arXiv:2301.01661v2 [cs.CV] 5 Sep 2023

RecRecNet：基于薄板样条模型和基于自由度的课程学习实现校正广角图像矩形化

KangLiao† LangNie† Chunyu Lin* Zishuo Zheng Yao Zhao 北京交通大学信息科学研究所 北京市先进信息科学与网络重点实验室{kang liao, nielang, cylin, yzhao}@bjtu.edu.cn

摘要

广角镜头在虚拟现实技术中具有广泛的应用前景，但会在拍摄的图像中引入严重的径向畸变。为了还原真实场景，先前的研究致力于校正广角图像的内容。然而，这种校正方案不可避免地会使图像边界发生形变，进而改变相关几何分布，并误导当前的视觉感知模型。在本研究中，我们提出了一种新的学习模型——矩形化校正网络（*RecRecNet*），以构建兼顾内容与边界的双赢表示。具体而言，我们提出了一个薄板样条（*TPS*）模块，用于对图像矩形化过程中的非线性、非刚性变换进行建模。通过学习校正图像上的控制点，我们的模型能够灵活地将源结构变换至目标域，并实现端到端的无监督形变。为降低结构近似的复杂度，我们进一步启发 *RecRecNet* 通过基于自由度（*DoF*）的课程学习来学习渐进式形变规则。通过在每个课程阶段逐步增加自由度，即从相似变换（*4*自由度）到单应变换（*8*自由度），网络能够探索更精细的形变，从而在最终的图像矩形化任务中实现快速收敛。实验表明，在定量和定性评估中，我们的方法均优于对比方法。代码和数据集可在 *https://github.com/KangLiao929/RecRecNet*。

1. 引言

由于视场（FoV）比传统镜头更大，广角镜头已广泛应用于计算成像、虚拟现实（VR）和自动驾驶等诸多领域 [21, 13, 50, 31]。通过

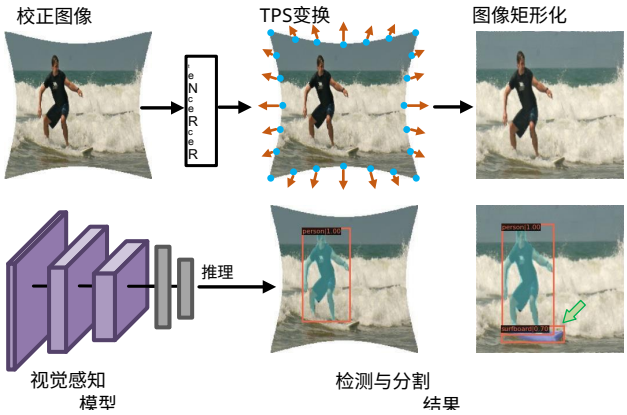


图 1. 现有校正方法致力于校正广角图像中的内容，但不可避免地会使图像边界发生畸变。这种变形边界会导致主流视觉感知模型 [17]的性能大幅下降。通过提出 RecRecNet，我们能够同时获得内容无畸变、边界规整的双赢表示。箭头标示了我们在感知方面的提升（例如，重新识别出冲浪板）。

采用特定映射后，广角镜头可以生成凸形的非直线图像，而非具有透视直线的图像。然而，这种映射机制违背了针孔相机假设，导致捕获的图像出现严重的径向畸变。

为消除广角镜头引起的畸变，人们提出了畸变校正算法。例如，传统方法利用标定靶标（例如，棋盘格） [54, 21] 和特定的场景假设 [10, 2]来标定广角镜头。近年来，深度学习为这一领域带来了新的启发。它能够基于学习到的语义特征，自动校正广角图像中的径向畸变。为追求更好的性能，近期研究通过多样化特征 [52, 49], 、学习范式 [30], 、标定表示 [26, 34], 和畸变先验 [7, 29],等改进学习框架。

我们注意到，上述畸变校正方法主

*通讯作者。 †前两位作者贡献相同。

要侧重于校正图像内容。通过将畸变图像向主点收缩，所有像素的空间分布都被重新排列，从而生成校正后的结果。然而，这种校正方式不可避免地会使图像边界发生畸变，导致图像呈现出视场狭窄、形状不规则的视觉效果。此外，我们发现，这种变形边界会显著影响当前主流的视觉感知模型，因为其中大多数模型都是在矩形图像数据集上训练的。如图1所示，Mask R-CNN [17] 会被校正后的图像所迷惑，导致边界附近出现分割和检测结果缺失（冲浪板被忽略）。尽管一些研究 [28] 采用图像外绘策略填充变形边界之外的空白区域， [40, 45], 但生成的内容通常是虚构的，并会扭曲原有的语义表征。

在本文中，我们探索了一种兼顾广角图像内容与边界的双赢校正表示，且无需引入任何额外的语义信息 [16, 37]。对于这种新表示，失真内容得到了校正，同时图像边界保持笔直，如图1所示。为此，我们提出了一种矩形化校正网络（称为**RecRecNet**）。具体而言，我们提出了一种薄板样条（TPS）运动模块，用于构建矩形化校正后广角图像的非线性、非刚性变换。通过学习校正图像上的控制点，该模块能够灵活地将源结构变形至目标域。控制点分布于整幅图像中，而图1 为便于直观说明，仅展示了其中一部分。

随后，我们设计了一种基于自由度的课程学习，进一步促使 RecRecNet 掌握渐进式变形规则，减轻复杂结构近似的负担，并为训练提供更好的起点。课程学习的一般范式 [3]为学习模型的训练引入了一种启发式且有目的的策略。模型可按照课程中的不同阶段，逐步学习各种样本或知识，从而更快收敛，这一过程模拟了人类和动物的学习方式。在此基础上，我们逐步提高所设计阶段/变换的自由度（DoF），从相似变换（4自由度）提升至单应变换（8自由度）。此外，我们的课程学习过程呈现出由简单到复杂的顺序：图像边界依次从直线、斜线变化到曲线。因此，RecRecNet 能够发现更多几何细节，并在最终的矩形化阶段更快收敛。

在致力于实现双赢的校正表示的同时，我们还深入分析了变形边界为何会导致感知变形，即，为何变形的图像边界会显著影响视觉感知模型。我们惊讶地发现，变形边界会在原始特征图上引入新特征，进而形成新语义，或导致非显著特征出现盲区。我们还证明，这种效应普遍存在于其他研究领域。

我们的主要贡献总结如下：

- 我们提出了一种矩形化校正网络（RecRecNet），用于对校正后的广角图像进行双赢表征。据我们所知，此方法此前尚未在相关文献中得到研究。
- 我们提出了一种薄板样条（TPS）运动模块，以灵活地建模非线性、非刚性的矩形化变换。我们还深入分析了变形后的图像边界为何会显著影响视觉感知模型。
- 为掌握渐进式变形规则并减轻复杂结构近似的负担，我们设计了基于自由度的课程学习，以引导RecRecNet。

2. 相关工作

2.1. 畸变校正

广角镜头的畸变校正方法可分为传统方法和基于学习的方法。大多数传统技术侧重于检测手工设计特征 [35, 24, 10, 2]。然而，这些技术常因铅垂线和曲线等特定约束而失效。近年来，研究者开始利用深度学习进行盲畸变校正。该方法无需任何场景假设，即可校正广角图像的径向畸变。例如，基于回归的方案 [39, 4, 52, 34, 29] 会预测输入图像的畸变参数，然后利用估计出的参数离线或在线校正畸变。基于重建的方案 [30, 26, 51, 7] 则直接学习从畸变图像到校正结果的像素级映射或位移场。

尽管这些校正方法具有良好的灵活性和效率，但在校正内容时不可避免地会使图像边界发生畸变。因此，边界处的畸变分布会干扰视觉感知网络，导致性能下降。一种直接的方法是将校正图像裁剪成矩形 [15], ，但这会舍弃有价值的内容，也违背了广角镜头的初衷。

2.2. 课程学习

课程学习技术已在机器学习的各个领域得到有效应用。Elman 等人 [11] 提出了一个课程学习的概念框架

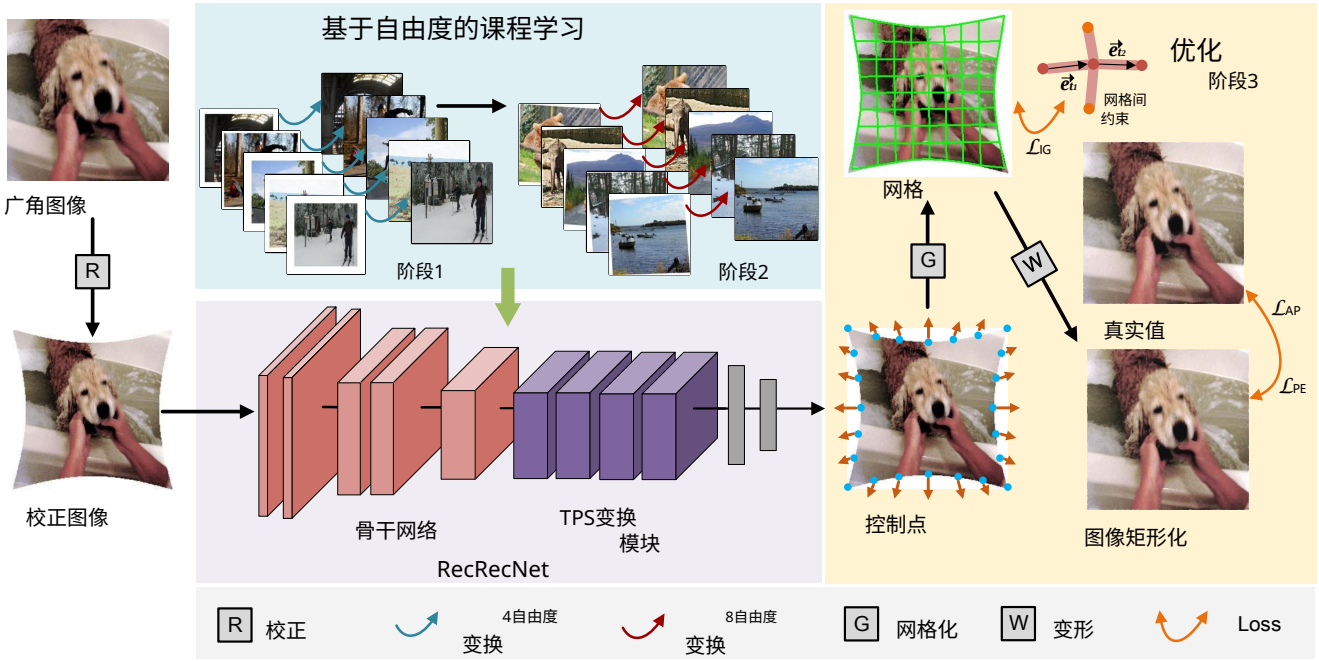


图2. 所提出框架的概述。我们的矩形化校正网络（RecRecNet）旨在构建一种兼顾图像内容和图像边界的双赢表征。RecRecNet 基于薄板样条（TPS）运动模型学习灵活的局部变换，预测图像边界上的控制点，并利用网格对目标图像进行变形。该网络通过外观约束 $\mathcal{L}_{AP}$、感知约束 $\mathcal{L}_{PE}$ 和网格间约束 $\mathcal{L}_{IG}$ 进行联合优化。为了进一步缓解结构近似带来的挑战，我们启发 RecRecNet 通过基于自由度的课程学习来学习渐进变形规则。

，强调从较低难度开始，再逐步提升到更具挑战性的案例。Bengio等人 [3]首次将课程学习正式定义为一种策略，即逐步提高训练过程中所用数据样本的复杂度。此后，大多数研究者遵循这一范式，将课程学习应用于众多领域 [42, 46, 38]。尤其是在计算机视觉领域，研究者探索了课程学习，以启发网络进行图像分类 [6, 36, 14], 、目标检测 [25, 47, 43],和语义分割 [53, 41, 12],等任务。

在这项工作中，我们将课程学习引入一项新任务。我们根据变换的自由度，将矩形化校正划分为三个层级。所提出的课程能够引导网络按由简到繁的顺序学习具有挑战性的形变，从而加快训练收敛。

3. 方法

3.1. 问题定义

给定一幅边界发生形变的校正后广角图像 $I_{Rec} \in \mathbb{R}^{h\times w\times 3}$，我们的图像矩形化校正网络（RecRecNet）旨在构建兼顾图像内容与图像边界的双赢表示，生成矩形的校正图像 $I_{Rec^2} \in \mathbb{R}^{h\times w\times 3}$。图2展示了所提出框架的整体概览。具体来说，

图像矩形化模块以 I_{Rec} 为输入，并基于薄板样条（TPS）运动模型 $\mathcal{T}$ 学习灵活的局部变换 [5, 20, 33, 9]。与以往的 TPS 工作相比，我们的方法旨在构建非线性非刚性矩形化变换。在弯曲图像边界上准确定位控制点是最大的挑战。借助外观约束和网格间约束，我们可以通过端到端无监督方式学习控制点，同时减少形变。为了进一步缓解结构近似带来的挑战，我们通过基于自由度的课程学习，引导网络学习渐进变形规则。

通用多项式相机模型 [21] 广泛用于近似广角图像的径向畸变。我们将镜头的投影方式与其畸变参数 $\mathcal{K} = [k_1, k_2, k_3, \cdots]$ 之间的关系表示如下：

$$r(\theta) = \sum_{i=1}^{N} k_i \theta^{2i-1}, N = 1, 2, 3, \cdots, \tag{1}$$

其中， r 是图像中主点与像素之间的距离， θ 表示光轴与入射光线之间的夹角。这种投影会产生凸形的非直线投影图像，因此拍摄的图像会受到严重的几何畸变。假设广角图像和

校正图像的像素坐标分别为 $\boldsymbol{p}=(x,\ y)$ 和 $\boldsymbol{p}_{rec}=(x_{rec},\ y_{rec})$，则校正方案可描述如下：

$$\boldsymbol{p}_d=\mathcal{R}(\boldsymbol{p}_{rec},\mathcal{K})=\begin{pmatrix}u_0\\v_0\end{pmatrix}+\frac{r(\theta)\boldsymbol{p}_{rec}}{\|\boldsymbol{p}_{rec}\|_2}, \tag{2}$$

其中（u_0，v_0）表示主点。利用上述公式，可以通过双线性插值对广角图像进行校正。通过将广角图像向主点方向收缩，重新排列所有像素的空间分布，从而获得校正后的内容。然而，这种校正方式不可避免地会使图像边界发生畸变。

3.2. 使用TPS变换进行矩形化

如公式2所示，广角图像的校正是一种非线性非刚性映射函数，无法通过简单的变换来表示。此外，利用估计参数构建精确的参数化模型来监督矩形化过程也颇具挑战性。从能量最小化的角度来看，我们可以通过采用均匀模型最小化变换能量，来近似矩形化校正：

$$\begin{aligned}\varepsilon&=\varepsilon_{\mathcal{T}}+\lambda\varepsilon_d,\\ \mathcal{T}:(x_{rec},y_{rec})&\mapsto(x_{rec^2},y_{rec^2}),\end{aligned} \tag{3}$$

其中，ε 表示期望变换的总能量。$(x_{rec},\ y_{rec})\in\Omega_S$ 和 $(x_{rec^2},\ y_{rec^2})\in\Omega_T$ 分别表示源 S 域和目标 T 域中的点。$\varepsilon_{\mathcal{T}}$ 表示数据惩罚能量，ε_d 表示畸变能量。具有最低总能量的上述公式所对应的变换即为所需变换，其中设计了超参数 λ，用于平衡数据惩罚与畸变之间的能量。

具体而言，我们提出利用 TPS变换$\mathcal{T}'$ 来最小化上述能量函数，从而实现端到端无监督变形。TPS 变换灵活且非线性，因此能够表示更复杂的运动。给定对应图像中的两组控制点，它可以在两幅图像之间进行变形，同时将失真降至最低。假设 $\boldsymbol{q}=[q_1,q_2,\cdots,q_N]$ 和 $\boldsymbol{q}'=[q'_1,q'_2,\cdots,q'_N]$ 分别为源点和目标点，则数据项 $\varepsilon_{\mathcal{T}'}$ 可由控制点位移 $\sum_{i=1}^{N}\|\mathcal{T}'(\boldsymbol{q}_i)-\boldsymbol{q}'_i\|_2^2$的差异计算得到。在对齐控制点后，TPS 可求得一个使失真项 ε_d 最小的最优插值变换，具体如下：

$$\min\iint_{\mathbb{R}^2}\left(\left(\frac{\partial^2\mathcal{T}'}{\partial x^2}\right)^2+2\left(\frac{\partial^2\mathcal{T}'}{\partial x\partial y}\right)^2+\left(\frac{\partial^2\mathcal{T}'}{\partial y^2}\right)^2\right)dxdy. \tag{4}$$

上述公式利用二阶导数来刻画每个目标点的失真偏差，并约束其累积值达到全局最小。由此，我们可以得到一个由控制点参数化的空间变形函数，如下所示：

$$\mathcal{T}'(q)=A\begin{bmatrix}q\\1\end{bmatrix}+\sum_{i=1}^{N}w_iU\left(\|\boldsymbol{q}'_i-q\|_2\right), \tag{5}$$

其中 q 是位于校正后的广角图像中的一个点。$A\in\mathbb{R}^{2\times3}$ 和 $w_i\in\mathbb{R}^{2\times1}$ 是变换参数 [5],，可通过最小化式子推导得出3。此外，$U(r)$ 是径向基函数，用于表示控制点对 $\boldsymbol{q}$的影响：$U(r)=r^2$ 对数 r^2。

在 RecRecNet 中，我们定义 $N=(U+1)\times(V+1)$ 个控制点，这些控制点可以连接起来形成网格，网络则学习预测校正后的广角图像的网格。随后，薄板样条将预测网格变换为目标矩形图像上的规则网格，其中控制点在图像上均匀分布。

对于网络，我们首先选用 ResNet50 [18] 作为 RecRecNet 的骨干网络，以提取高层语义特征。假设输入校正图像的尺寸为$W\times H$，骨干网络输出的特征图尺寸为 $\frac{W}{16}\times\frac{H}{16}$。随后，将这些特征图输入运动头，以预测校正图像上的$(U+1)\times(V+1)$ 个控制点。将 4 个核尺寸为 3 且带有批归一化的卷积层堆叠起来，以聚合高层特征，所有层的通道维度均设为 512。随后，使用 3 个全连接层，其单元数依次为 4096、2048、$(U+1)\times(V+1)\times2$来预测所有控制点的坐标。

3.3 基于自由度的课程学习

如上所述，我们提出利用 TPS 变换，通过一组控制点灵活地近似从校正后的图像到矩形化结果的运动。为了进一步降低这一变换带来的挑战，我们为 RecRecNet 引入了一种课程学习策略。

3.3.1 图像变换的形式化表述

图像变换包含平移、缩放、旋转、剪切等不同元素。这些元素的任意组合都可以形成具有一定自由度（DoF）的新变换。通常，自由度是指一组独立位移，它们能够完整描述系统发生位移或变形后的位置。二维图像中存在一些经典变换：平移变换（2自由度）、欧几里得/刚性变换（平移 + 旋转，3自由度）、相似变换（平移 +旋转 + 缩放，4自由度）、仿射变换（平移 +旋转 + 缩放 +剪切，6-

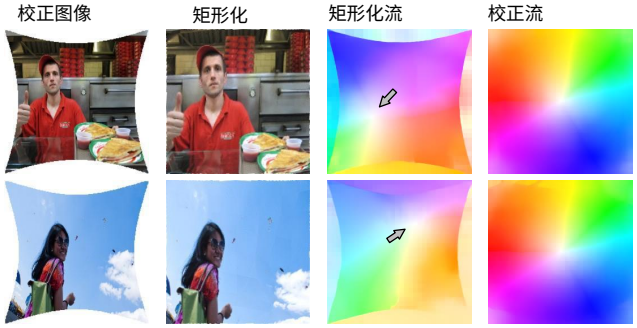


图 3. 校正后的广角图像的矩形化流和校正流可视化。我们用箭头标出了不同矩形化流中的不同径向中心。

自由度），以及射影变换（平移、旋转、缩放和无穷远线各有两个自由度，8自由度）。

校正后的广角图像与其矩形化图像之间的图像变换此前从未得到研究和形式化。与畸变校正相比，由于图像边界更为重要，矩形化变换似乎是一种逆向的非对称变形。为了直观地理解这种变换，我们将校正后的广角图像与其矩形化图像之间的像素位移场可视化。具体而言，我们利用最先进的光流估计网络 RAFT [44] 来展示运动差异，并将其命名为矩形化流（校正图像 $\mapsto$ 矩形化图像）。如图 3所示，与校正流（广角图像 $\mapsto$ 校正图像）相比，我们有以下观察结果：（1）矩形化流和校正流都呈现径向空间分布，其中四个主要方向分别朝向或背离图像的四个角。（2）每个样本的矩形化流都有一个不固定的径向中心，该中心由图像的内容和布局决定。相比之下，校正流的径向中心相对固定，这与广角图像发生径向畸变之前的几何先验有关。

因此，我们可以得出结论：图像内容高度决定了矩形化流的分布。使用经典图像变换及其组合来表示这种变换十分困难。不过，这种非线性、非刚性的变换可以在局部解耦为不同的变换元素，例如平移、缩放和剪切。启发网络学习一些基本变换，并逐步将其迁移到更具挑战性的情形中，将很有意义。

3.3.2 课程选择

为图像矩形化任务选择有效的课程时，我们应遵循两项原则：第一，课程中的各个阶段都应包含与目标密切相关的知识；第二，课程中的各个阶段应按由简到繁的顺序排列。为此，我们设计了一种基于自由度的课程，以启发 RecRecNet：我们逐步增加所设计变换的自由度，从相似变换（4自由度）到单应变换（8自由度），最终过渡到矩形化变换。

我们认为，所提出的课程具有以下优势：(1) 如第 3.3.1 节所述，自由度直接体现了图像变换的难度。随着自由度增加，变换呈现出更复杂的全局映射和更多样的运动。因此，RecRecNet 能够学习渐进变形规则。(2) 对于局部几何结构，即 即图像边界的形状，我们的课程呈现出从直线、斜线到曲线的渐进变化。这种设计可以提升 TPS 变换中控制点的定位能力，并使 RecRecNet 具备边界感知能力。更多详情和实验将在第 4.4 节中介绍。

3.4. 训练损失

外观损失： 我们约束矩形化图像I_{Rec^2} 在像素级别上接近真实值 I_{GT}。外观损失 $\mathcal{L}_{AP}$ 可通过 I_{Rec^2} 与 I_{GT} 之间的差异，使用 $\mathcal{L}_1$ 范数计算得出。

感知损失： 随后，我们最小化 I_{Rec^2} 与 I_{GT} 之间的 $\mathcal{L}_2$ 距离，以保证矩形化结果在高层语义感知上显得自然：

$$\mathcal{L}_{PE} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} ||\phi_{i,j}(I_{Rec^2}^{x,y}) - \phi_{i,j}(I_{GT}^{x,y})||_2, \tag{6}$$

其中，矩形化校正图像与真实值之间的差异在特征图 $\phi_{i,\ j}$上进行最小化。该特征图取自 VGG19 网络中第 j 个卷积层（激活后）、第 i个最大池化层之前。

网格间网格损失： 为避免矩形化后出现内容失真，预测网格不应发生大幅变形。因此，我们设计了网格间网格项$\mathcal{L}_{IG}$ 来约束预测网格的形状，促使相邻网格进行一致变换。两个连续变形网格 $\{\vec{e}_{t1}, \vec{e}_{t2}\}$ 的边缘受到约束，使其共线，具体如下：

$$\mathcal{L}_{IG} = \frac{1}{M} \sum_{\{\vec{e}_{t1}, \vec{e}_{t2}\} \in m_s} (1 - \frac{\langle \vec{e}_{t1}, \vec{e}_{t2} \rangle}{\| \vec{e}_{t1} \| \cdot \| \vec{e}_{t2} \|}), \tag{7}$$

其中， M 表示网格 m_s中两条相邻边组成的元组数量。当最大化上述余弦表示时，即其值为 1，且对应的两条边共线。因此，损失达到最小值，图像内容保持一致。

总体训练损失可通过以下方式计算：

$$\mathcal{L} = \lambda_{AP}\mathcal{L}_{AP} + \lambda_{PE}\mathcal{L}_{PE} + \lambda_{IG}\mathcal{L}_{IG}, \tag{8}$$



图4. 与图像矩形化工作 He等人 [16]的比较。为进行全面评估，我们将测试数据集划分为复杂场景（左）和简单场景（右）。

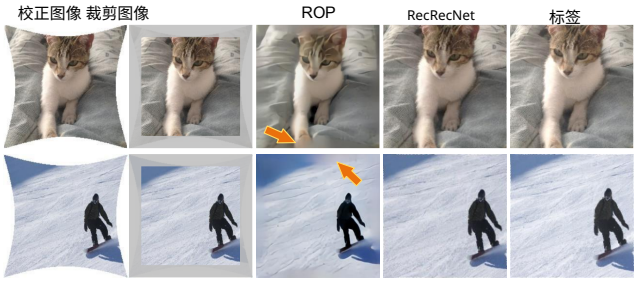


图5. 与校正外扩工作 ROP [28]的比较。我们的 RecRecNet 可以拉直校正图像中的不规则图像边界，而不会引入新内容。

其中， λ_{AP}、 λ_{PE} 和 λ_{IG} 是用于平衡外观损失、感知损失和网格间网格损失的权重，根据经验分别设为 1、 $1e^{-4}$ 和 1。

4. 实验

4.1. 实现细节

数据集构建： 沿用现有方法 [39,52, 30, 51], , 我们首先使用 4^{th} 阶多项式模型合成广角图像，依据公式 1。然后，我们依据公式 2对广角图像进行畸变校正。由于非线性和非刚性特征，要为校正图像的矩形化建立准确的变换模型很困难。我们注意到，He等人在计算机图形学领域提出了一种经典的全景图像矩形化方法 [16]，通过优化能量函数并进行保持直线的网格形变，使拼接图像变得规整。

因此，我们在校正图像上采用相同的能量函数，以适应我们的任务。然而，保持线性结构的能力 [16] 受直线检测的限制。因此，部分矩形化图像仍存在不可忽视的畸变。为解决这一问题，我们仔细筛选所有结果，并重复三次筛选流程，最终从 30,000 张源图像中得到 5,160 张训练数据，从 2,000 张源图像中得到 500 张测试数据。每次人工操作约需 10 秒。有关数据集的更多详情见补充材料。我们希望发布该数据集，以推动相关领域的发展。

实验配置： 我们使用初始值为 10^{-4} 的指数衰减学习率，并采用 Adam[23] 训练 RecRecNet。批量大小设为 16，总训练轮次设为 260。具体而言，根据所提出的基于自由度的课程，RecRecNet 分三个阶段进行训练。4自由度课程、8自由度课程和最终矩形化课程的训练轮次经经验划分，分别为 30、50 和 180。

4.2. 矩形化校正结果

RecRecNet是首个针对校正后的广角图像进行矩形化的工作。我们将我们的方法与经典的图像矩形化工作 He等人 [16] 以及针对校正图像的最先进外扩工作进行比较 [28]。在第一次比较中（图4），我们根据图像特征将测试数据集划分为复杂场景和简单场景。我们还展示了裁剪结果，即直接丢弃

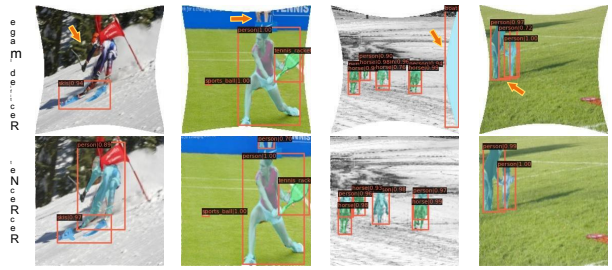


图6. RecRecNet 有助于下游视觉感知任务。箭头标出了校正图像中的失败结果。

变形边界周围内容的结果，这种做法在以往的工作中很常见。实验结果表明，尽管RecRecNet基于He等人筛选后的标签进行训练 [16], , 但它能够学习到更高的性能上限，并泛化到更多样化的场景。相比之下，He等人 [16] 的方法由于采用保线规则，无法对布局复杂或背景单调的场景进行矩形化。在第二次比较中（图5），我们可以观察到ROP [28] 通过外推图像内容，使校正图像的变形边界变得规整。但生成的结果通常并不真实，还会扭曲原始语义表征，导致边界附近出现令人不适的语义断裂。在不引入新颖且模糊内容的前提下，我们的RecRecNet通过灵活的TPS变换，为校正图像构建了双赢的表征，其外观与矩形化标签最为接近。此外，图6 显示，RecRecNet能够显著助力下游视觉任务。

表 1. 对比方法的定量评估。

指标	Crop	填充	ROP [28]	He [16]	Ours
峰值信噪比 ↑	11.51	12.07	13.90	15.36	18.68
结构相似性 ↑	0.1907	0.2775	0.3516	0.4211	0.5450
AP ↑	21.8	23.9	30.8	34.7	41.3
平均交并比 ↑	18.6	20.1	26.1	30.2	37.8

此外，我们从两个方面对不同方法进行定量评估比较。对于矩形化结果，我们选用峰值信噪比（PSNR）和结构相似性（SSIM）来衡量图像质量。对于视觉任务，我们采用平均精度（AP）和平均交并比（mIoU）来评估视觉模型的检测与分割性能 [17]。比较结果列于表1中，其中我们采用镜像填充作为填充方法，简单地填充变形边界之外的空白区域。可以看出，评估结果表明，RecRecNet不仅在低层次视觉重建方面具有优势，在高层次语义恢复方面也表现出色。

4.3. 为什么变形边界会导致感知变形

虽然我们的任务是将校正图像进行矩形化，但我们仍然对感知性能背后的秘密感到好奇

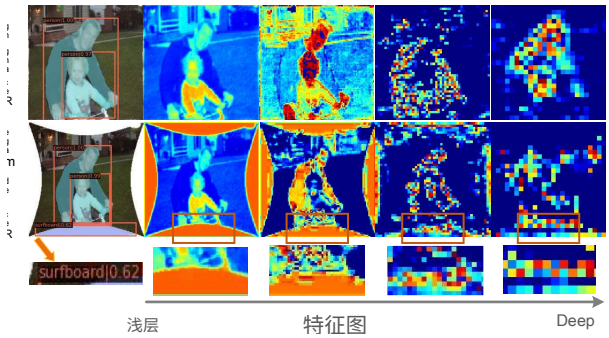


图7. Mask R-CNN 的失败案例 [17]（关于错误感知）。我们发现，变形边界会引入新的
特征到原始特征图中。

视觉感知模型性能下降的原因。在本节中，我们将对这种影响进行深入分析。首先，我们概括了感知模型的两种典型失效情况：错误感知和感知缺失。具体而言，我们重点关注 Mask R-CNN 的检测性能 [17]，并在图6和图7中展示了一些样本。我们可以观察到，该模型无法检测物体，尤其是那些位于边界附近的物体。通过将变形边界进行矩形化，感知性能得以出人意料地恢复。

在实验中，我们发现变形边界会向特征图中引入新特征，这些特征会（i）形成新语义（导致错误感知），或（ii）造成原始特征的盲区（导致感知缺失）。特别是，在图7中，我们展示了采用 ResNet 骨干网络的 Mask R-CNN 从浅层到深层的特征图101 [18]。在最外侧边界和变形边界处，分别可以观察到明显的线状伪影和曲线伪影。具体而言，线状伪影 [1]本质上是由零填充产生的。当卷积核通过零填充提取边界处的特征时，零值与原始内容之间的急剧变化会被错误地识别为边缘。同样，校正图像的变形边界也会引发这种边缘效应。随着卷积层的增加，这种效应会逐渐放大，并在特征图上构造出新特征。

对于错误感知案例，网络错误地将边界及其周围区域识别为冲浪板。由于边缘效应会在浅层特征图上引入新特征，并形成新语义

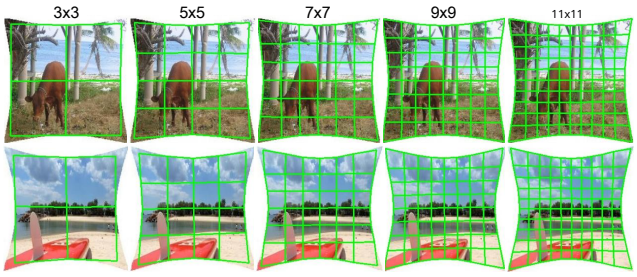


图8. 薄板样条中控制点大小的消融研究。

Tabl表 2. 控制点数量的消融研究 i n 薄板样条。

Size	峰值信噪比 ↑	结构相似性 ↑	FID ↓	LPIPS ↓
3×3	17.96	0.5134	25.00	0.1347
5×5	18.41	0.5288	21.49	0.1221
7×7	18.59	0.5385	19.62	0.1153
9×9	**18.68**	**0.5450**	**19.01**	**0.1136**
11×11	18.64	0.5392	19.16	0.1139

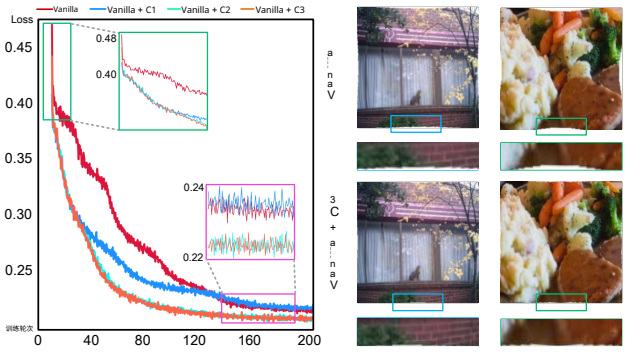


图9. 图像矩形化任务中课程学习的消融研究。C1、C2 和 C3 表示不同的课程。

在深层特征图中。更多结果见补充材料。此外，变形边界不仅存在于校正图像中，也存在于图像变形、图像拼接和滚动快门校正等其他研究领域。因此，边缘延拓效应会在特征图上引入新特征，并进一步影响下游感知性能。

4.4. 消融研究

控制点：如表2和图8所示，实验表明，增加控制点有助于提升 RecRecNet 对变形边界的结构近似能力和定位能力。此外， 9×9在所有评估指标上均取得了最佳性能。

课程学习： 我们将不使用课程学习的 RecRecNet 定义为 Vanilla，并将使用不同课程的 RecRecNet 定义为：2 自由度 $\mapsto$ 4-自由度 Vanilla + C1、2自由度 $\mapsto$ 8- 自由度 Vanilla + C2，以及 4自由度 $\mapsto$ 8-自由度 Vanilla + C3。在图 9（左）中，训练损失曲线表明，课程学习能够为训练提供更好的初始化点，并加快矩形化任务的收敛速度。与课程设置相比，Vanilla 训练大约需要多 40 个训练轮次才能达到相近的训练性能，但容易在验证数据上过拟合。此外，合理的课程设计（例如 Vanilla + C3）有助于提升 RecRecNet 感知边界的能力，从而增强结构恢复性能，如图 9（右）所示。

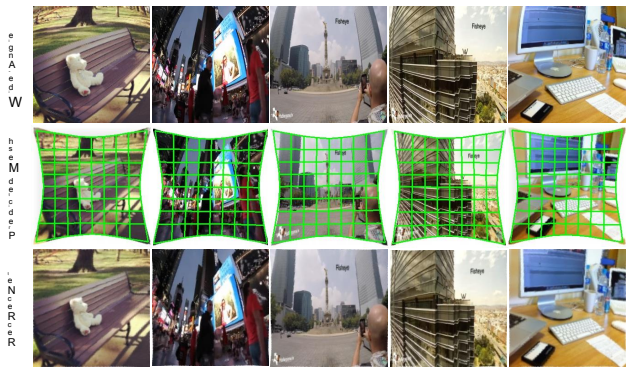
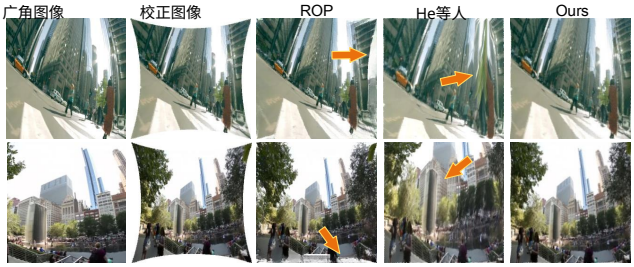


图10. 对真实世界广角图像和其他类型合成数据集进行的跨域评估。



Fi图 11. 跨域评估的比较 结果。

4.5. 跨域评估

为评估泛化能力，我们收集了 300 个来自最先进校正方法的校正后的广角图像结果 [29, 51]。这些结果源自各种类型的数据集和真实世界广角镜头，例如 Rokinon 8mm Cine Lens、Opteka 6.5mm Lens 和 GoPro。这些镜头通常用于获取全景图像。图10 显示，尽管 RecRecNet 仅使用一种相机模型的合成图像数据集进行训练，但仍能很好地泛化到其他领域和不同的校正结构。我们还展示了一些对比结果，如图 11所示，其中可以观察到以往方法存在语义破碎和物体畸变。

5. 结论

在这项工作中，我们针对广角图像提出了一项新的矩形化校正任务。尽管以往文献从未研究过这一任务，我们证明了常见的校正后广角图像中的变形边界会显著影响视觉感知模型。为了解决这一问题，我们提出为校正后的广角图像构建一种双赢表征，并设计了全新的 RecRecNet。借助灵活的薄板样条变换运动模型，RecRecNet 能够以无监督端到端方式，将局部形变从变形边界转化为直线边界。此外，我们通过基于自由度的课程学习，引导 RecRecNet 学习渐进变形规则，从而缓解

非线性和非刚性变换的复杂性。我们还提供了详细分析，解释变形图像边界为何会导致当前视觉感知发生形变。在未来的工作中，我们计划将这一方法扩展为适用于任意变形图像的通用范式，并进一步研究图像边界与视觉感知性能之间的关系。此外，将我们的图像矩形化算法嵌入在线数据增强流程，用于训练视觉模型，也颇具研究价值。

参考文献

[1] Bilal Alsallakh、Narine Kokhlikyan、Vivek Miglani、Jun Yuan 和 Orion Reblitz-Richardson。注意填充——卷积神经网络可能产生盲区。载于国际学习表征会议，2020 7。J. P. Barreto 和 H. [2] Araujo。中心折反射线图像的几何属性及其在标定中的应用。《IEEE模式分析与机器智能汇刊》，27(8):1327–1333，2005。1, 2[3] Yoshua Bengio、J´erˆome Louradour、Ronan Collobert 和 Jason Weston。课程学习。载于国际机器学习会议论文集，第41–48页，2009。2, 3[4] Oleksandr Bogdan、Viktor Eckstein、Francois Rameau 和 Jean-Charles Bazin。Deepcalib：一种用于广角相机自动内参标定的深度学习方法。载于第15届ACM SIGGRAPH欧洲视觉媒体制作会议论文集，第1–10页，2018。2, 12[5] F.L. Bookstein。主变形：薄板样条与变形分解。《IEEE模式分析与机器智能汇刊》，11 585，1989 6 567 3, 4 ( ): – 。[6] Paola Cascante-Bonilla、Fuwen Tan、Yanjun Qi 和 Vicente Ordonez。课程标注：重新审视半监督学习中的伪标注。载于AAAI人工智能会议论文集，第35卷，第6912–6920页，2021。3[7] Chun-Hao Chao、Pin-Lun Hsu、Hung-Yi Lee 和 Yu-Chiang Frank Wang。用于鱼眼图像校正的自监督深度学习。载于ICASSP 2020——2020年IEEE国际声学、语音与信号处理会议（ICASSP），第2248–2252页，2020。1, 2[8] Daniel DeTone、Tomasz Malisiewicz 和 Andrew Rabinovich。深度图像单应性估计。arXiv预印本 arXiv:1606.03798，2016。12[9] Daniele Detone、Tomasz Malisiewicz 和 Andrew Rabinovich。用于几何匹配的卷积神经网络架构。IEEE计算机视觉与模式识别会议（CVPR）论文集，第1965–1974页，2018。3[10] Frederic Devernay 和 Olivier Faugeras。直线必须保持笔直。《机器视觉与应用》，13(1):14–24，2001。1, 2[11] Jeffrey L. Elman。神经网络中的学习与发展：从小处着手的重要性。《认知》，48:71–99，1993。2 [12] Zhengyang Feng、周倩玉、程光亮、谭鑫、施剑平和马立庄。通过动态自训练和类别均衡课程进行半监督语义分割。arXiv预印本 arXiv:2004 5, 2020 08514, 1 2 3 . ( ): . [13] 高少华、Kailun Yang、石浩、王凯威和白健。全景成像及其在场景理解中的应用综述。IEEE仪器与测量汇刊，71:1–34，2022年。1[14] 郭晟、黄伟林、张昊志、庄辰帆、董登科、Matthew R Scott和黄定龙。CurriculumNet：从大规模网络图像中进行弱监督学习。载于《欧洲计算机视觉会议（ECCV）论文集》，第135–150页，2018年。3[15] 何恺明、常慧文和孙剑。内容感知旋转。载于《IEEE国际计算机视觉会议论文集》，第553–560页，2013年。2, 12[16] 何恺明、常慧文和孙剑。通过变形实现全景图像矩形化。《ACM图形学汇刊（TOG）》，32(4):1–10，2013年。2, 6, 7, 12, 13[17] 何恺明、Georgia Gkioxari、Piotr Doll´ar和Ross Gir-shick。Mask R-CNN。载于《IEEE国际计算机视觉会议论文集》，第2961–2969页，2017年。1, 2, 7, 13, 14[18] 何恺明、张祥雨、任少卿和孙剑。用于图像识别的深度残差学习。载于《2016年IEEE计算机视觉与模式识别会议（CVPR）》，第770–778页，2016年。4, 7[19] Md Amirul Islam、贾森和Neil DB Bruce。卷积神经网络编码了多少位置信息？载于《国际学习表征会议》，2019年。14[20] Max Jaderberg、Karen Simonyan、Andrew Zisserman等。空间变换网络。载于《神经信息处理系统进展》，第2017–2025页，2015年。3[21] Juho Kannala和Sami S Brandt。适用于常规镜头、广角镜头和鱼眼镜头的通用相机模型与标定方法。《IEEE模式分析与机器智能汇刊》，28(8):1335–1340，2006年。1, 3[22] Osman Semih Kayhan和Jan C van Gemert。论卷积神经网络中的平移不变性：卷积层可以利用绝对空间位置。载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第14274–14285页，2020年。14[23] Diederik P Kingma和Jimmy Ba。Adam：一种随机优化方法。arXiv预印本 arXiv:1412 6980, 2014 6 . . [24] H. I. Koo。相机拍摄的印刷文档图像分割与校正。《IEEE多媒体汇刊》，15(3):647–660，2013年。2[25] 李思扬、朱祥鑫、黄勤、徐浩和C-C Jay Kuo。用于弱监督目标检测的多实例课程学习。arXiv预印本 arXiv:1711 09191, 2017 3 . . [26] 李晓宇、张博、 Pedro V. Sander 和廖静。通过深度学习对图像进行盲几何失真校正。载于《IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集》，

2019年6月。1, 2[27] Kang Liao、Chunyu Lin、Lixin Liao、Yao Zhao和Weiyao Lin。用于训练失真感知桶形畸变校正模型的多级课程。载于《IEEE/CVF国际计算机视觉会议论文集》，第4389–4398页，2021年。12, 15[28] Kang Liao、Chunyu Lin、Yunchao Wei、Feng Li、Shangrong Yang和Yao Zhao。通过曲线感知外推实现完整场景与规则形状的失真校正。载于《IEEE/CVF国际计算机视觉会议论文集》，第14569–14578页，2021年。2, 6, 7[29] Kang Liao、Chunyu Lin和Yao Zhao。用于失真校正的深度序数失真估计方法。《IEEE图像处理汇刊》，30:3362–3375，2021年。1, 2, 8[30] Kang Liao、Chunyu Lin、Yao Zhao和Moncef Gabbouj。Dr-GAN：使用条件生成对抗网络实时自动校正径向畸变。《IEEE视频技术电路与系统汇刊》，30(3):725–733，2020年。1, 2, 6, 12[31] Kang Liao、Lang Nie、Shujuan Huang、Chunyu Lin、Jing Zhang、Yao Zhao、Moncef Gabbouj和Dacheng Tao。深度学习在相机标定及其他领域的应用：综述。arXiv预印本 arXiv:2303.10559，2023年。1[32] Tsung-Yi Lin、Michael Maire、Serge Belongie、James Hays、Pietro Perona、Deva Ramanan、Piotr Doll´ar和C Lawrence Zitnick。Microsoft COCO：上下文中的常见物体。载于《欧洲计算机视觉会议》，第740–755页。Springer，2014 12 。Ziheng Liu、Ping Luo、[33] Xiaokang Wang和Xiaoou Tang。用于面部表情识别的薄板卷积神经网络。载于《IEEE计算机视觉与模式识别会议》，第8253–8261页。IEEE，2018年。3[34] Manuel Lopez、Roger Mari、Pau Gargallo、Yubin Kuang、Javier Gonzalez-Jimenez和Gloria Haro。带径向畸变的深度单图像相机标定。载于《IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集》，2019年6月。1, 2[35] Rui Melo、Michel Antunes、Jo˜ao Pedro Barreto、Gabriel Falc˜ao和Nuno Gonc¸alves。使用铅垂线方法从单帧进行无监督内参标定。载于《IEEE国际计算机视觉会议》，第537–544页，2013年。2[36] Pietro Morerio、Jacopo Cavazza、Riccardo Volpi、Ren´e Vidal和Vittorio Murino。课程式随机失活。载于《IEEE国际计算机视觉会议论文集》，第3544–3552页，2017年。3[37] Lang Nie、Chunyu Lin、Kang Liao、Shuaicheng Liu和Yao Zhao。用于图像拼接的深度图像矩形化：一种学习基线。载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第5740–5748页，2022年。2[38] Anastasia Pentina、Viktoriia Sharmanska和Christoph H Lampert。多项任务的课程学习。载于《IEEE计算机视觉与模式识别会议论文集》，第5492–5500页，2015年。3[39] Jiangpeng Rong、Shiyao Huang、Zeyu Shang 和 Xianghua Ying。利用合成图像训练的卷积神经网络进行径向镜头畸变校正。发表于亚洲计算机视觉会议，第 35–49 页。Springer，2016年。2，6，12[40] Mark Sabini 和 Gili Rusak。突破边界：利用生成对抗网络进行图像外扩。arXiv 预印本 arXiv:1808 08483，2018年 2。[41] Christos Sakaridis、Dengxin Dai 和 Luc Van Gool。用于夜间图像语义分割的引导式课程模型适配与不确定性感知评估。发表于 IEEE/CVF国际计算机视觉会议论文集，第 7374 7383, 2019 3 –。Yangyang [42] Shi、Martha Larson 和 Catholijn M Jonker。基于课程学习的循环神经网络语言模型适配。《计算机语音与语言》，33(1):136–154，2015年。3[43] Petru Soviany、Radu Tudor Ionescu、Paolo Rota 和 Nicu Sebe。用于跨域目标检测的课程式自步学习。《计算机视觉与图像理解》，204:103166，2021年。3[44] Zachary Teed 和 Jia Deng。RAFT：用于光流的循环全对场变换。发表于欧洲计算机视觉会议，第 402–419 页。Springer，2020年。5[45]Piotr Teterwak、Aaron Sarna、Dilip Krishnan、Aaron Maschinot、David Belanger、Ce Liu 和 William T Free-man。Boundless：用于图像扩展的生成对抗网络。发表于 IEEE/CVF国际计算机视觉会议论文集，第 10521–10530 页，2019年。2[46] Radu Tudor Ionescu、Bogdan Alexe、Marius Leordeanu、Marius Popescu、Dim P Papadopoulos 和 Vittorio Fer-rari。这能有多难？估计图像中视觉搜索的难度。发表于 IEEE计算机视觉与模式识别会议论文集，第 2157–2166 页，2016年。3[47] Jiasi Wang、Xinggang Wang 和 Wenyu Liu。基于课程学习的弱监督与半监督 Faster R-CNN。发表于2018年第24届国际模式识别会议（ICPR），第 2416–2421 页。IEEE，2018年。3[48] Rui Xu、Xintao Wang、Kai Chen、Bolei Zhou 和 Chen Change Loy。位置编码作为生成对抗网络中的空间归纳偏置。发表于 IEEE/CVF计算机视觉与模式识别会议论文集，第 13569–13578 页，2021年。14[49] Zhucun Xue、Nan Xue、Gui-Song Xia 和 Weiming Shen。学习校准直线以进行鱼眼图像校正。发表于 IEEE计算机视觉与模式识别会议论文集，第 1643–1651 页，2019年。1[50] Kailun Yang、Jiaming Zhang、Simon Reiß、Xinxin Hu 和 Rainer Stiefelhagen。捕获全方位范围上下文以实现全景分割。发表于 IEEE/CVF计算机视觉与模式识别会议论文集，第 1376–1386 页，2021年。1[51] Shangrong Yang、Chunyu Lin、Kang Liao、Chunjie Zhang 和 Yao Zhao。利用外观流进行鱼眼图像校正的渐进互补网络。发表于 IEEE/CVF计算机视觉

与模式识别会议（*CVPR*）论文集，第 6348–6357 页，2021年6月。2，6，8，12，15[52] Xiaoqing Yin、Xinchao Wang、Jun Yu、Maojun Zhang、Pas-cal Fua 和 Dacheng Tao。Fisheyerecnet：用于鱼眼图像校正的多上下文协作深度网络。发表于欧洲计算机视觉会议论文集（*ECCV*），2018年9月。1，2，6，12[53] Yang Zhang、Philip David 和 龚博清。用于城市场景语义分割的课程域适配。发表于*IEEE*国际计算机视觉会议论文集，第 2020–2030 页，2017年。3[54] 张正友。一种灵活的新型相机标定技术。*IEEE*模式分析与机器智能汇刊，22(11):1330–1334，2000年。1

A. 补充材料

A.1. 概述

本文档提供以下补充内容：

- 矩形化的合理性说明详情（第A.2节）。
- 数据集构建详情（第A.3节）。
- 对比方法和我们的方法的更多定性结果（第A.4节）。
- 更多视觉感知结果（第A.5节）。
- 更多跨域评估结果（第A.6节）。
- 用户研究（第A.7节）。

A.2. 图像矩形化的合理性

在正文中，我们已经证明，我们的图像矩形化方法能够显著促进下游视觉任务，提升目标检测和语义分割的性能。除了对视觉任务的益处之外，我们认为图像矩形化算法还有其他值得研究的理由。首先，图像矩形化能为人类提供视觉上令人愉悦的结构，也能为设备提供常规格式。以往研究 [16, 15] 表明，大多数用户更喜欢矩形边界，用于发布、分享和打印照片。例如，Flickr（flickr.com）上标签为“panorama”的图像中，超过 99% 都具有矩形边界。而且，人类视觉系统对不规则线条更加敏感。此外，矩形图像非常适合主流显示窗口和屏幕。相反，非矩形图像中的空白区域会占用无效空间，使数据存储和压缩效率低下。

A.3. 数据集构建细节

我们构建了一个矩形化校正数据集，这是该研究领域的首个此类数据集，并希望将其公开，以推动相关研究的发展。具体而言，我们通过以下四个步骤构建数据集：**(i) 广角图像与校正图像的合成**。由于收集大规模的成对广角图像及其校正真实值极具挑战性，我们遵循现有的畸变校正方法 [39, 4, 52, 30, 51] 来合成数据集。首先，从 MS-COCO 数据集 [32]中收集原始图像。我们采用 4^{th} 阶多项式模型来近似广角图像的径向畸变，已有研究验证该模型能够以较高精度适用于大多数投影模型。具体而言，从以下范围内随机生成四个畸变参数：$k_1 \in [-1\times10^{-4}, -1\times10^{-8}]$、$k_2 \in [1\times10^{-12}, 1\times10^{-8}]$ 或 $\in [-1\times10^{-8}, -1\times10^{-12}]$、$k_3 \in [1\times10^{-16}, 1\times10^{-12}]$ 或 $\in [-1\times10^{-12}, -1\times10^{-16}]$，以及 $k_4 \in [1\times10^{-20}, 1\times10^{-16}]$或 $\in [-1\times10^{-16}, -1\times10^{-20}]$。随后，我们对广角图像进行畸变校正，得到校正图像。**(ii) 对校正图像进行矩形化**。由于校正图像具有非线性和非刚性特征，要为其矩形化构建精确的变换模型十分困难。我们注意到，计算机图形学领域有一种经典的全景图像矩形化技术 He等人 [16]，该技术通过采用保线网格变形优化能量函数，使拼接图像变得规整。因此，我们在校正图像数据集上采用相同的能量函数，以适配矩形化校正任务。然而，He等人 [16] 的方法保留线性结构的能力受到直线检测的限制。因此，部分矩形化校正图像存在不可忽视的畸变。为解决这一问题，我们仔细筛选所有矩形化结果，并将筛选过程重复三次，最终从 30,000 张源图像中获得 5,160 个训练样本，并从 2,000 张源图像中获得 500 个测试样本。每次人工操作约需 10 秒。所有图像的尺寸均为 256×256。**(iii) 跨域验证**。除了合成数据集之外，我们还收集了 300 张由最先进的校正方法 [27, 51]生成的校正后广角图像结果。这些结果来自其他类型的数据集以及真实世界的广角镜头，例如 Rokinon 8mm Cine Lens、Opteka 6.5mm Lens 和 GoPro。**(iv) 基于自由度的课程数据集**。为降低图像矩形化任务中结构近似的难度，我们提出了一种基于自由度（DoF）的课程学习方法。具体而言，我们采用三个课程阶段来引导 RecRecNet，即从相似变换（4自由度）到单应变换（8自由度），再到矩形化变换。因此，我们还构建了两个数据集（4自由度数据集和8自由度数据集），用于提供基础变换知识，每个数据集均包含 5,000 对图像。在变换合成过程中，我们随机扰动原始图像的四个角点，并按照先前的工作 [8]将其变形为目标图像。

A.4. 更多定性比较结果

如图12所示，我们展示了更多定性对比结果。RecRecNet 能够对各种场景中的校正后的广角图像进行矩形化。可以观察到，RecRecNet 将变形边界拉直，并重新排列图像内容，使其保持不失真，从而为广角图像提供了双赢的校正表示。相比之下，以往的方法无法在校正后图像的图像边界和图像内容之间取得平衡，其结果通常会出现内容不完整或失真分布。例如，He等人生成的一些矩形化结果 [16]会旋转原始场景

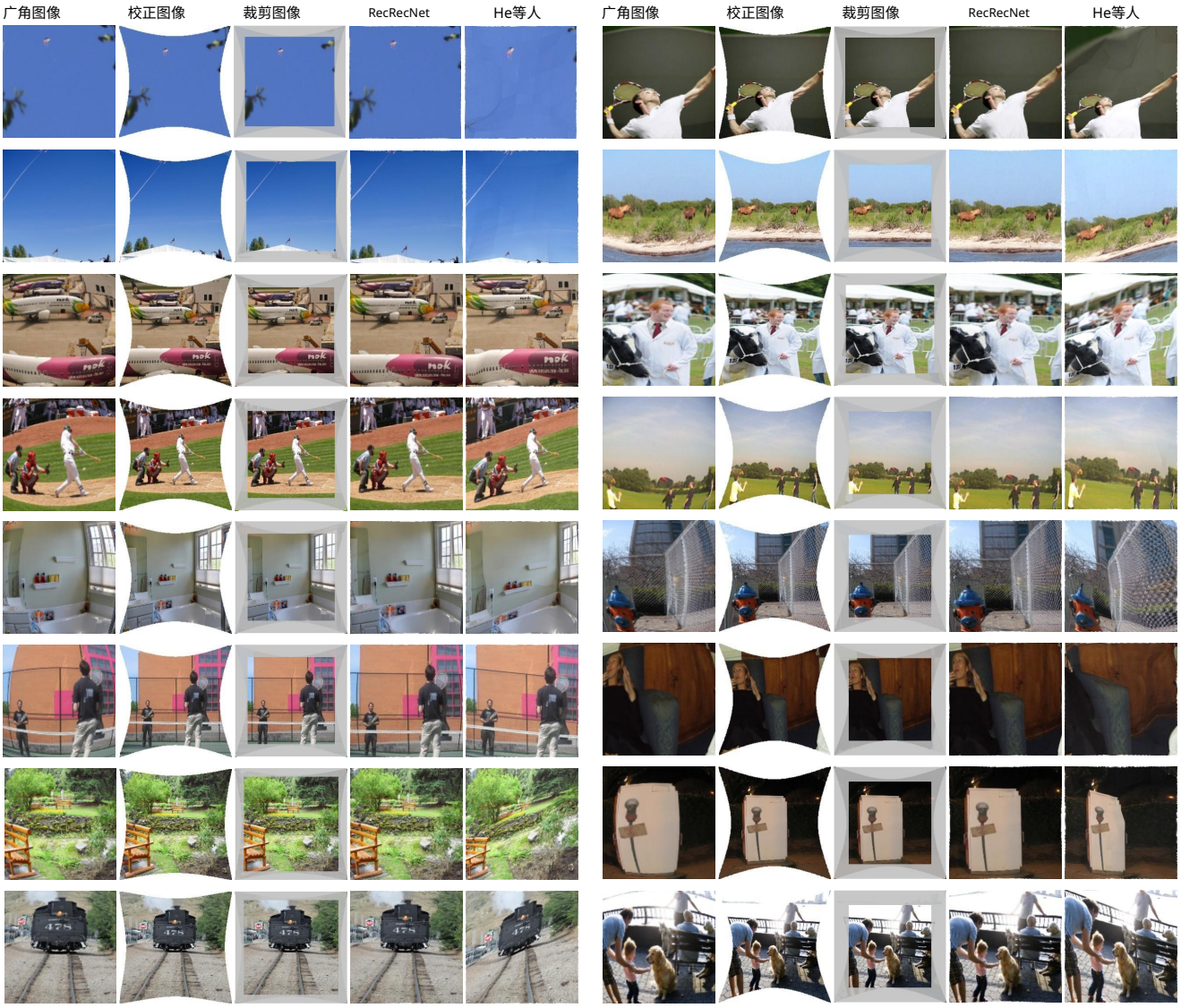


图 12. 与 He等人相比的更多定性结果。 [16]从左到右依次展示广角图像、校正图像、裁剪后的校正图像、RecRecNet的矩形化结果，以及He等人的矩形化结果。 [16]

并因其保持直线的网格变形而扭曲物体。

A.5. 更多视觉感知结果

如图 13所示，我们展示了更多视觉感知结果，包括由广受欢迎的感知模型 Mask R-CNN 得出的目标检测和语义分割结果 [17]。可以观察到，原始校正图像中的变形结构误导了学习模型，导致错误感知和感知缺失。得益于所提出的灵活 TPS 变换模块和基于自由度的课程学习，我们的 RecRecNet 能够通过消除变形问题，显著助力下游视觉任务

边界问题。感知性能得以恢复，尤其是在图像边界附近。校正图像中性能下降的一个直接原因是，感知模型在训练期间从未见过边界变形的图像，测试数据与训练数据之间的领域差距使其难以处理校正图像。尽管如此，我们仍希望进一步探索其深层原因，并揭示变形几何为何会导致视觉感知发生偏差。

图 14 展示了一个示例，用于说明为什么变形的图像边界会导致感知变形。正如正文所述，我们发现，变形边界会在特征图上引入 新特征 ，这些新特征会（i）形成新语义（导致错误的

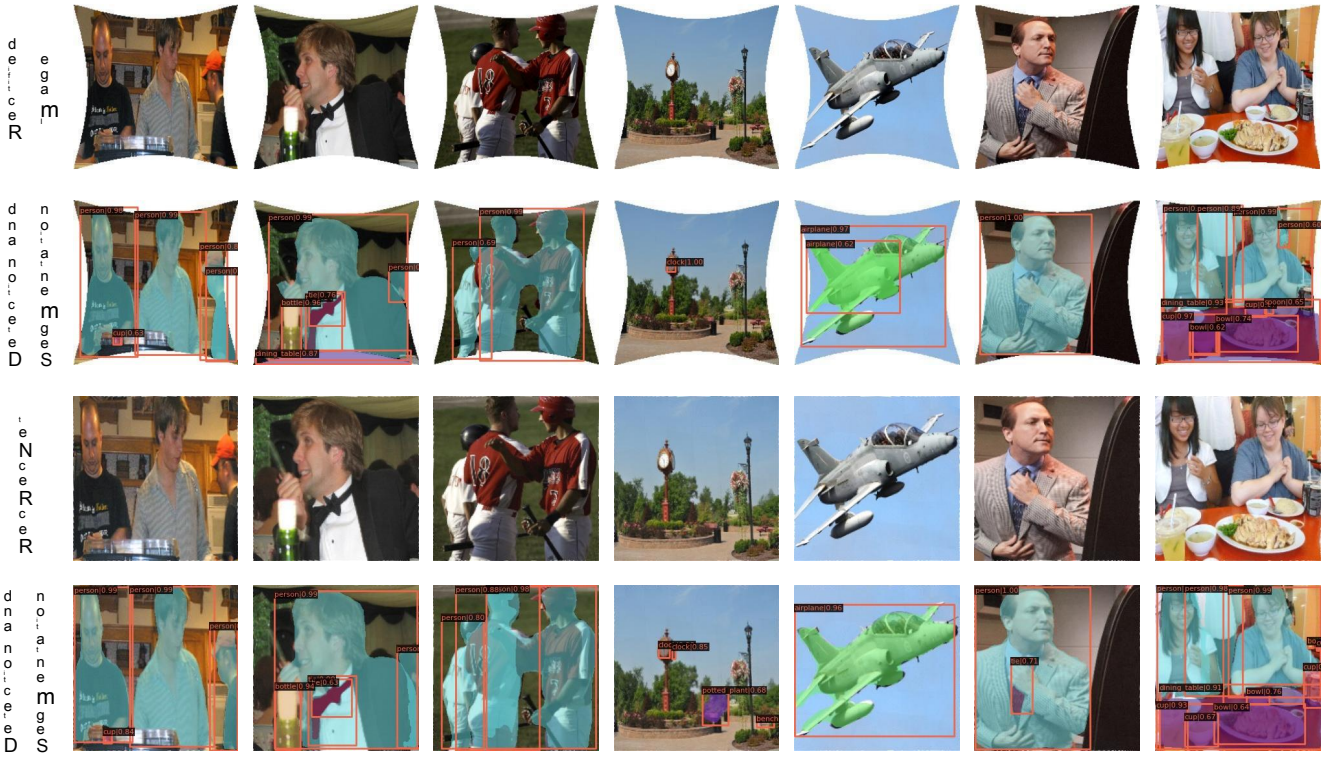


图13。更多检测和分割结果，由 Mask R-CNN 得出 [17] . 我们展示了校正后的广角图像及其视觉感知结果，以及我们的矩形化结果，以及其视觉感知结果，从上到下。

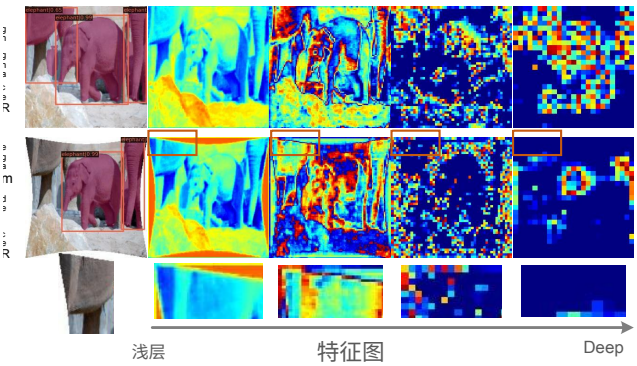


图14. Mask R-CNN [17] 在感知缺失方面的失败案例。我们发现，变形边界会为原始特征图引入新特征。因此，由于深层特征图中的原始特征受到干扰，网络无法识别变形图像边界附近的大象。

感知），或（ii）导致原始特征出现盲区（导致感知缺失）。在浅层特征图中，可以分别在最外侧边界和变形边界处观察到明显的线状伪影和曲线伪影。尤其是，线状伪影本质上是由零填充产生的。零填充是主流卷积神经网络架构的基本组成部分，其作用是通过向特征图添加零值，使整个网络中的特征图尺寸保持不变。当卷积核使用零填充来提取边界处的特征时，零值与原始内容之间的突变会被错误地识别为边缘。与常规图像相比，校正图像中的边缘效应延拓得更快。因此，推理过程中会涉及更多新特征。我们将其称为双边缘扩展效应。

近期研究还表明，零填充可能会在卷积神经网络 [19, 22, 48]中无意间泄露位置信息。对于具有变形边界的校正图像，我们认为，引入的新特征可能会干扰原有的位置信息。因此，感知模型无法构建准确的空间分布，也无法理解相邻语义之间的关系。此外，我们发现，变形边界引入的新语义特征会与其附近的特征共同得到理解。换言之，神经网络能够感知物体关系和场景上下文。

对于感知缺失的情况，如果边界附近的特征不显著，新引入的特征就可能覆盖这些特征，并为感知模型造成盲区。因此，网络无法识别变形图像边界附近的大象。我们认为

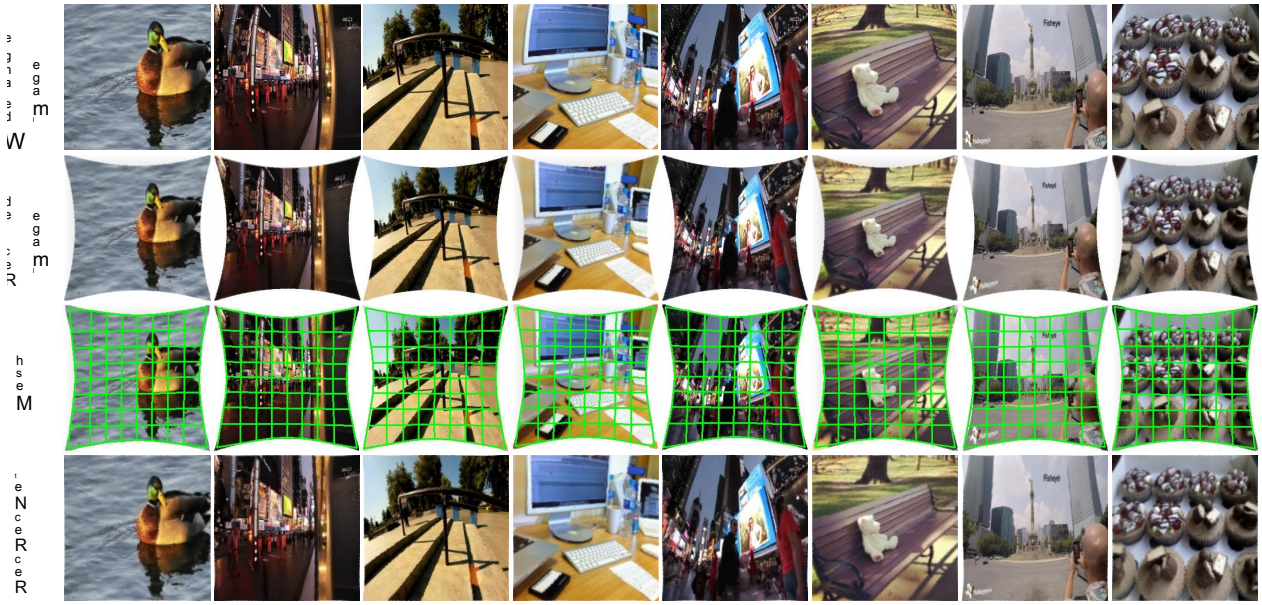


图15. 更多跨域评估结果。从上到下依次展示广角图像、校正图像、预测网格以及 RecRecNet 生成的矩形化结果。

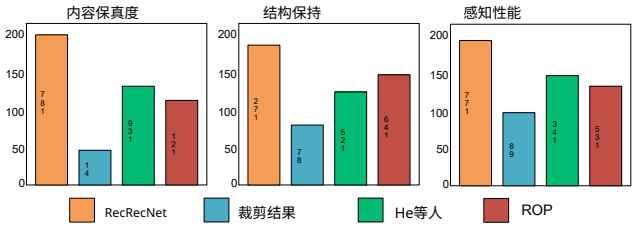


图16. 矩形化校正结果的用户研究。

这种现象可能出现在大多数图像变换场景中，例如图像变形。在这些场景中，边界会发生变形，以权衡图像内容方面的不同要求。

A.6. 更多跨域评估结果

如第A.3节所述，我们从最先进的校正方法 [27, 51] 中收集了 300 个校正后的广角图像结果，以进行跨域评估。这些方法的结果来自其他类型的合成数据集，以及采用不同相机模型的真实世界广角镜头。图15 展示了跨域评估结果。尽管从未见过新的数据域，我们的 RecRecNet 通过学习灵活的 TPS 变换，仍能展现出良好的泛化能力。我们还可以观察到，矩形化结果呈现出合理的全局分布和视觉上令人愉悦的局部细节。此外，预测网格位于校正图像的最大空间范围内，证明了所学习的矩形化变换是有效的。

A.7. 用户研究

生成矩形图像的主要目的是满足视觉感受和视觉感知需求，因此我们开展了一项用户研究，从三个方面评估矩形化方法：内容保真度、结构保持和感知性能。具体而言，内容保真度指对图像内容（如物体和背景）的衡量。结构保持需要进行评估，尤其是矩形化图像边界，即边界是否笔直。感知性能指直观的目标检测和语义分割结果。随后，我们按随机顺序收集了 200 个样本，并邀请 10 位具备视觉专业知识的志愿者，在不同情境下对结果进行投票。如图 16所示，RecRecNet 在三项测试中均获得最高票数，展现出卓越的场景忠实度。