用户的体验质量。

在文献中，根据是否有参考光场图像可用，光场图像质量评估可分为全参考任务和无参考任务。全参考光场图像质量评估假设在质量评估过程中可以完全访问参考图像，而无参考光场图像质量评估则无需任何参考图像 [29]。因此，无参考光场图像质量评估在实际应用中至关重要，也是本研究的主要关注点。然而，无参考光场图像质量评估面临若干技术挑战。首先，如何利用不同子视图之间的相关性来量化光场图像的角度质量特征颇具挑战。除可通过传统指标评估的空间质量外，光场图像质量评估指标还需反映角度质量，即用户观看光场图像时各子视图之间的质量一致性

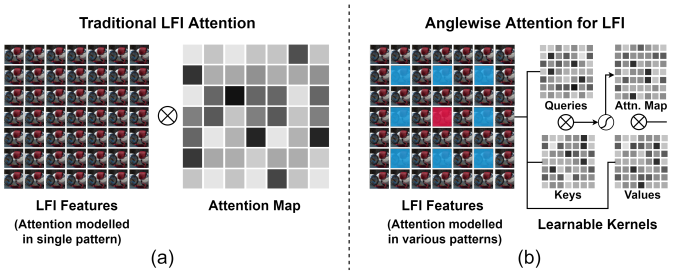

图 1：光场图像的不同注意力模型：(a) 现有的光场图像注意力内核直接将光场图像与学习得到的注意力图相乘。(b) 所提出的内核引入可学习的键、查询和值，以更好地对不同模式下的自注意力进行建模（以不同颜色标示）。

[27]。其次，尽管注意力机制在自然语言处理 [17, 32] 和传统计算机视觉 [6, 15,44], 领域得到了充分研究，但要对光场中独特的“角度注意力”建模仍然很困难。事实上，目前尚无有效的自注意力机制能够从光场图像中提取角度特征，以用于质量评估。若能开发出此类注意力机制，便可选择性地聚焦于关键子视图，从而实现注意力感知的特征提取。第三，由于光场图像的数据量庞大，涉及角度域和空间域的特征提取计算成本高昂。因此，亟需降低这项计算成本，并开发适用于实际应用的轻量级指标。

为应对上述挑战，我们提出“角度注意力”这一新颖概念，以改进无参考光场图像质量评估。具体而言，我们提出三种新的角度注意力核，包括角度自注意力、角度网格注意力和角度中心注意力。角度自注意力用于在角度域中构建自注意力机制，并提取全局多角度特征。另一方面，角度网格注意力和角度中心注意力旨在选择性地聚焦于关键子视图，以较低的计算成本提取关键多角度特征。所提出的角度注意力核在无参考光场图像质量评估中的优势有三点。首先，所实现的自注意力机制能够通过考虑子视图特征所有可能组合之间的相关性，更好地反映角度质量。其次，所提出的角度网格

• *Qiang Qu*、*Yuk Ying Chung* 和 *Weidong Cai* 任职于澳大利亚悉尼大学。电子邮箱：{*vincent.qu* | *vera.chung* |*tom.cai*}*@sydney.edu.au*。

• *Xiaoming Chen* 任职于中国北京工商大学。电子邮箱：*xiaoming.chen@btbu.edu.cn*。

• *Xiaoming Chen* 是通讯作者。

• 项目网址：*https://github.com/VincentQQu/LFACon*

稿件收到日期：*2022*年 *xx xxx*；录用日期：*2023*年 *xxxxx*。出版日期：*2023*年 *xx xxx*；当前版本日期：*2023* 年 *xx xxx*。如需获取本文的重印本，请发送电子邮件至：*reprints@ieee.org*。数字对象标识符：*xx.xxxx/TVCG.2023.xxxxxxx*

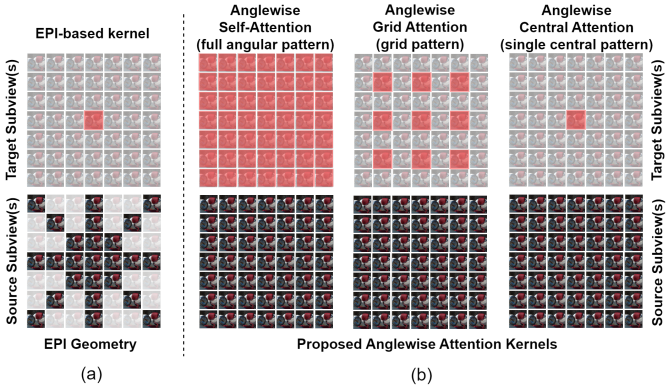

图 2：注意力计算中的角度域覆盖范围：(a) 基于 EPI 的内核仅利用部分源子视图（取决于相对极线几何）来处理目标子视图。(b) 所提出的三种内核利用所有源子视图，以三种模式计算目标子视图的自注意力，并可灵活调整复杂度。

注意力和角度中心注意力能够选择性地聚焦于光场图像特征，并对齐不同子视图间的角度特征，从而改进角度域中的特征提取。第三，三种角度注意力核均可灵活地与其他轻量级前置与后置注意力核结合。因此，这些注意力核能够将自注意力与现有光场图像轻量级内核（例如 Qu 等人提出的光场深度可分离卷积（LF-DSC））的优势结合起来，形成一个广义核。这有望带来更高效、适用于实际应用的无参考光场图像质量评价指标。 [24],

需要阐明所提出的逐角度注意力核与现有光场图像注意力方法之间的区别。如图 1(a) 所示，现有的光场图像注意力机制 [12,19,30] 直接将光场图像与学习得到的注意力图相乘。而我们的注意力核则利用强大的多头自注意力机制，以更好地对跨子视图的多种模式下的自注意力进行建模，如图 1(b) 所示。还值得注意的是，我们的注意力核在注意力计算中涵盖完整的角度域，这与现有的光场图像处理技术不同。例如，极平面成像（EPI）是光场图像处理中应用最广泛的技术之一 [27,38],，但如图 2(a) 所示，它仅根据 EPI 几何利用部分子视图。我们的注意力核如图 2(b) 所示，可利用所有子视图，以灵活的复杂度计算不同模式下的多头自注意力（详情见第 3 节）。因此，所提出的内核有望更好地反映所有子视图之间的角度质量一致性。

基于所提出的核，我们设计并实现了一种学习模型，称为光场注意力卷积神经网络（LFACon），将其作为一种改进的无参考光场图像质量评价指标。为了评估所提出的指标，我们在三个主流公开数据集上进行了大量实验。在实验中，我们首先开展消融研究，以确定适用于所提出的角度注意力核的前后注意力核。随后，我们进行了大量基准测试实验，将所提出的指标与传统图像质量评估指标和最先进的无参考光场图像质量评价指标进行比较。实验结果表明，所提出的指标优于最先进的指标，并且在大多数失真类型上取得了最佳性能。在计算成本方面，所提出的指标还具有相对较短的计算时间和较低的训练成本。

我们的工作贡献总结如下：

- 我们提出了角度注意力这一新颖概念，并为无参考光场图像质量评价设计了三种新型角度注意力核。所提出的新型核构成了一种多头自注意力机制，可实现高效的全局或选择性光场图像特征提取。

- 我们通过消融研究，寻找最优的轻量级前后注意力核，用于所提出的注意力核。确定的轻量级前置与后置注意力核可与任意所提出的注意力核组合成一个广义核，从而降低特征提取的计算成本。

- 我们提出了 LFACon 指标，它将所提出的角度注意力核与确定的前核和后核进行有效组合。大量实验验证了该指标，结果表明，它的性能显著优于最先进的无参考光场图像质量评估指标。

2 相关工作

2.1 光场成像

与捕获光线二维投影的传统摄影不同，光场图像（LFI）呈现来自多个方向的光线。由于数据表示的维度增加，光场摄影能够捕获更全面的视觉数据 [10,18,40]。光场图像的理论基础由全光函数描述，通常表示为 l_λ(x,y,z θ,ϕ,λ,t)，其中 l_λ[W /m2/sr/nm/s] 表示单位时间内的光谱辐亮度，而 (x,y,z) 表示空间域。光场中的每个点都可由一个四维坐标 (u,v,x,y) 表示，其中 (x,y) 表示空间域，(u,v) 表示角度域 [10]。从数字成像的角度来看，光场图像中的每个像素都可通过两个二维坐标定位：一个是用于标识光场图像中子视图的角度坐标，另一个是用于指示该子视图内像素确切位置的空间坐标，这与传统二维图像类似。

2.2 注意力机制

自注意力可用于学习源材料中的对齐关系，并已被证明在自然语言处理领域非常有效，尤其是在 Transformer 模型 [32]问世之后。事实上，近期研究表明，自注意力在计算机视觉任务中也具有优势。Zhang 等人 [44] 首次将多头自注意力引入二维生成对抗网络。视觉 Transformer [6]及其移位窗口变体 Swin Transformer [15], 已被证实在多种计算机视觉应用中的表现优于传统卷积。然而，这些 Transformer 需要更多的可训练参数。由于训练成本高得惊人，直接将 Transformer 用于高维光场数据并不现实。因此，我们尝试以更恰当的方式将自注意力引入光场。尽管已有若干研究将注意力机制应用于光场 [12,19,30],，但这些方法并未利用强大的多头自注意力。图 1 展示了现有光场图像注意力方法与我们提出的角度自注意力方法之间的区别。此外，现有方法将注意力机制与卷积独立使用。相比之下，我们提出的角度注意力能够将注意力机制的优势与可分离卷积的轻量特性结合于一个广义核中（详情见第 3.1 节）。

2.3 图像质量评估

客观图像质量评估方法可根据是否能获取原始参考图像 [9,29], 分为全参考指标和无参考指标。一方面，全参考图像质量评估指标假设在预测质量分数时可以完整获取参考图像。另一方面，无参考指标在不使用原始参考图像的情况下评估图像质量，因此实现起来更具挑战性，但更适用于现实场景。因此，本研究聚焦于无参考指标。一方面，文献中提出了许多用于二维图像的传统指标，包括 ERGAS [33], UQI [35], 结构相似性指数（SSIM） [36], 多尺度SSIM [37], RASE [7],VSI [45],VIF [26], PSNR-B [43], 和 BRISQUE [22]。另一方面，越来越多面向LFI的指标也已被提出

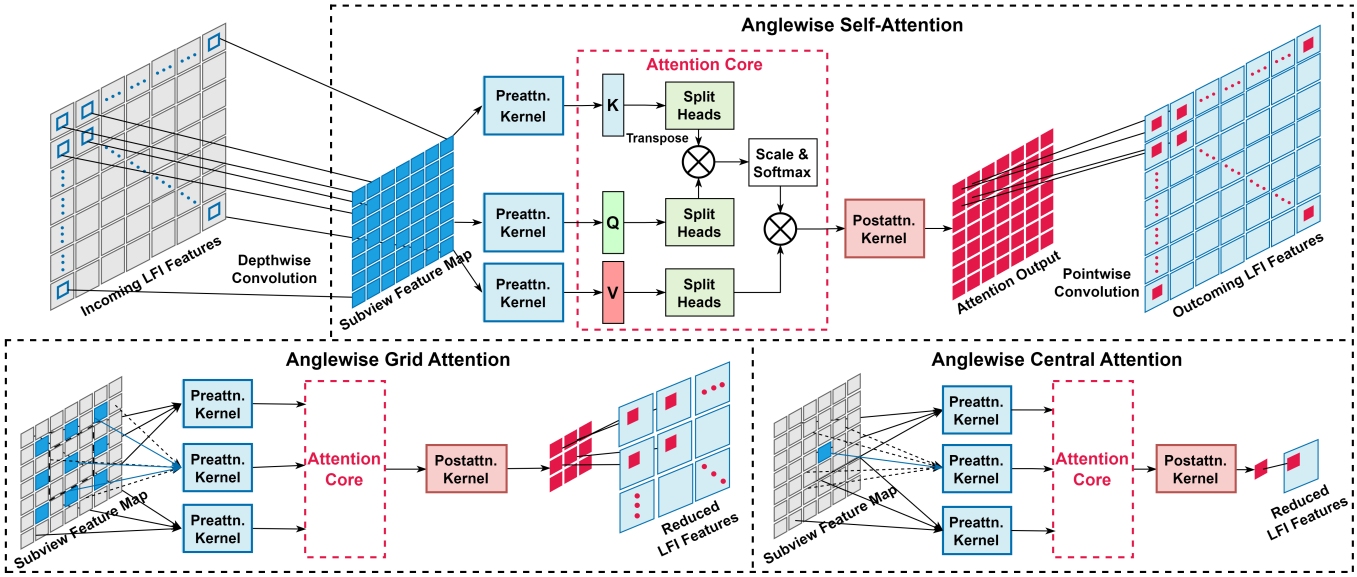

图3：所提出的角度注意力核结构，包括角度自注意力、角度网格注意力和角度中心注意力。所提出的角度注意力核包含强大的多头自注意力核和轻量级卷积操作（位于前后注意力核中），并将它们封装为单个广义核，以实现高效的特征提取。

。Huang等人 [9] 概述了这些指标，包括其分类、评估标准以及一些公开可用的光场图像数据集。近期的无参考光场图像质量评估研究包括 BELIF [27], NR-LFQA [29], Tensor-NLFQ [46], 和 ALAS-DADS [24]。然而，这些指标未能以全面的方式兼顾空间质量和角度一致性，因此无法充分反映光场图像的质量。尽管如此，在 ALAS-DADS 中，Qu 等人 [24] 将深度可分离卷积（DSC）扩展到光场图像的空间域和角度域，并提出了光场版本的 DSC，即 LF-DSC。该方法能够以较低的计算成本提取空间和角度特征。在本研究中，我们认为 LF-DSC 是适用于所提出的角度注意力核中前后注意力核的一个有前景的候选方案（详见并报告于第4节）。

3 方法论

近期研究表明，自注意力有益于二维图像处理 [6, 15]。本研究将自注意力的应用范围扩展到光场图像的角度域，并首次提出角度注意力这一新概念。此外，我们提出了三种新型角度注意力核，包括角度自注意力、角度网格注意力和角度中心注意力。角度自注意力旨在全局提取多角度特征，而角度网格注意力和角度中心注意力则旨在实现选择性特征集中。在这些新型核的基础上，我们提出了一种学习模型 LFACon，作为一种改进的无参考光场图像质量评估指标。

3.1 角度注意力

如第 3.1 节所述，与以往基于注意力的核不同，本研究提出的角度注意力核本质上利用多头自注意力，从光场图像中提取多角度特征。这些核旨在有选择地突出光场图像特征中的关键部分，获取不同角度子视图特征之间的对应关系，并在特征提取过程中降低维度。具体而言，下面将详细介绍提出的三个核。

3.1.1 角度自注意力

图3展示了所提出的角度自注意力模型的设计。假设输入的5D光场图像特征为$LF \in \mathbb{R}^{u\times v\times x\times y\times c}$（两个角度维度 u ×v、两个空间维度 x ×y，以及一个通道维度 c），我们首先通过深度卷积操作从这些特征中生成子视图特征图。随后，将此特征图变换到三个空间 K,Q,V $\in \mathbb{R}^{u\times v\times x\times y\times c}$，即键、查询和值。该变换通过三个独立的可学习前注意力核 $\text{PreAttn}_i\text{(LF)}, i \in \{k,q,v\}$ 完成。第二，需要确定适用于角度自注意力的前后注意力核。根据相关研究 [32, 44], ，直观而言，候选的前后注意力核可以是线性神经投影或 1×1 卷积层。然而，由于近期提出的轻量级 LF-DSC 内核 [24] 显然是用于光场图像的高效低复杂度特征提取器，我们提出将 LF-DSC 作为前后注意力核的另一种有前景的候选方案。为评估这些候选方案，我们进行了消融研究，证实了 LF-DSC 的优越性能（结果见第4节）。第三，在计算注意力值之前，我们将 K,Q,V 的空间通道维度展平（使得 K,Q,V $\in \mathbb{R}^{A\times d}$）。第四，我们将查询、键和值拆分成多个头，并分别应用注意力函数。多个拆分头使角度自注意力能够协同关注来自不同子视图不同表示子空间的感知特征，从而扩展内核容纳更广泛子视图的能力（而非只过度关注相邻子视图）。最后，我们通过对键和查询应用缩放点积，并接着使用后注意力核 PostAttn(X)，计算注意力输出。

假设我们应用具有h 个头的注意力机制，例如$h = 8$。对于每个头 H^j，$j \in [0, h-1]$:

$$H_j = softmax(\frac{QK^T}{\sqrt{d}})V, \tag{1}$$

其中

$$K,Q,V = Z_i = PreAttn_i(LF), \quad i \in \{k,q,v\},$$
$$Z_i \in \mathbb{R}^{A\times d},\ A = u\times v,\ d = \frac{x\times y\times c}{h}. \tag{2}$$

注意力输出的计算方式如下:

$$Output_{Attn} = PostAttn(Concat(H_1, ...H_h)). \tag{3}$$

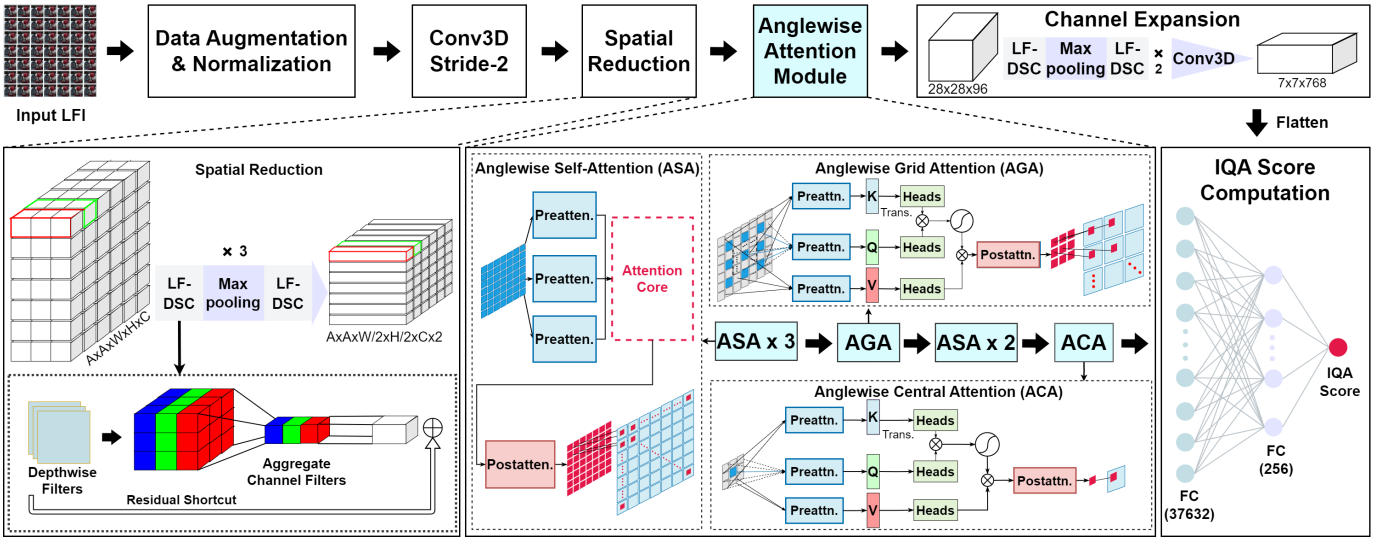

图4：所提出的 LFACon 指标的结构，其中集成了所提出的角度注意力核。

输出$_{Attn}$ 随后经过逐点卷积进行线性投影，从而得到输出的光场图像特征。

3.1.2 角度网格注意力

在光场处理中，选择性地聚焦于关键特征往往至关重要。这是因为在实际应用中，光场处理任务要求输出的维度显著更低，例如，深度估计需要输出二维结果 [2], 1分类需要输出 [16], 维结果，而光场图像质量评估需要输出一维结果 [27]。为此，我们提出角度网格注意力，以从较低维度中提取多角度特征。与将自注意力全局应用于整个角度域（如角度自注意力所做的那样）不同，角度网格注意力仅以网格模式对一组必要的角度特征进行选择性注意。我们设计这一基于网格的注意力模型，而非全局注意力，因为在实际应用中，大多数角度信息都可以从相邻角度子视图特征的视差中获得。如图3所示，我们仅选取网格模式中最重要的角度特征，因为它们能够合理地表征整个角度域。

因此，角度网格注意力的注意力头计算方式不同于角度自注意力中的公式 1 和 2。假设我们采用一个 $b \times b$ 邻域窗口集合 W和步长 s来执行角度网格注意力。对于每个局部窗口 $w \in W$，g 表示中心子视图特征。那么，由 h 个注意力头（即 $m \in [0,h\{v27>$）中的每个局部聚合注意力头 H_m 为：

$$H_m = Concat(H^w) \quad \forall w \in W, \tag{4}$$

$$H^w = softmax(\frac{Q^w K^{wT}}{\sqrt{d}})V^w, \tag{5}$$

其中

$$K^w, V^w = PreAttn_i(w), \quad i \in \{k, v\}, \tag{6}$$

$$Q^w = PreAttn_q(g). \tag{7}$$

角度网格注意力的其余处理过程与角度自注意力几乎完全相同，不同之处在于，角度网格注意力的输出维度显著降低。

3.1.3 角度中心注意力

如图3所示，角度中心注意力的设计继承自角度网格注意力，但进一步集中了特征。我们不再以网格模式从多个子视图中选取角度特征，而是选择中心子视图中的特征，并应用全局自注意力来近似整个角度域。

角度中心注意力的公式与角度网格注意力相同，唯一的区别是，在角度自注意力中，查询位于查询Q 的中心。

3.2 LFACon 指标

基于所提出的新内核，我们提出了一种新的学习模型 LFACon，作为改进的无参考光场图像质量评价指标。图4展示了 LFACon 的结构，具体介绍如下。

表1：多个候选前后注意力核组合的均方根误差 ↓ 对比（最佳结果以粗体标示）。

内核	Win5-LID	SMART	MPI-LFA	参数级别
线性 + 线性	0.4558	0.8853	0.8714	High
Conv3D + Conv3D	0.3972	0.7960	0.8209	Low
Conv3D + LF-DSC	000...333444000333	0.7206	000...777555333999	Low
LF-DSC + LF-DSC	0.3630	000...777000222777	0.7631	中等

首先，我们在训练集内进行归一化。对于光场图像k，

$$\hat{p}_k(D) = \frac{p_k(D) - \mu(D)}{\sigma(D) + 1}, \quad \forall D = (u, v, x, y, c), \tag{8}$$

其中，$(u,\ v)$ 用于标识子视图，而 $(x,\ y,\ c)$ 表示子视图中像素的空间坐标和颜色通道。其次，我们通过执行步长为 2 的三维卷积操作，再结合三维卷积操作与三次“LF-DSC +最大池化 + LF-DSC”操作 [24]，来降低模型输入的空间维度。这些组合操作能够快速降低空间维度，以便进行空间特征提取。组合操作完成后，空间维度从 434×434 降至 28×28。第三，采用经验证的所提注意力核排列来计算角度注意力（结果见第4节）。在此过程中，先通过三个角度自注意力核处理空间降维后的光场图像，以提取角度特征。随后，利用一个角度网格注意力核处理提取的特征，以聚合角度特征，同时将角度维度从 7×7 降至 3×3，然后再经过另外两个角度自注意力核。接着，利用一个角度中心注意力核将角度维度进一步降至 1×1。最后，我们通过通道扩展将通道数从 96 增至 384，以便进行后续的图像质量分数预测，再通过全连接神经网络输出最终的质量分数。值得注意的是，在每个 LF-DSC 步骤结束时，我们选择层归一化而非

表 2：基准图像质量评估指标在 Win5-LID、SMART 和 MPI-LFA 数据集上的定量比较（均方根误差/斯皮尔曼等级相关系数/皮尔逊线性相关系数/离群率）。最佳结果以粗体标出，最后一行表示相较于次优结果的改进。

		Win5-LID				SMART				MPI-LFA			
Type	度量指标	均方根误差 ↓ 斯皮尔曼等级相关系数 ↑ 皮尔逊线性相关系数 ↑			OR ↓	均方根误差 ↓ 斯皮尔曼等级相关系数 ↑ 皮尔逊线性相关系数 ↑			OR ↓	均方根误差 ↓ 斯皮尔曼等级相关系数 ↑ 皮尔逊线性相关系数 ↑			OR ↓
全参考	PSNR	0.7770	0.6579	0.6622	0.0028	1.8047	0.6559	0.5825	0.0049	1.8207	0.4191	0.3945	0.0633
	PSNR-B	0.7883	0.6467	0.6496	0.0028	1.8084	0.6614	0.5802	0.0000	1.8301	0.4012	0.3833	0.0000
	SSIM	0.8202	0.5970	0.6117	0.0170	1.8640	0.5183	0.5433	0.0171	1.9160	0.4025	0.2549	0.1024
	多尺度结构相似性指数	0.8036	0.6350	0.6320	0.0568	1.7955	0.5317	0.5882	0.0318	1.9275	0.4451	0.2316	0.1304
	MSE	0.8644	0.6579	0.5522	0.0483	1.8478	0.6559	0.5544	0.0465	1.9805	0.4191	0.0300	0.1378
	滑动窗口均方根误差	0.8436	0.6128	0.5814	0.0312	1.8522	0.5902	0.5514	0.0196	1.9716	0.3860	0.0996	0.0819
	RASE	0.9272	0.4608	0.4475	0.0057	2.1805	0.1370	0.1884	0.0954	1.9770	0.3415	0.0668	0.1155
	SAM	0.8528	0.5904	0.5687	0.0142	2.0123	0.4422	0.4226	0.0147	1.9599	0.4093	0.1468	0.0875
	SCC	0.8504	0.5532	0.5722	0.0000	2.1430	0.2383	0.2615	0.0098	1.8280	0.3716	0.3858	0.0708
	ERGAS	0.9262	0.4633	0.4495	0.0057	2.0759	0.2218	0.3546	0.0269	1.9768	0.3419	0.0684	0.1117
	UQI	0.9287	0.4865	0.4446	0.0597	2.0871	0.1944	0.3412	0.0465	1.9785	0.3512	0.0544	0.1322
	VIF	0.7970	0.6241	0.6396	0.0000	2.0456	0.4049	0.3888	0.0000	1.7479	0.4156	0.4709	0.0000
无参考	BRISQUE	0.5816	0.8279	0.8242	0.0000	1.2250	0.7680	0.8229	0.0000	1.4433	0.5503	0.6691	0.0000
	NR-LFQA	0.5638	0.9258	0.8746	0.0000	0.8689	0.8884	0.9213	0.0000	1.1302	0.6954	0.8215	0.0000
	Tensor-NLFQ	0.4482	0.9008	0.9075	0.0000	1.1997	0.8383	0.8489	0.0000	1.2829	0.6425	0.7618	0.0000
	ALAS-DADS	0.4371	0.9047	0.9068	0.0000	0.9749	0.8480	0.8984	0.0000	1.2987	0.7000	0.7552	0.0000
	提出的 LFACon	000.....333444000333	000...999444999444	000...999555111888	000...000000000000	000...777222000666	000.....888999000777	000...999555000888	000...000000000000	000...777553333999	000...888222666222	000...999333333555	000...000000000000
	相较于第二名的提升	**+28%**	**+0.024**	**+0.044**	–	**+21%**	**+0.002**	**+0.030**	–	**+50%**	**+0.126**	**+0.112**	–

批归一化，因为层归一化不要求较大的批次大小 [1]。这解决了光场图像归一化中因维度较高而导致的批次较小问题。此外，鉴于其已展现出避免使用预激活函数的能力 [25]，我们选择了一种平滑激活函数 Swish：

$$Swish(x) = \frac{x}{1+e^{-x}}. \tag{9}$$

最后，采用均方误差（MSE）作为训练的损失函数。

4 实验

4.1 实验设置

为验证所提出的注意力内核和 LFACon 指标的有效性，我们在三个公开可用的主流光场图像数据集上进行了全面实验，包括 Win5-LID [28], SMART [23], 和 MPI-LFA [13]。根据 Yeung 等人 [41], 的方法，我们将所有光场图像调整并重塑为相同的大小 $7\times7\times434\times434\times3$，以保留最相关的图像区域。

4.1.1 数据集

Win5-LID 数据集包含 6 个真实场景和 4 个合成场景作为参考光场图像。随后，对这 10 张参考光场图像应用六种失真类型和五个失真级别，生成了 220 张失真光场图像。这六种失真类型分别为高效视频编码、JPEG 2000、线性插值（LN）、最近邻（NN）插值、EPICNN 和 USCD。参与者使用双刺激连续质量测量，在 5 点离散量表上评估这 220 张光场图像的质量。针对每张光场图像，收集了平均意见分数。SMART 数据集基于 16 张原始光场图像构建，另有 256 张光场图像经四种失真类型变换而成，分别为 HEVC 帧内编码、JPEG、JPEG 2000 和 SSDC。主观评价采用 Bradley-Terry 评分系统收集。MPI-LFA 数据集由 14 张无失真光场图像组成。通过应用六种失真生成了 336 张失真光场图像，包括高效视频编码、DQ、OPT、LINEAR、NN 和高斯。为评估光场图像的质量，采用了带有二选一强迫选择的成对比较法。记录刚可察觉差异值作为质量指标。

这三个数据集的真实值（人类视觉感知）值得进一步说明。Win5-LID 数据集的图像质量评分采用平均意见分数，范围为 1 到 5，值越高表示质量越好。相比之下，SMART 数据集采用 Bradley-Terry 评分，通常范围为 -13 到 0，数值越大表示质量越好。MPI-LFA 数据集使用的质量指标是刚可察觉差异分数，通常范围为 -9 到 0，值越大表示质量越好。

4.1.2 数据集增强与评估指标

我们通过将光场图像垂直翻转和/或分别旋转 90、180 和 270 度来增强数据集。因此，数据集规模扩大了八倍：Win5-LID 数据集增至 1760 张光场图像，SMART 数据集增至 2048 张光场图像，MPI-LFA 数据集增至 2688 张光场图像。这种数据集增强方法不仅有助于模型训练，还能让模型更好地学习图像质量模式 [24]。我们使用均方根误差（RMSE） [5], 、Spearman 秩序相关系数（SRCC） [47], 、Pearson 线性相关系数（PLCC） [5], 以及基于 Tukey 围栏的离群值比率（OR） [31] 来评估图像质量评价指标的准确性。例如，当 RMSE 值降低而 SRCC 或 PLCC 值升高时，指标的输出更加准确，或与真实值（人类视觉感知）更加一致。

4.2 实验结果

4.2.1 前注意力与后注意力选择的消融研究

如前所述，我们开展了消融研究，以确定适用于所提出的角度注意力核的前后注意力核。具体而言，我们考察了四种前后注意力核组合：1）两者均采用线性神经投影（Linear +Linear）；2）两者均采用 1×1 3D 卷积层（Conv3D +Conv3D）；3）采用一个 1×1 3D 卷积层和一个 LF-DSC（Conv3D +LF-DSC）；以及 4）两者均采用 LF-DSC（LF-DSC + LF-DSC）。其均方根误差结果见表1。我们发现，“Conv3D +LF-DSC”组合在三个数据集中的两个上取得了最佳结果，且参数数量较少。因此，我们将该组合作为角度注意力核的默认前后注意力核，并将其集成到 LFACon 指标中。

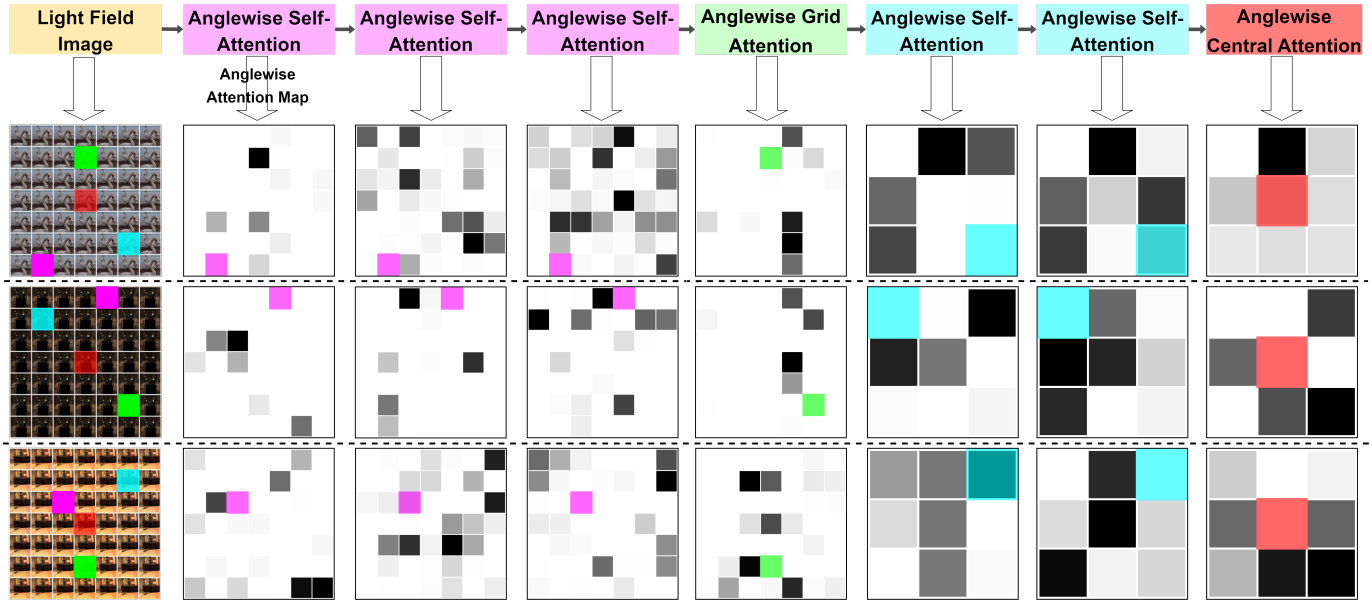

图 5：所提出的角度注意力可视化：三行分别展示了 LFACon 中所集成注意力核的工作流程和注意力图，涵盖三个示例光场图像。每行中，第一列显示示例光场图像，并以红色、粉色、蓝色和绿色标出四个具有代表性的查询位置。其余七列展示这些查询位置对应的角度注意力图（在不同阶段计算得出）。彩色位置对应查询位置。其他位置以不同灰度级标记，以可视化不同的注意力权重。

4.2.2 图像质量评估性能

为验证所提出的 LFACon 指标的有效性，我们将其与 12 种全参考图像质量评估指标和 4 种无参考图像质量评估指标进行比较，其中包括 3 种先进的无参考光场图像质量评估指标，即 NR-LFQA [29], Tensor-NLFQ [46], 和 ALAS-DADS [24]。表2展示了所测试指标在 Win5-LID、MPI-LFA 和 SMART 数据集上的性能。所有受测的无参考光场图像质量评估指标都是监督学习模型，其训练集与测试集的比例与 LFACon 所用比例相同。从表2可以看出，LFACon 在所有基准指标中表现最佳。在 Win5-LID 上，与次优模型相比，LFACon 在均方根误差、斯皮尔曼等级相关系数和皮尔逊线性相关系数方面分别提升了 28%、0.024 和 0.044。在 SMART 数据集上，LFACon 的均方根误差较次优结果改善了 21%。在 MPI-LFA 数据集上，LFACon 的提升更为显著，其均方根误差、斯皮尔曼等级相关系数和皮尔逊线性相关系数分别提升了 50%、0.126 和 0.112。总体而言，LFACon 显著优于所有先进的无参考光场图像质量评估模型，包括近期提出的 ALAS-DADS 模型。值得注意的是，与 LFACon 类似，ALAS-DADS 也采用了 LF-DSC。LFACon 的性能优于 ALAS-DADS，这清楚地表明，融入我们提出的注意力核能够显著提升光场图像质量评估性能。

图 5 展示了所提出的角度注意力核在 LFACon 中针对三个示例光场图像计算出的工作流程和注意力图。直接观察注意力图，很难发现注意力权重中存在任何明显模式，这表明注意力建模具有挑战性。这验证了采用最先进的注意力机制进行注意力建模的必要性。实验结果证实，所提出的角度注意力机制在对此类注意力进行建模时，具备充分的特征提取、特征集中和学习能力。

图 6 展示了一些主观结果，以定性说明所提出的 LFACon 指标的有效性。我们展示了来自三个数据集的三个代表性光场图像。每幅光场图像都经过特定类型和程度的失真处理，并在一行中以其四个子视图呈现。同时，我们给出了 LFACon 对每幅光场图像预测的质量分数，并与最先进的光场图像质量评估指标进行比较。这些结果表明，LFACon 预测的分数更接近真实值。例如，名为“希腊语”的光场图像样本因 JPEG2000失真而出现明显伪影。LFACon 预测其分数为 1.8509，非常接近真实值 1.8696。然而，包括 NR-LFQA、Tensor-NLFQ 和 ALAS-DADS 在内的其他度量指标分别将其质量高估为 2.3758、2.0048 和 2.3429。

4.2.3 对各种失真类型的有效性 es

表 3 展示了所评估的图像质量评价指标在各种失真类型上的性能。从表中可以看出，LFACon 在大多数失真类型上的准确性排名第一。在 10 种失真类型中，它有 8 种的均方根误差最低，并且在 7 种失真类型上的斯皮尔曼等级相关系数最高。还可以注意到，对于 EPICNN 这一失真类型，NR-LFQA 的性能更优。我们认为，这是因为 NR-LFQA 指标采用了基于EPI的处理，而这可能使其在同样使用 EPI 的 EPICNN 上表现占优。此外，LFACon 在 MPI-LFA 数据集上展现出显著优势。例如，对于高效视频编码、OPT 和高斯这几种失真类型，LFACon 的均方根误差分别比排名第二的结果低 87%、53% 和 40%。为了更直观地展示结果，图 7 对比了 LFACon 与最先进的光场图像质量评估指标在不同失真类型上的均方根误差/斯皮尔曼等级相关系数表现。结果清楚地表明，对于大多数失真类型，LFACon 的性能最佳，均方根误差最小且斯皮尔曼等级相关系数最高。

图8展示了在 MPI-LFA、Win5-LID 和 SMART 数据集上，针对每种失真类型，采用所测试的无参考光场图像质量评估指标得到的预测图像质量分数与真实质量分数的散点图。每个子图中的红线表示斜率为 1 的假设理想预测线，即预测质量分数等于真实质量分数的位置。以 MPI-LFA 数据集上的 NR-LFQA、Tensor-NLFQ 和 ALAS-DADS 子图为例，可以发现这些指标会产生有偏的预测，即当质量分数低于 -2 时（其中 0 表示质量最佳）会低估质量分数，而在其他情况下则会高估。相比之下，LFACon 成功减少了

表 3：不同失真类型下具有竞争力的图像质量评价指标的 RMSE/SRCC 对比。最佳结果以粗体标出。LFACon 的排名显示在倒数第二列。最后一列列出了 LFACon 相较于第二佳指标的改进（若 LFACon 排名第一），或与最佳指标的差距（若 LFACon 并非最佳）。

失真	评估	PSNR	SSIM	多尺度结构相似性指数	BRISQUE	NR-LFQA	Tensor-NLFQ	ALAS-DADS	**LFACon**	**排名**	**提升**
DQ	均方根误差 ↓	1.7402	1.8509	1.8526	1.6057	1.0085	1.2906	1.4027	000...777777333999	1	+30%
	斯皮尔曼等级相关系数 ↑	0.2474	0.2378	0.2302	0.4694	0.6002	0.5099	0.5109	000..... 888555111888	1	+0.25
高斯	均方根误差 ↓	1.5908	1.6245	1.6305	1.4613	0.8616	1.0431	0.9532	000...666111111333	1	+40%
	斯皮尔曼等级相关系数 ↑	0.3462	0.3195	0.3276	0.3200	0.5458	000..... 666999999444	0.5024	0.6359	2	−0.06
HEVC	均方根误差 ↓	2.4933	2.4766	2.5510	2.0591	1.4641	1.3750	1.1699	000..... 666222555000	1	+87%
	斯皮尔曼等级相关系数 ↑	0.6381	0.7578	0.7843	0.2767	0.5031	0.5997	0.7276	000..... 888333777000	1	+0.05
LINEAR	均方根误差 ↓	2.1339	2.3727	2.3605	1.3453	1.2970	1.1480	1.3792	000...888333555888	1	+37%
	斯皮尔曼等级相关系数 ↑	0.4856	0.4586	0.5014	0.4329	0.7518	0.8460	0.8615	000...888777999333	1	+0.02
NN	均方根误差 ↓	1.4988	1.5541	1.5674	0.8855	0.7986	1.3433	1.3310	000...666888888555	1	+15%
	斯皮尔曼等级相关系数 ↑	0.5683	0.5466	0.5809	0.7243	0.7898	0.6039	0.7291	000...888551111222	1	+0.06
OPT	均方根误差 ↓	1.6742	1.7398	1.7486	1.5653	1.2926	1.3476	1.2747	000...888222888111	1	+53%
	斯皮尔曼等级相关系数 ↑	0.2472	0.1980	0.2442	0.4168	0.5407	0.4829	0.6722	000...666999222222	1	+0.02
EPICNN	均方根误差 ↓	1.3001	1.1649	1.0664	1.0193	000...111555666555	0.2934	0.4448	0.2325	2	−32%
	斯皮尔曼等级相关系数 ↑	0.7683	0.7450	0.8470	0.4909	000...999333999777	0.8624	0.6996	0.9258	2	−0.01
JPEG2000	均方根误差 ↓	0.7857	0.8714	0.8491	0.5773	0.3059	000...222999777666	0.4450	0.3781	3	−21%
	斯皮尔曼等级相关系数 ↑	0.7116	0.6018	0.6947	0.8344	000...999666111888	0.9501	0.9231	0.9455	3	−0.02
USCD	均方根误差 ↓	0.8289	1.0166	0.9608	0.7798	0.9931	1.1885	0.4553	000..... 333888333999	1	+18%
	斯皮尔曼等级相关系数 ↑	0.6702	0.5793	0.5560	0.8667	0.8126	0.3426	0.9066	000...999222444333	1	+0.02
SSDC	均方根误差 ↓	1.7447	1.8537	1.7757	1.2179	1.1728	1.8321	1.0344	000..... 777111111666	1	+45%
	斯皮尔曼等级相关系数 ↑	0.6333	0.3896	0.4733	0.7136	0.7825	0.7066	0.7311	000...888777111555	1	+0.09

所有失真类型中此类有偏预测模式。例如，图8显示，在这三个数据集上，与其他指标相比，LFACon 的预测质量分数明显更加集中，也更接近假设的理想预测线。这表明，对于所有失真类型，LFACon 不仅能以更小的偏差进行预测，还能降低方差，并显著减少离群值。

表4：LFACon 对不同失真级别的预测结果的斯皮尔曼等级相关系数 ↑ 比较。

失真	跳过 1	跳过 4	跳过 7	跳过 10	跳过 17	跳过 24
DQ	0.7015	000...888333333444	0.5865	0.5969	0.7874	0.4389
高斯	0.1271	0.6983	000...777555999000	0.2206	0.3591	−0.8660
HEVC	0.5988	0.0684	0.3015	0.6357	0.2052	000..... 777888000777
LINEAR	0.3409	0.6792	0.2838	0.3503	0.6303	000...777555444444
NN	0.4943	0.7734	000...888444444888	0.4869	0.1017	0.6955
OPT	−0.0238	0.5628	0.3685	0.4771	0.0881	000...888666333444

4.2.4 不同失真级别下的变化

我们还希望评估 LFACon 在不同失真级别下的性能，以进一步了解其特性。为使相邻失真级别之间的变化幅度较小且具有可比性，MPI-LFA 数据集的作者先针对 10 个失真级别开展了简短的试验研究，随后手动选定最终级别 [13]。各类失真的最终级别选择结果按从中等失真到严重失真的顺序依次为跳过 1、跳过 4、跳过 7、跳过 10、跳过 17 和跳过 24。图 9 展示了不同失真级别（如低、中等和高失真级别）如何影响同一场景中不同视点的子视图。

表4展示了 LFACon 在不同失真级别下的 SRCC 结果（而非质量分数），其中粗体值表示每行（即每种失真类型）中的最高 SRCC 值。SRCC 值越高，表示预测的质量分数越接近真实质量分数。显然，总体而言，LFACon 在 DQ、高斯和神经网络的低失真级别下表现更好，但在高效视频编码、LINEAR 和 OPT 的高失真级别（即跳过 24）下预测更准确。即使在最严重的失真级别下，LFACon 也能够做出相当准确的预测。

表5：无参考光场图像质量评估指标的计算时间比较。

	Win5-LID		SMART		MPI-LFA	
度量指标	总计（秒）	每项（秒）	总计（秒）	每项（秒）	总计（秒）	每项（秒）
NR-LFQA	3387	9.622	4274	10.450	4797	8.933
Tensor-NLFQ	3085	8.764	3441	8.413	4298	8.004
ALAS-DADS	313	0.889	369	0.902	375	0.885
LFACon	**218**	000...666111999	**247**	000...666000444	**326**	000...888000777
改进	**+43.6%**		**+49.4%**		**+15.0%**	

4.2.5 计算效率评估

表5展示了所测试的无参考光场图像质量评估指标在各数据集上输出图像质量分数所需的计算时间。我们发现，LFACon 预测一幅光场图像的质量分数仅需约 0.6 秒。平均而言，这比最先进的 ALAS-DADS 指标快约 36%，也远快于 NR-LFQA 和 Tensor-NFLQ；后两者预测质量分数均需 8 秒以上。

4.2.6 讨论

我们的实验结果表明，提出的 LFACon 指标优于传统图像质量评估指标和最先进的无参考光场图像质量评价指标。我们观察到，对于大多数失真类型，LFACon 都能达到最高准确率。我们认为，LFACon 的卓越性能很大程度上得益于采用了提出的角度注意力核。如第 3.1 节所述，这些内核能够从光场中并行提取多角度特征，捕获反映不同失真类型下光场图像角度质量的关键信息。此外，由于提出的注意力核能够在特征提取过程中选择性地聚焦于关键子视图，提取的特征也进一步集中于重要的质量线索。此外，正如文中所讨论的，

希腊语（JPEG2000失真）来自 Win5-LID



Subview (1, 1) Subview (1, 5) Subview (5, 1) Subview (5, 5)

真实质量分数：1.8696（位于 [1, 5]）

指标	NR-LFQA	Tensor-NLFQ	ALAS-DADS	**LFACon**
质量分数	2.3758	2.0048	2.3429	111...888555000999

巴塞罗那之夜（NN失真）来自 MPI-LFA



Subview (1, 1) Subview (1, 5) Subview (5, 1) Subview (5, 5)

真实质量分数：-0.2349（位于 [-9, 0]）

指标	NR-LFQA	Tensor-NLFQ	ALAS-DADS	**LFACon**
质量分数	-0.5605	-1.3282	-0.6537	**−0.2040**

MSU0481（SSDC失真）来自 SMART



Subview (1, 1) Subview (1, 5) Subview (5, 1) Subview (5, 5)

真实质量分数：-2.4350（在 [-13, 0] 中）

指标	NR-LFQA	Tensor-NLFQ	ALAS-DADS	**LFACon**
质量分数	-4.0230	-3.6972	-2.9612	**−2.4344**

图 6：从三个数据集中各选取一幅光场图像，并展示其在角位置 (1, 1)、(1, 5)、(5, 1) 和 (5, 5) 处的子视图。LFACon 预测的质量分数（粗体）与真实值（下划线）以及最先进的光场图像质量评估指标给出的质量分数进行了对比。**预测分数越接近真实值，指标就越准确**。

第 3.1 节中，所提出的注意力核可以轻松集成到现有的轻量级前注意力核和后注意力核中。我们使用集成后的内核对 LFACon 进行实验，结果表明其速度显著快于基准指标。

我们认为，我们的研究不仅推动了无参考光场图像质量评估的发展，也为更广泛的光场图像研究奠定了理论基础。本研究提出的新概念和框架，例如角度注意力核及相应的 LFACon学习模型，可自然地扩展到其他光场图像研究领域，如超分辨率 [3, 42] 和深度估计 [8, 34]。例如，在光场图像超分辨率任务中，所提出的注意力核可改造为轻量级生成式光场图像特征提取器，并利用计算得到的角度注意力帮助提升光场图像的分辨率。此外，光场图像的深度估计需要处理多个子视图，与光场图像质量评估类似。在这种情况下，可将 LFACon 用作通用学习框架，重新调整其中整合的注意力核，以适应深度估计任务。最后，将所提出的角度注意力应用于其他新兴的三维场景表示也切实可行。例如，为了支持自由视点，神经辐射场 [21] 能够合成复杂三维

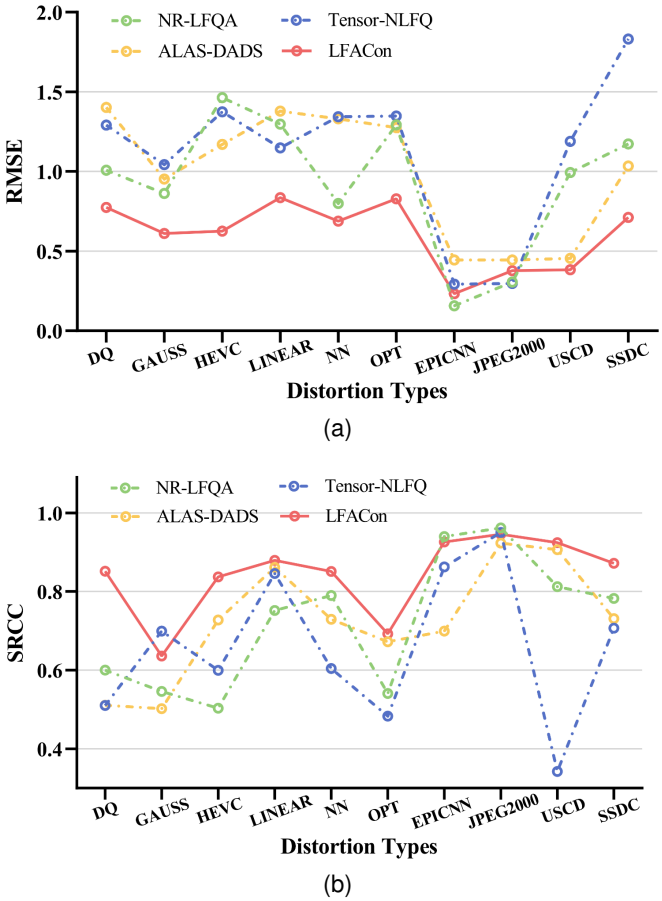

图 7：不同失真类型下最先进的光场图像质量评估指标的可视化比较：(a) 均方根误差 ↓ 比较。(b) 斯皮尔曼等级相关系数 ↑ 比较。

场景的新视图，只需使用稀疏输入视图集。所提出的角度注意力核可针对神经辐射场进行定制，例如用于建模输入视图的注意力，或反映合成新视图的多视图一致性。

5 结论

本文提出了角度注意力这一新颖概念，并设计了三种具体的角度注意力核。验证结果表明，所提出的注意力核能够通过对光场图像中的自注意力进行建模，有效提取质量特征。基于这些新颖的注意力核，我们提出了用于无参考光场图像质量评价的 LFACon 指标。我们进行了大量实验，结果表明，LFACon 优于最先进的度量指标，在大多数失真类型上均取得最佳性能，同时计算时间更短。

在未来的工作中，我们将探索所提出的注意力核在其他光场处理任务中的广泛应用，例如超分辨率和深度估计。我们还可能尝试将所提出的角度注意力机制应用于其他三维场景表示，例如神经辐射场。

致谢

本研究部分受北京自然科学基金（项目编号：4222003）和国家自然科学基金委员会（项目编号：62177001）资助。

参考文献

[1] J. L. Ba、J. R. Kiros 和 G. E. Hinton。层归一化。*arXiv* 预印本 *arXiv:1607.06450*，2016年。

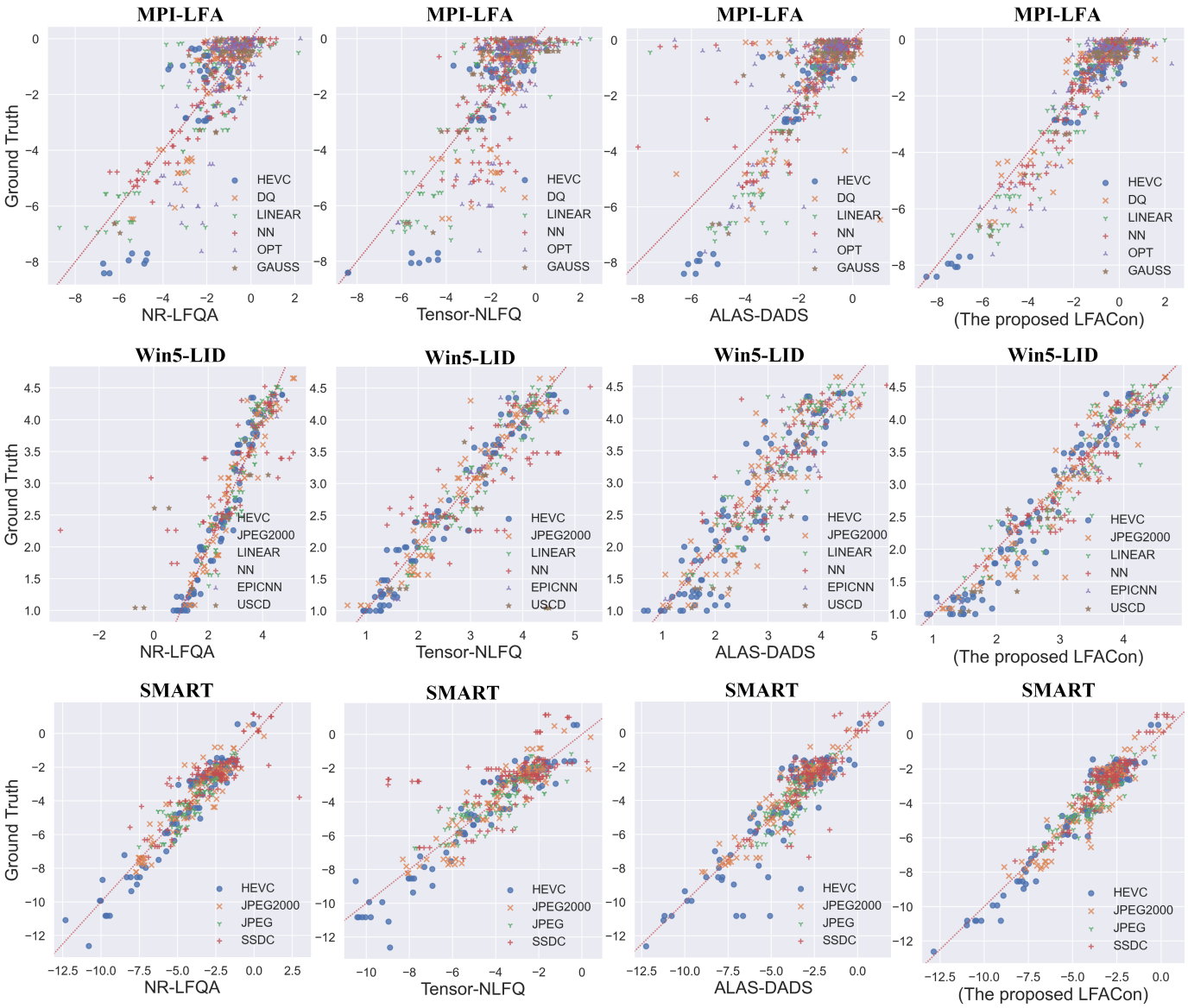

图8：在 MPI-LFA、Win5-LID 和 SMART 数据集上，使用受测无参考光场图像质量评估指标得到的真实图像质量分数与预测质量分数的散点图。不同失真类型以不同符号和颜色标示。每个子图中的红线表示假设的理想预测线。**点越接近红线，性能越好**。

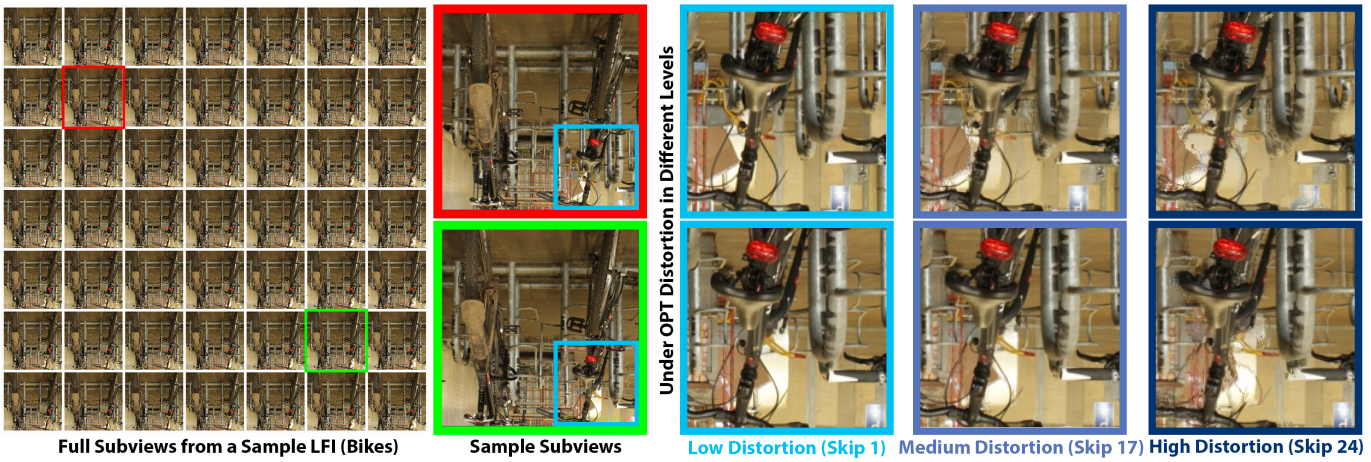

图 9：来自 MPI-LFA 的一个光场图像样本（Bikes），展示了两个角位置的子视图细节，随后是六张放大图像，用以呈现低（跳过 1）、中等（跳过 17）和高（跳过 24）失真级别的影响。

[2] J. Chen、J. Hou、Y. Ni 和 L.-P. Chau。基于部分遮挡区域超像素正则化的精准光场深度估计。IEEE 图像处理汇刊，2018年。[3] Y. Chen、G. Jiang、M. Yu、H. Xu 和 Y.-S. Ho。利用异构成像实现深度光场空间超分辨率。IEEE 可视化与计算机图形学汇刊，2022年。[4] M. Date、M. Isogai 和 H. Kimata。利用视觉等效光场实现全视差桌面式三维显示。载于2019年 IEEE 虚拟现实与三维用户界面会议（VR），第1297–1298页。IEEE，2019年。[5] F. M. Dekking、C. Kraaikamp、H. P. Lopuha¨a 和 L. E. Meester。概率与统计学现代导论：理解原因与方法。施普林格科学与商业媒体，2005年。[6] A. Dosovitskiy、L. Beyer、A. Kolesnikov、D. Weissenborn、X. Zhai、T. Un- t erthiner、M. Dehghani、M. Minderer、G. Heigold、S. Gelly 等。图像值得16×16个词：用于大规模图像识别的 Transformer。arXiv 预印本 arXiv:2010.11929，2020年。[7] M. Gonz´alez-Aud´ıcana、J. L. Saleta、R. G. Catal´an 和 R. Garc´ıa。基于小波分解，使用改进的 I HS 和 PCA 融合实现多光谱图像与全色图像的融合。IEEE 地球科学与遥感汇刊，42(6)：1291–1299，2004年。[8] K. Han、W. Xiang、E. Wang 和 T. Huang。一种新颖的遮挡感知投票代价，用于光场深度估计。IEEE 模式分析与机器智能汇刊，2021年。[9] H. Huang、H. Zeng、Y. Tian、J. Chen、J. Zhu 和 K.-K. Ma。光场图像质量评价：综述。载于2020年 IEEE 多媒体信息处理与检索会议（MIPR），第348–353页。IEEE，2020年。[10] I. Ihrke、J. Restrepo 和 L. Mignard-Debise。光场成像原理：简要回顾25年的研究。IEEE 信号处理杂志，33(5)：59–69，2016年。[11] Y. Itoh、T. Amano、D. Iwai 和 G. Klinker。高斯光场：光学透视头戴式显示器的视点相关模糊估计。IEEE 可视化与计算机图形学汇刊，22(11)：2368–2376，2016年。[12] J. Jin、J. Hou、J. Chen、S. Kwong 和 J. Yu。通过混合透镜的注意力引导融合实现光场超分辨率。载于第28届 ACM 多媒体国际会议论文集，第193–201页，2020年。[13] V. Kiran Adhikarla、M. Vinkler、D. Sumin、R. K. Mantiuk、K. Myszkowski、H.-P. Seidel 和 P. Didyk。面向稠密光场的质量指标。载于 IEEE 计算机视觉与模式识别会议论文集，第58–67页，2017年。[14] C. Koniaris、M. Kosek、D. Sinclair 和 K. Mitchell。具有实时视点相关重建的压缩动画光场。IEEE 可视化与计算机图形学汇刊，25 1680，2018年 4 1666( )：–。[15] Z. Liu、Y. Lin、Y. Cao、H. Hu、Y. Wei、Z. Zhang、S. Lin 和 B. Guo。Swin Transformer：使用移位窗口的分层视觉 Transformer。载于 IEEE/CVF 国际计算机视觉会议论文集，第10012–10022页，2021年。[16] Z. Lu、H. W. Yeung、Q. Qu、Y. Y. Chung、X. Chen 和 Z. Chen。利用四维光场和交错卷积神经网络改进图像分类。多媒体工具与应用，78 29227，2019年 20 29211( )：–。[17] M.-T. Luong、H. Pham 和 C. D. Manning。基于注意力的神经机器翻译的有效方法。arXiv 预印本 arXiv:1508.04025，2015年。[18] X. Lv、X. Wang、Q. Wang 和 J. Yu。基于光场超像素超图表示的四维光场分割。IEEE 可视化与计算机图形学汇刊，27(9)：3597–3610，2020年。[19] S. Ma、L. Zhu、X. Chen、X. Yan、S. Wang、P. Yang 和 B. Xu。ARFNet：用于光场显著目标检测的注意力导向细化与融合网络。IEEE 系统期刊，2022年。[20] X. Meng、R. Du、J. F. JaJa 和 A. Varshney。用于光场的三维内核注视点渲染。IEEE 可视化与计算机图形学汇刊，27(8)：3350–3360，2020年。[21] B. Mildenhall、P. P. Srinivasan、M. Tancik、J. T. Barron、R. Ramamoorthi 和 R. Ng。NeRF：将场景表示为用于视图合成的神经辐射场。ACM 通讯，65(1)：99–106，2021年。[22] A. Mittal、A. K. Moorthy 和 A. C. Bovik。空间域中的无参考图像质量评价。IEEE 图像处理汇刊，21(12)：4695–4708，2012年。[23] P. Paudyal、F. Battisti、M. Sj¨ostr¨om、R. Olsson 和 M. Carli。面向压缩光场图像的感知质量评价。IEEE《广播学报》，63(3):507–522，2017年。[24] Q. Qu、X. Chen、V. Chung 和 Z. Chen。基于深度与角度可分离卷积及辅助学习的光场图像质量评价。IEEE《广播学报》，67(4):837–850，2021年。[25] P. Ramachandran、B. Zoph 和 Q. V. Le。搜索激活函数。arXiv预印本 arXiv:1710.05941，2017年。[26] H. R. Sheikh 和 A. C. Bovik。图像信息与视觉质量。IEEE 图像处理汇刊，15(2):430–444，2006年。[27] L. Shi、S. Zhao 和 Z. Chen。BELIF：一种基于张量结构变化指数的光场图像盲质量评价器。载于：2019年 IEEE 国际图像处理会议（ICIP），第3781–3785页。IEEE，2019年。[28] L. Shi、S. Zhao、W. Zhou 和 Z. Chen。光场图像的感知评价。载于：2018年第25届 IEEE 国际图像处理会议（ICIP），第41–45页。IEEE，2018年。[29] L. Shi、W. Zhou、Z. Chen 和 J. Zhang。基于空间角度测量的无参考光场图像质量评价。IEEE《视频技术电路与系统汇刊》，30 4128，2019 11 4114( ): – 。[30] Y.-J. Tsai、Y.-L. Liu、M. Ouhyoung 和 Y.-Y. Chuang。用于光场视差估计的基于注意力的视图选择网络。载于：AAAI人工智能大会论文集，第34卷，第12095–12103页，2020年。[31] J. W. Tukey 等。探索性数据分析，第2卷。马萨诸塞州雷丁，1977年。[32] A. Vaswani、N. Shazeer、N. Parmar、J. Uszkoreit、L. Jones、A. N. Gomez、Ł. Kaiser 和 I. Polosukhin。注意力就是一切。神经信息处理系统进展，30，2017年。[33] L. Wald。高分辨率合成图像的质量：是否存在简单的判据？载于：第三届“地球数据融合：点测量、栅格地图与遥感图像的合并”会议，第99–103页。SEE/URISCA，2000年。[34] T.-C. Wang、A. A. Efros 和 R. Ramamoorthi。使用光场相机进行遮挡感知深度估计。载于：IEEE国际计算机视觉会议论文集，第3487–3495页，2015年。[35] Z. Wang 和 A. C. Bovik。一种通用图像质量指数。IEEE《信号处理快报》，9(3):81–84，2002年。[36] Z. Wang、A. C. Bovik、H. R. Sheikh 和 E. P. Simoncelli。图像质量评估：从误差可见性到结构相似性。IEEE 图像处理汇刊，13(4):600–612，2004年。[37] Z. Wang、E. P. Simoncelli 和 A. C. Bovik。用于图像质量评估的多尺度结构相似性。载于：第三十七届阿西洛马会议：信号、系统与计算机，第2卷，第1398–1402页。IEEE，2003年。[38] S. Wanner 和 B. Goldluecke。四维光场的全局一致深度标记。载于：2012年 IEEE 计算机视觉与模式识别会议，第41–48页。IEEE，2012年。[39] G. Wu、B. Masia、A. Jarabo、Y. Zhang、L. Wang、Q. Dai、T. Chai 和 Y. Liu。光场图像处理：综述。IEEE《信号处理精选主题期刊》，11(7):926–954，2017年。[40] R. Yang、X. Huang、S. Li 和 C. Jaynes。迈向光场显示：通过投影仪集群实现自动立体渲染。IEEE 可视化与计算机图形学汇刊，14(1):84–96，2007年。[41] H. W. F. Yeung、J. Hou、J. Chen、Y. Y. Chung 和 X. Chen。通过空间角度线索的深度由粗到细建模，实现快速光场重建。载于：欧洲计算机视觉会议（ECCV）论文集，第137–152页，2018年。[42] H. W. F. Yeung、J. Hou、X. Chen、J. Chen、Z. Chen 和 Y. Y. Chung。利用高效深度空间角度可分离卷积实现光场空间超分辨率。IEEE 图像处理汇刊，28 2330，2018 5 2319( ): – 。[43] C. Yim 和 A. C. Bovik。去块效应图像的质量评估。IEEE 图像处理汇刊，20(1):88–98，2010年。[44] H. Zhang、I. Goodfellow、D. Metaxas 和 A. Odena。自注意力生成对抗网络。载于：国际机器学习会议，第7354–7363页。PMLR，2019年。[45] L. Zhang、Y. Shen 和 H. Li。VSI：一种用于感知图像质量评价的视觉显著性诱导指数。IEEE 图像处理汇刊，23(10):4270–4281，2014年。[46] W. Zhou、L. Shi、Z. Chen 和 J. Zhang。张量导向的无参考光场图像质量评价。IEEE 图像处理汇刊，29:4070–4084，2020年。[47] D. Zwillinger 和 S. Kokoska。CRC标准概率与统计表及公式。CRC出版社，1999年。