Computer Science and Application 计算机科学与应用, 2026, 16(2), 314-327
Published Online February 2026 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2026.162061

Hans汉斯

高效并转融全景玻璃精分网络

黄智鸿[1]，李科霖[1]，黄舒荧[2]，余昊辉[1]，常青玲[1*]，崔　岩[1,3]

[1]五邑大学中德人工智能研究院，广东 江门
[2]华南农业大学继续教育学院，广东 广州
[3]珠海市四维时代网络科技有限公司，广东 珠海

收稿日期：2026年1月5日；录用日期：2026年2月5日；发布日期：2026年2月13日

摘　要

精准分割环境中的玻璃物体，是提升自动驾驶、深度感知等视觉系统性能的关键环节。然而当前主流的深度学习分割方法，其训练与推理几乎完全建立在传统的透视图像之上，这类图像的有限视野与局部的上下文信息，使其在处理开阔场景中尺度多变、距离各异的玻璃目标时显得力不从心。虽然全景成像能提供无死角的全局环境感知，但其中玻璃区域因透视产生的剧烈形变，与其自身的透光、反射等固有光学特性相互作用，构成了一个极度复杂的视觉分析难题，远超传统透视图像所面临的挑战。为系统性地解决上述难题，本文提出了一种新的网络架构——高效并转融全景玻璃精分网络。该神经网络架构系统集成了注意力机制、转置卷积、深度可分离卷积和空卷积等先进操作，分别设计了三个模块：高效并行卷融深度可分模块、高效转卷双支融调模块和高效并转累融精调模块，用于对主干网络提取的特征进行再处理。我们在PanoGlass V2等基准数据集上的实验表明，本方法关键指标显著优于现有技术，IoU、MAE与F-Score分别达到91.37%、95.49%与0.0060，验证了其高效性与优越的泛化能力，为复杂场景下的全景视觉应用提供了可靠解决方案。

关键词

玻璃分割，深度学习，神经网络，全景语义分割

Efficient and Integrated Panoramic Glass Precision Sorting Network

Zhihong Huang[1], Kelin Li[1], Shuying Huang[2], Haohui Yu[1], Qingling Chang[1*], Yan Cui[1,3]

[1]Sino-German Artificial Intelligence Research Institute, Wuyi University, Jiangmen Guangdong
[2]College of Continuing Education, South China Agricultural University, Guangzhou Guangdong
[3]Zhuhai 4DAGE Technology Co., Ltd., Zhuhai Guangdong

Received: January 5, 2026; accepted: February 5, 2026; published: February 13, 2026

[*]通讯作者。

文章引用: 黄智鸿, 李科霖, 黄舒荧, 余昊辉, 常青玲, 崔岩. 高效并转融全景玻璃精分网络[J]. 计算机科学与应用, 2026, 16(2): 314-327. DOI: 10.12677/csa.2026.162061

Abstract

Accurately segmenting glass objects in the environment is a crucial step in enhancing the performance of visual systems such as autonomous driving and deep perception. However, the current mainstream deep learning segmentation methods rely almost entirely on traditional perspective images for training and inference. The limited field of view and local contextual information of such images make them inadequate in handling glass targets with varying scales and distances in open scenes. Although panoramic imaging provides a comprehensive and unobstructed view of the environment, the severe deformation of glass areas due to perspective, coupled with their inherent optical properties such as light transmission and reflection, poses an extremely complex visual analysis challenge, far exceeding the challenges faced by traditional perspective images. To systematically address the aforementioned challenges, this paper proposes a novel network architecture—Efficient and Integrated Panoramic Glass Precision Sorting Network. This neural network architecture integrates advanced operations such as attention mechanisms, transposed convolution, depthwise separable convolution, and spatial convolution, and designs three modules: the Efficient Parallel-to-Global Deepwise Separable Module, the Efficient Transposed-to-Global Dual-Stream Fusion Module, and the Efficient Parallel-to-Global Accumulative Fusion Fine-tuning Module, for reprocessing the features extracted by the backbone network. Our experiments on benchmark datasets such as PanoGlass V2 demonstrate that the key metrics of this method significantly outperform existing techniques, achieving IoU, MAE, and F-Score of 91.37%, 95.49%, and 0.0060, respectively. This verifies its efficiency and superior generalization ability, providing a reliable solution for panoramic vision applications in complex scenes.

Keywords

Glass Segmentation, Deep Learning, Neural Network, Panoramic Semantic Segmentation

1. 介绍

从我们每日触手可及的透明水杯，到构建建筑空间的玻璃幕墙与门窗，由透明玻璃构成的物体已深度嵌入人类生活的方方面面。与此同时基于计算机视觉的智能系统，如自动驾驶汽车与环境深度感知设备，正日益广泛地部署于现实世界。然而一个显著的技术挑战随之浮现：这些系统如何可靠地检测并辨识这些视觉上“隐形”的玻璃物体。其核心难点在于视觉传感器必须精确解析玻璃的物理存在，并将其与场景中的其他实体或背景进行有效区分。若视觉系统在此环节发生误判或漏检，可能在关键应用中引发不可预知的风险与安全隐患，凸显了提升透明物体感知能力的紧迫性与重要性。例如，自动驾驶系统[1]可能会撞上商场的玻璃墙，深度估计系统[2]可能会错误地估计玻璃区域的深度。全景图像具有广阔的视野和丰富的信息，基于全景图像的玻璃分割可以为自动驾驶、深度估计等任务提供准确的数据，从而提高这些任务的准确性和效率。基于全景图像的玻璃分割对计算机视觉的发展具有重要意义，因为它有助于推动相关技术的进步和创新。

在视觉感知系统中，实现对玻璃材质物体的精准辨识与轮廓分割，是一项公认的艰巨任务。这一挑战在全景图像中尤为突出，其中玻璃扭曲等问题使得现有方法难以应对。玻璃卓越的透光能力使得玻璃的外

观表征高度不稳定，随着观测角度与周遭环境的改变而持续波动。而光线在玻璃表面发生的反射会扭曲甚至复制背后的景物信息，导致其视觉特征极度混乱与不可靠，这为鲁棒的特征提取设置了巨大障碍。近年来，针对玻璃识别和分割的问题，研究者们提出了多种解决方案。其中，一些方法通过识别反射和玻璃边界来尝试解决玻璃分割的问题。例如，Translab [3]和 Tran2seg [4]等研究为此做出了贡献。此外，RGB-T [5]和 RGB-P [6]等方法则利用 RGB 图像、热图、深度图等辅助图像来进行玻璃分割。然而，这些方法在处理全景图像时面临着玻璃扭曲等复杂问题的挑战，全景图像中的这些变化往往导致分割精度下降。

为了解决这个问题，我们设计了一种新的网络架构——高效并转融全景玻璃精分网络。结合注意力机制、转置卷积、深度可分离卷积和空卷积等多种特征提取方法，设计了三个特征提取模块：高效并行卷融深度可分模块、高效转卷双支融调模块和高效并转累融精调模块，分别用于解决玻璃透明性、玻璃反射性和全景图像玻璃扭曲这三个问题。试验结果验证了这种方法的有效性。

综上所述，本文的主要工作如下：

- 首先，我们提出了一种新的网络模型架构——高效并转融全景玻璃精分网络，它使用单个 RGB 全景图像作为输入，可以用较少的参数实现高精度的全景图像玻璃分割。
- 其次，我们设计了三个新的模块：高效并行卷融深度可分模块、高效转卷双支融调模块和高效并转累融精调模块，用于对主干网络提取的特征进行再处理。这三个模块是针对全景图像玻璃分割中的关键问题而设计的，如玻璃透明性、玻璃反射性和全景图像玻璃扭曲。
- 最后，我们在我们团队的数据集 PanoGlass V2 [7]上进行了大量实验，证明了高效并转融全景玻璃精分网络在全景玻璃分割中实现了最佳性能。同时以其他数据集为实验对象，对典型词袋特征压缩算法的性能进行比较性研究报道，证明高效并转融全景玻璃精分网络具有同等的竞争力。

本文其余部分的结构如下。在第二节中，我们讨论了与全景图像玻璃分割任务相关的工作。在第三节中，我们详细介绍了我们的网络模型高效并转融全景玻璃精分网络。在第四节中，我们展示并分析了高效并转融全景玻璃精分网络在 PanoGlass V2 [7]数据集和其他数据集上的实验结果。在第五节中，我们总结了我们的工作，并讨论了未来的工作。

2. 相关工作

2.1. 玻璃分割

全景图像中玻璃的扭曲导致其外观多变，增加了准确分割的难度，具有重要的研究价值。目前的玻璃分割方法包括基于 RGB 图像进行处理。在基于 RGB 图像作为输入的模型中，GDNet [8]是首个基于深度学习的玻璃检测模型，配备了大视场上下文特征集成模块，以探索丰富的上下文线索。此后，TransLab [3]和 Trans2Seg [4]利用边界线索改善玻璃分割。GSDNet [9]通过结合边界和反射线索来区分玻璃和非玻璃。EBLNet [10]设计了边缘感知的基于点的 GCN 模块以增强边界提示。PGSNet [11]进一步将边界和位置信息嵌入不同层的特征图，并提出 DE 和 FEBF 模块以提取更多线索。GlassSegNet [12]受人类检测玻璃行为的启发，先假设玻璃存在再进行纠正。这些 RGB 方法大多依赖反射和边界线索，在特殊场景如黑暗环境中效率较低。为了充分运用 RGB 图像的观察角度和场景复杂度限制，研究人员开始采用结合上下文特征融合和边界监督学习。富上下文聚合模块(RCAM) [9]通过融合周围信息提高分割精度。基于反射率的优化模块(RRM) [9]专注于检测玻璃反射光。TransLab [3]和 Trans2Seg [4]聚焦于提取玻璃边缘特征。此外，基于分段的改进差分模(RDM) [10]和图形卷积网络(GCN) [10]也用于边缘感测。PanoGlassNet [13]是首个基于全景 RGB 和光强度图的玻璃检测模型，创新性地以单骨干网络处理多模态数据，有效利用 RGB 和光强度图，实现对全景图中玻璃的高精度检测。

然而，仅依靠 RGB 图像可能无法充分捕捉和利用玻璃的复杂特性。针对玻璃检测任务中单一模态信息

的局限性，学术界逐步转向多模态融合策略以挖掘互补特征。RGB-T 模型[5]利用玻璃的热辐射特性，通过结合 RGB 和热图像来增强训练效果。RGB-P 模型[6]则聚焦于玻璃材料的特性，如光的偏振效应和强度变化，并通过 RGB-P 图像进行训练，以期获得更准确的检测结果。这些模型通常采用双编码器或多编码器架构，以便更全面地提取特征。然而，这种方法的复杂性在实际应用中导致了一些局限性，模型的参数量显著增加，这不仅加大了计算负担，还可能影响实时处理的效率，这都极大限制了它们在实际应用中的可用性。

鉴于此，我们提出一种面向全景 RGB 图像的创新玻璃分割架构——高效并转融全景玻璃精分网络。该模型摒弃了对特殊成像设备或数据对的依赖，转而聚焦于普通全景图像本身。其核心在于直接利用全景视野中玻璃区域固有的透视畸变与多尺度特性，设计了一种高效的特征融合与提取机制，通过精简的网络结构设计，该方案大幅压缩了参数量与计算复杂度，从而在确保分割精度的前提下，显著提升了模型的实用性与泛化能力。高效并转融全景玻璃精分网络这种轻量化设计不仅降低了部署门槛，也使其能够更好地适应需要快速响应和大规模处理的实际应用场景，为解决全景环境下的玻璃感知问题提供了一条高效路径。

2.2. 全景分割

针对全景图像的分割技术，主要有基于针孔图像和畸变宽视场图像的分割模型。Eder 等人[14]通过划分全景图像为二十面体的局部平面图像，对其进行了深入研究，他们采用的多面体构建法全景表示被广泛应用于计算机视觉领域。Lee 等人[15]则进一步采用球形多面体结构来表示可比较的全向透视，这一方法能够更精确地模拟真实世界的三维场景。Semih Orhan 等人[16]通过引入扭曲传感模块来模拟矩形卷积运算，该模块使得图像的精细结构能够得到更准确的解析。Sun 等人[17]通过高度压缩模块进行底层特征的表示学习，预测密集特征的离散分布，为后续的图像分割提供了坚实的基础。Zheng 等人[18]则致力于融合互补的水平特征和垂直特征，综合利用图像的多尺度信息，提高全景图像分割的精度和鲁棒性。Shen 等人[19]则对传统模块进行了革新，引入了全新的全景 Transformer 模块。Xiong 等人[20]利用注意力机制来指导网络的学习过程，从而实现更精确的全景分割。

3. 方法

3.1. 整体架构

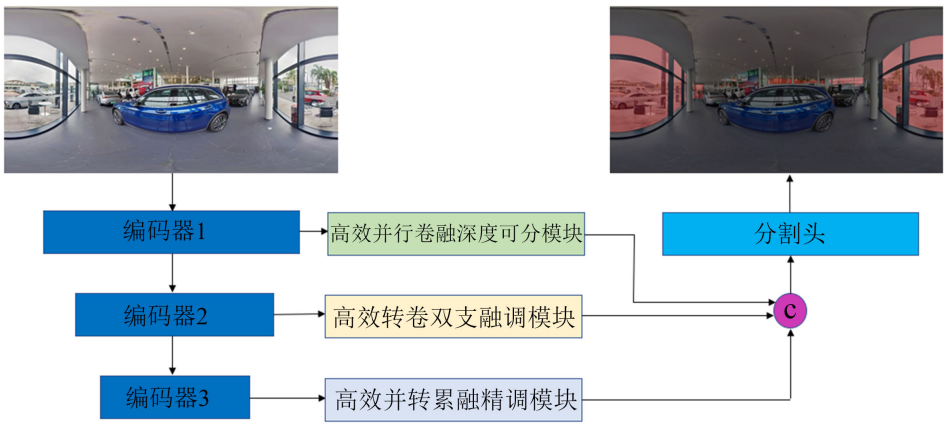

Figure 1. Overall architecture diagram of efficient and integrated panoramic glass precision sorting network
图 1. 高效并转融全景玻璃精分网络的整体架构图

我们的网络架构如图 1 所示。该网络的主要结构包括骨干网和我们提出的三种结构，高效并行卷融深度可分模块、高效转卷双支融调模块和高效并转累融精调模块。在主干网络结构中，我们使用 poolformer

主干进行特征提取。这是我们通过大量的实验发现的，该结构可以用较少的参数实现高性能。

3.2. 高效并行卷融深度可分模块

高效并行卷融深度可分模块的结果如图 2 所示。由于玻璃允许光线穿透，其外观特征高度依赖于背景环境，导致玻璃区域在特征空间中表现出极高的方差。传统的单路径卷积结构在处理此类特征时，容易忽略微弱的玻璃边缘线索，或将其误认为背景物体。为了增强网络对透明介质微弱特征的提取能力，我们设计了高效并行卷积深度可分模块，旨在通过多尺度并行感知与深度可分卷积的结合，实现对透明区域特征的精炼与下采样过程中的信息补偿。高效并行卷融深度可分模块模型输入特征图首先进入四个并行的 3 × 3 卷积分支，四个分支的输出在通道维度进行拼接，随后经过统一的批归一化处理。为了在提升性能的同时控制计算开销，模块核心采用了 3 × 3 的深度可分卷积。最后通过一个 1 × 1 卷积层进行通道整合，输出优化后的特征图，输出的优化特征图能有效区分透明玻璃与纯空旷区域，显著提升了在复杂背景下透明物体的分割精度。

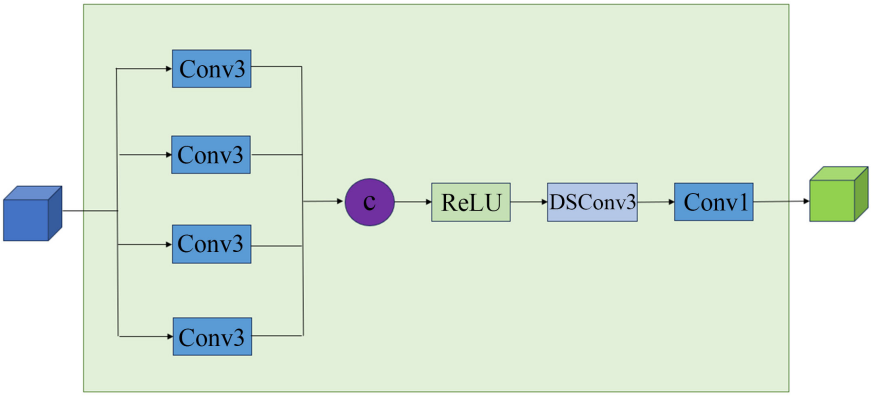

Figure 2. Efficient parallel volume fusion depth-separable module structure diagram
图 2. 高效并行卷融深度可分模块结构图

3.3. 高效转卷双支融调模块

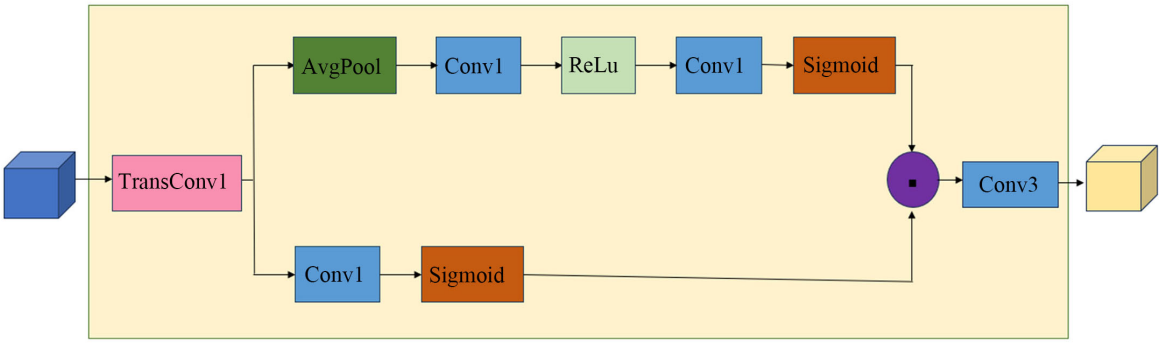

Figure 3. Structural diagram of efficient dual-branch fusion and adjustment module for roll conversion
图 3. 高效转卷双支融调模块结构图

高效转卷双支融调模块的结果如图 3 所示。为了有效分离反射干扰并增强真实玻璃区域的特征，我们提出了高效转置卷积双支融合调节模块，利用双重注意力引导机制实现特征的动态增强。高效转置卷积双支融合调节模块首先采用 1 × 1 转置卷积对输入特征进行初步的空间维度调整，为后续的注意力计算

提供更丰富的空间上下文。特征分上下支路进行注意力引导，上支路经过平均池化压缩空间维度，随后通过连续的 1 × 1 卷积与 ReLU 激活，最后经由 Sigmoid 函数生成通道注意力权重；下支路的特征直接经过 1 × 1 卷积与 Sigmoid 激活。最后两个支路的输出进行逐元素相乘，实现双重注意力引导。融合后的特征最后通过一个 3 × 3 卷积层进行平滑处理，输出最终的增强特征。该模块的这种设计使得模型在面对强反射干扰时，依然能保持极高的分割鲁棒性。

3.4. 高效并转累融精调模块

高效并转累融精调模块的结果如图 4 所示。传统的正方形卷积核在处理全景图像等具有几何畸变的特征时，其规则的采样网格与物体在畸变空间中的实际形状无法对齐，导致特征提取效率低下。为解决这一问题，我们设计了高效并转累加融合精调模块。该模块的核心在于利用转置卷积的可学习重采样特性，建立从畸变特征空间到校正特征空间的映射关系。转置卷积通过其分数步长和可学习的核参数，能够对输入特征图执行一种柔性的、非均匀的空间重分布操作。这种操作在数学上可以近似模拟或学习全景畸变投影模型的局部逆变换。通过部署多个并行的转置卷积路径，模块能够从不同几何假设出发，对扭曲特征进行多视角的适应性采样与变换。随后将这些并行路径输出的、经过初步几何校正的特征图进行累加融合，从而集成不同变换视角下的有效信息，最终在特征层面实现对全景畸变的精确补偿与几何结构的精调，为后续处理提供对齐更佳的特征表示。输入特征图分别经过一个不经过任何卷积处理，直接作为跳跃连接参与后续融合的分支和三个并行的转置卷积分支，四个分支的输出并非简单的拼接，而是通过三个串联的加法节点进行逐级累加。累加后的特征经过通道拼接和 ReLU 激活，最后通过一个 1 × 1 卷积层进行最终的特征映射。该模块的这种设计确保了网络在处理大幅度扭曲的玻璃区域时，依然能够维持特征的连贯性。

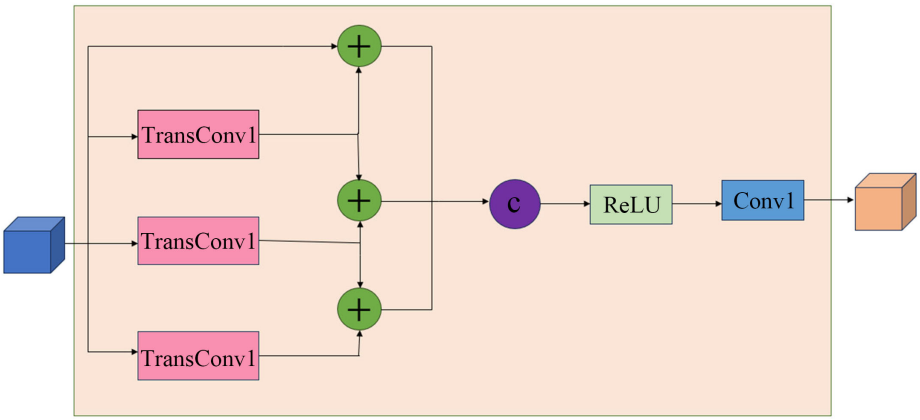

Figure 4. Efficient parallel-to-serial and fine-tuning module structure diagram
图 4. 高效并转累融精调模块结构图

4. 实验

4.1. 实验详细信息

我们在 mmsegmentation 上训练和评估了高效并转融全景玻璃精分网络，mmsegmentation 是一个集成的语义分割框架，提供了在 ImageNet 上预先训练的多个主干，如 ResNet，MobileNet，DenseNet，VisionTransformer，VGG 等。具体来说，我们使用了以下设置：环境是 mmcv 2.1.0，mmsegmentation 1.2.2，PyTorch 2.1.2，Python 3.8.19，CUDA 12.2。对于训练改进，我们进行了缩放(scale = (2048, 512)，scale range

= (0.5, 2.0))，随机裁剪(size = (512, 512))，随机翻转(prob = 0.5)，以及图像测量失真和填充(size = (512, 512))。在测试阶段，我们执行了随机翻转和翻转。在优化器方面，我们使用 AdamW，学习率为 0.00006，权重衰减率为 0.01，批量大小为 4，学习速率调度使用 1.0 的多衰减功率、线性预热、5000 的预热步数和 1e−6 的预热比。我们使用 CrossEntropyLoss 作为损失函数。我们在 PanoGlass V2 [7]上进行了训练，这是一个由全景图像组成的玻璃分割数据集，共有 1983 张图像用于训练，416 张图像用于测试。并且数据集根据室内和室外场景进行区分，可以分别进行测试。训练持续 320,000 次迭代，在 NVIDIA GeForce RTX 3060 Ti 和 64 GB RAM 上花费约 34.5 小时。

4.2. 评估指标

我们使用了玻璃分割 RGB-T [5]和 Trans 10 K-v2 [4]中使用的指标来评价我们的模型和我们的数据集 PanoGlass V2 [7]上选择的方法。我们使用的主要指标是 IoU，它也广泛用于语义分割和玻璃分割。IoU 的定义为：

$$\mathrm{IoU}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}} \tag{1}$$

其中，TP、FP、FN 分别为真阳性、假阳性和假阴性像素的数量。除了 IoU，我们还使用平均绝对误差(MAE)度量，其定义为：

$$\mathrm{MAE}=\frac{1}{\mathrm{H}\times\mathrm{W}}\sum_{i=1}^{\mathrm{H}}\sum_{j=1}^{\mathrm{W}}\left|P(i,j)-G(i,j)\right| \tag{2}$$

其中 $P(i,j)$ 是模型预测的位置(i,j)处的目标的概率，$G(i,j)$ 是位置(i,j)处的真实标签，H 和 W 是图像的高度和宽度。此外，我们使用 F 分数作为另一个度量标准，这是平均准确率和平均召回率的调和平均值。F 分数定义为：

$$\mathrm{Fscore}=\frac{\left(1+\beta^2\right)\mathrm{Precision}\times\mathrm{Recall}}{\beta^2\times\mathrm{Precision}\times\mathrm{Recall}} \tag{3}$$

当 β 设置为 1 时，Precision 和 recall 分别定义为 $\mathrm{Precision}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}$ 和 $\mathrm{recall}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$。

4.3. 定量评价

为了保证比较的公平性，除 RGB-T [5]、Trans2Seg [4]、translab [3]外，在训练过程中，我们还采用 mmsegmentation tool 对所有模型进行了预评估，其提供了所有评估指标。接下来，我们将高效并转融全景玻璃精分网络与 16 种最先进的语义分割方法和 4 种玻璃分割方法在 PanoGlass V2 [7]上进行比较。语义分割方法包括：ResNeSt [21]，CCNet [22]，Segformer [23]，Fpn [24]，Mae [25]，Poolformer [26]，DeepLabv3+ [27]，Pointrend [28]，Vit [29]，Swin [30]，STDC [31]，Twins [32]，ConvNext [33]，SegNext [34]。玻璃分割方法有：RGB-T [5]、translab [4]、Trans 2Seg [3]、Panoglassnet [13]。实验结果示于表 1 中。我们使用 mmsegmentation 实现和测试了各种经典的语义分割模型，并使用公开的代码版本实现和测试了所有的玻璃检测方法。

结果表明，我们的模型高效并转融全景玻璃精分网络在 PanoGlass V2 [7]上的 IoU 达到 91.37%，MAE 达到 0.0060，Fscore 达到 95.49%，在这三个指标中均达到最优值，均优于其他模型。通过对实验结果的仔细分析，很明显，我们的模型相比所有其他模型实现了更好的性能。虽然一些模型的精度接近我们的水平，但它们依赖于大量的参数，这给实际应用带来了相当大的压力。相比之下，我们的模型以更少的

参数提供了卓越的性能。相反，尽管一些方法使用的参数比我们的模型少，但它们的精度低于我们的结果。为了进一步确定模型的性能，我们在 PanoGlass V2 [7]的室内和室外数据集上进行了测试，结果如表 2 和表 3 所示。

Table 1. Quantitative comparison of models on the PanoGlass V2 dataset [7]
表 1. 数据集 PanoGlass V2 [7]上模型的定量比较

	Methods	Backbone	IoU↑	Fscore↑	MAE↓	Param (M)	Flops (G)
Glass	TransLab [3]	ResNet50 [35]	81.81	84.75	0.0083	40.147	61.6
	Trans2Seg [4]	ResNet50 [35]	85.22	93.47	0.0067	85.01	83.59
	RGB-T [5]	ResNet50 [35]	87.27	94.25	0.0071	327.72	222.29
	Panoglassnet [13]	MSCAN [31]	*89.21*	*94.81*	*0.0063*	427.5	581
Semantic	Vit [29]	Vit [29]	57.68	73.58	0.0264	144.06	393.84
	BiSeNetv2 [27]	BiSeNetv2 [27]	62.45	75.89	0.2341	13.23	*11.02*
	Mae [25]	Vit [29]	65.86	81.76	0.0260	604	162
	STDC [31]	STDCNet [31]	75.78	86.64	0.0099	*12.6*	11.78
	ResNeSt [21]	ResNet [21]	85.21	92.89	0.0096	69.9	263.64
	Fpn [24]	ResNet101 [35]	85.33	91.02	0.0099	49.7	51
	Pointrend [28]	ResNet101 [35]	85.54	91.29	0.0101	44.5	15.4
	Segformer [23]	MIT [23]	85.66	91.49	0.0098	**3.72**	**7.89**
	Poolformer [26]	Poolformer [26]	86.08	92.42	0.0086	15.65	30.74
	Twins [32]	PCPVT [32]	87.11	93.3	0.0068	132.67	282.37
	CCNet [22]	BiSeNetv1 [22]	87.8	93.51	0.0076	47.59	201
	SegNext [34]	MSCAN [34]	88.22	93.81	0.0074	27.56	32.48
	Swin [30]	Swin [30]	88.75	93.93	0.0070	233.85	409.53
	ConvNext [33]	ConvNext [33]	89.12	94.31	0.0066	80.95	257
	Ours	Poolformer [26]	**91.37**	**95.49**	**0.0060**	77.06	82.008

注：所有方法都使用 MMSEGMENTATION 进行了训练和验证。第 2 行至 5 行是玻璃分割方法的结果，第 6 行至 19 行是经典语义分割方法的结果。我们保留每个指标的最佳值，并加粗显示每个指标的最佳值，用斜体表示每个指标的次优值。

Table 2. Performance of different methods on indoor panoramic image dataset
表 2. 不同方法在室内全景图像数据集上的性能

	Methods	IoU↑	Fscore↑	MAE↓
Glass	TransLab [3]	82.67	85.38	0.0095
	RGB-T [5]	86.32	95.02	*0.0057*
	Trans2Seg [4]	88.47	95.55	0.0061
	Panoglassnet [13]	*90.02*	*95.77*	0.0061
	Vit [29]	54.77	70.17	0.0302

续表

Semantic	BiSeNetv2 [27]	69.7	82.91	0.0205
	Mae [25]	69.87	84.2	0.0192
	STDC [31]	70.49	85.38	0.0171
	ResNeSt [21]	74.62	88.51	0.0141
	Fpn [24]	83.55	90.13	0.0091
	Pointrend [28]	83.71	90.39	0.0091
	Segformer [23]	87.02	90.55	0.0081
	Twins [32]	87.24	93.18	0.0073
	CCNet [22]	87.5	92.99	0.0077
	SegNext [34]	88.38	93.83	0.0071
	Swin [30]	89.27	94.27	0.0063
	ConvNext [33]	89.42	94.58	0.0063
	Poolformer [26]	89.69	95.01	0.0058
	Ours	**91.21**	**95.91**	**0.0045**

注：对于不同方法在室内全景图像数据集上的性能，我们用粗体字表示每个指标的最优值，用斜体字表示每个指标的次优值。

Table 3. Performance of different methods on outdoor panoramic image dataset
表 3. 不同方法在室外全景图像数据集上的性能

	Methods	**IoU↑**	**Fscore↑**	**MAE↓**
Glass	TransLab [3]	71.65	82.31	0.0005
	RGB-T [5]	72.79	81.42	0.0005
	Trans2Seg [4]	74.97	83.41	0.0004
	Panoglassnet [13]	76.21	85.49	0.0004
Semantic	Vit [29]	28.52	43.92	0.0013
	BiSeNetv2 [27]	33.57	50.25	0.0011
	Mae [25]	34.42	58.55	0.0011
	STDC [31]	35.59	56.19	0.0011
	ResNeSt [21]	36.11	53.09	0.0011
	Fpn [24]	52.74	65.37	0.0009
	Pointrend [28]	53.85	67.27	0.0009
	Segformer [23]	72.91	81.25	0.0004
	Twins [32]	73.75	85.47	0.0004
	CCNet [22]	73.79	87.61	0.0004
	SegNext [34]	74.07	87.08	0.0004
	Swin [30]	74.77	86.39	0.0004
	ConvNext [33]	75.66	86.41	0.0004
	Poolformer [26]	*77.69*	*87.47*	*0.0004*
	Ours	**83.51**	**90.27**	**0.0003**

注：对于不同方法在室外全景图像数据集上的性能，我们用粗体表示每个指标的最优值，用斜体表示每个指标的次优值。

我们的模型在室内数据集上 IoU 表现为 91.21，MAE 为 0.0045 和 Fscore 为 95.91，在室外数据集上表现为 83.51 的 IoU、0.0003 的 MAE 和 90.27 的 Fscore。该模型的实验结果也是最优的。

我们评估了高效并转融全景玻璃精分网络在不同玻璃数据集上的性能。我们选择了目前流行的公开可用的玻璃检测数据集，包括 GDD [8]、HSO [11]、RGB-T [5]。具体而言，GDD [8]是第一个已知的大规模人工标注数据集，其包含 2827 个室内图像和 1089 个室内图像，并且大多数现有的玻璃分割模型都基于该数据集。HSO [11]包含来自 MatterPort3D [36]、2D3DS [37]、ScanNet [38]、Sunrgbd [39]过滤图像的 9704 个片段。RGB-T [5]是一个基于玻璃物理特性的数据集，由 5518 个 RGB 图像和热图组成。我们收集了这些方法在公开数据集上的测试结果，如表 4~6 所示，并将我们的模型与它们进行了比较。

Table 4. Performance comparison on the open glass dataset RGB-T [5]
表 4. 在开放玻璃数据集 RGB-T [5]上的性能比较

Methods	IoU↑	Fscore↑	MAE↓
EBLNet [10]	80.22	88.31	0.113
RGB-T (only RGB) [5]	88.78	92.75	0.057
RGB-T [5]	*92.97*	*95.32*	*0.028*
Ours	**94.88**	**96.99**	**0.021**

注：对于不同模型在开放玻璃数据集 RGB-T [5]上的性能，我们的模型显示了最佳性能，我们用粗体字表示每个指标的最优值，用斜体字表示每个指标的次优值。

Table 5. Performance comparison on the open glass dataset HSO [11]
表 5. 在开放玻璃数据集 HSO [11]上的性能比较

Methods	IoU↑	Fscore↑	MAE↓
GDNet [8]	78.25	81.42	0.098
EBLNet [10]	79.45	——	0.093
PGSNet [11]	80.45	*83.61*	0.089
GlassSegNet [12]	*84.77*	——	*0.086*
Ours	**86.75**	**91.13**	**0.055**

注：对于不同模型在开放玻璃数据集 HSO [11]上的性能，我们的模型显示了最佳性能，我们用粗体字表示每个指标的最优值，用斜体字表示每个指标的次优值。

Table 6. Performance comparison on the open glass dataset GDD [8]
表 6. 在开放玻璃数据集 GDD [8]上的性能比较

Methods	IoU↑	Fscore↑	MAE↓
RGB-T (only RGB) [5]	86.77	——	0.070
GDNet [8]	87.42	89.23	0.067
PGSNet [11]	87.99	90.18	0.062
EBLNet [10]	88.22	*93.24*	*0.058*
GlassSegNet [12]	*90.53*	——	0.063
Ours	**91.95**	**94.57**	**0.051**

注：对于不同模型在开放玻璃数据集 GDD [8]上的性能，我们的模型显示了最佳性能，我们用粗体字表示每个指标的最优值，用斜体字表示每个指标的次优值。

在 RGB-T 数据集上，我们的模型实现了 94.88%的 IoU、0.021 的 MAE 和 96.99%的 Fscore；在 HSO 数据集上，我们的模型实现了 86.75%的 IoU、0.055 的 MAE 和 91.13%的 Fscore。在 GDD 数据集上，我们的模型实现了 91.95 的 IoU、0.051 的 MAE 和 94.57 的 Fscore。

实验结果表明我们所建立的模型在所有的玻璃分割数据集上都有较好的表现，优于我们所选择的其他模型。同时，我们的模型参数数量也具有竞争力，这表明高效并转融全景玻璃精分网络在玻璃分割任务中具有显著的性能优势，为玻璃分割领域的进一步研究和应用提供了有力的支持。

4.4. 消融实验

Table 7. Ablation experiment
表 7. 消融实验

Methods	IoU	Fscore	MAE
Base	86.08	92.42	0.0086
Base + 高效并行卷融深度可分模块	90.56	94.76	0.0071
Base + 高效转卷双支融调模块	90.94	95.12	0.0065
Base + 高效并转累融精调模块	91.09	95.76	0.0063
高效并转融全景玻璃精分网络	**91.37**	**95.49**	**0.0060**

注：消融实验分别验证了高效并行卷融深度可分模块、高效转卷双支融调模块和高效并转累融精调模块的性能改善。

在本节中，我们在数据集 PanoGlass V2 [7]上进行消融实验，以证明我们设计的模块(即高效并行卷融深度可分模块、高效转卷双支融调模块和高效并转累融精调模块)在模型中的作用，实验结果示于表 7 中。我们尝试在所有四个阶段中使用高效并行卷融深度可分模块、高效转卷双支融调模块或高效并转累融精调模块，但实验结果并不令人满意。高效并行卷融深度可分模块仅能增强透明区域与背景的特征区分度，但无法抑制反射引入的虚假纹理特征，易导致分割结果混入反射干扰；高效转卷双支融调模块可弱化反射噪声，却缺乏对全景图像几何畸变的校正能力，难以适配扭曲区域的特征对齐需求；高效并转累融精调模块虽能校正投影畸变，却缺乏对光学物理特性的显式建模能力。只有当这三个模块一起使用时，它们才能发挥最大效用，单一模块的孤立使用难以覆盖全景图像玻璃分割的全流程需求。

5. 结论

本文的主要贡献如下：(1) 提出了一种新的用于全景图像玻璃分割的网络结构。(2) 本文提出了三种新的特征处理结构高效并行卷融深度可分模块、高效转卷双支融调模块和高效并转累融精调模块，分别利用了通道注意、空间注意、深度可分离卷积、转置卷积和注意力机制等技术。(3) 在此基础上进行了大量的实验，验证了本文提出的网络结构和模块的优越性。当然，这项研究也有一些明显的局限性。虽然我们的特征处理模块使用了多种不同的处理方法来提高模型的整体性能，但这也导致了模型参数和计算量的增加。为了达到更好的性能，我们还需要进一步优化这些模块，减少参数的数量和计算量。其次，实验数据集中透明玻璃类型的样本分布不均衡，可能影响模型在特殊场景下的泛化能力，一个高质量的数据集应该包含很多图像，以提高模型的鲁棒性。展望未来，我们将着力突破上述瓶颈，致力于开发更为轻量、高效且精准的新一代玻璃感知模型。其潜在应用将拓展至智能建筑巡检、增强现实导航等智慧城市核心场景，从而提供更可靠的视觉技术支持。这项研究不仅推动了计算机视觉在透明物体理解方面的发展，也为探索多模态感知融合提供了有益的参考路径。

基金项目

本工作由中国国家重点研发计划(项目编号：2022YFA1602003)资助，该计划由中国科学院高能物理研究所主持。

参考文献

[1] Fu, X., Zhang, S., Chen, T., Lu, Y., Zhou, X., Geiger, A. and Liao, Y. (2023) Panopticnerf-360: Panoramic 3D-to-2D Label Transfer in Urban Scenes.

[2] Cassar, D.R. (2023) Glassnet: A Multitask Deep Neural Network for Predicting Many Glass Properties. *Ceramics International*, **49**, 36013-36024. https://doi.org/10.1016/j.ceramint.2023.08.281

[3] Xie, E., Wang, W., Wang, W., Ding, M., Shen, C. and Luo, P. (2020) Segmenting Transparent Objects in the Wild. In: Vedaldi, A., *et al*., Eds., *Computer Vision—ECCV* 2020, Springer International Publishing, 696-711. https://doi.org/10.1007/978-3-030-58601-0_41

[4] Xie, E., Wang, W., Wang, W., Sun, P., Xu, H., Liang, D., *et al*. (2021) Segmenting Transparent Objects in the Wild with Transformer. *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, Montreal, 19-27 August 2021, 1194-1200. https://doi.org/10.24963/ijcai.2021/165

[5] Huo, D., Wang, J., Qian, Y. and Yang, Y. (2023) Glass Segmentation with RGB-Thermal Image Pairs. *IEEE Transactions on Image Processing*, **32**, 1911-1926. https://doi.org/10.1109/tip.2023.3256762

[6] Mei, H., Dong, B., Dong, W., Yang, J., Baek, S., Heide, F., *et al*. (2022) Glass Segmentation Using Intensity and Spectral Polarization Cues. 2022 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), New Orleans, 18-24 June 2022, 12622-12631. https://doi.org/10.1109/cvpr52688.2022.01229

[7] Chang, Q., Meng, X., Hong, Z. and Cui, Y. (2024) ProgressiveGlassNet:Glass Detection with Progressive Decoder. 2024 *IEEE International Symposium on Parallel and Distributed Processing with Applications* (*ISPA*), Kaifeng, 30 October-2 November 2024, 917-925. https://doi.org/10.1109/ispa63168.2024.00122

[8] Mei, H., Yang, X., Wang, Y., Liu, Y., He, S., Zhang, Q., *et al*. (2020) Don't Hit Me! Glass Detection in Real-World Scenes. 2020 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 13-19 June 2020, 3687-3696. https://doi.org/10.1109/cvpr42600.2020.00374

[9] Lin, J., He, Z. and Lau, R.W. (2021) Rich Context Aggregation with Reflection Prior for Glass Surface Detection. *Proceedings of the IEEE*/*CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 19-25 June 2021, 13415-13424.

[10] He, H., Li, X., Cheng, G., Shi, J., Tong, Y., Meng, G., Prinet, V. and Weng, L. (2021) Enhanced Boundary Learning for Glass-Like Object Segmentation. *Proceedings of the IEEE*/*CVF International Conference on Computer Vision*, Montreal, 10-17 October 2021, 15859-15868.

[11] Yu, L., Mei, H., Dong, W., Wei, Z., Zhu, L., Wang, Y., *et al*. (2022) Progressive Glass Segmentation. *IEEE Transactions on Image Processing*, **31**, 2920-2933. https://doi.org/10.1109/tip.2022.3162709

[12] Zheng, C., Li, P., Zhang, X., Lu, X. and Wei, M. (2023) Don't Worry about Mistakes! Glass Segmentation Network via Mistake Correction. https://api.semanticscholar.org/CorpusID:258291912

[13] Chang, Q., Liao, H., Meng, X., Xu, S. and Cui, Y. (2024) Panoglassnet: Glass Detection with Panoramic RGB and Intensity Images. *IEEE Transactions on Instrumentation and Measurement*, **73**, 1-15. https://doi.org/10.1109/tim.2024.3390163

[14] Eder, M., Shvets, M., Lim, J. and Frahm, J. (2020) Tangent Images for Mitigating Spherical Distortion. 2020 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 13-19 June 2020, 12426-12434. https://doi.org/10.1109/cvpr42600.2020.01244

[15] Lee, Y., Jeong, J., Yun, J., Cho, W. and Yoon, K. (2019) SpherePHD: Applying CNNs on a Spherical PolyHeDron Representation of 360˚ Images. 2019 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Long Beach, 15-20 June 2019, 9181-9189. https://doi.org/10.1109/cvpr.2019.00940

[16] Orhan, S. and Bastanlar, Y. (2021) Semantic Segmentation of Outdoor Panoramic Images. *Signal, Image and Video Processing*, **16**, 643-650. https://doi.org/10.1007/s11760-021-02003-3

[17] Sun, C., Sun, M. and Chen, H. (2021) HoHoNet: 360 Indoor Holistic Understanding with Latent Horizontal Features. 2021 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Nashville, 20-25 June 2021, 2573-2582. https://doi.org/10.1109/cvpr46437.2021.00260

[18] Zheng, Z., Lin, C., Nie, L., Liao, K., Shen, Z. and Zhao, Y. (2023) Complementary Bi-Directional Feature Compression for Indoor 360˚ Semantic Segmentation with Self-Distillation. 2023 *IEEE*/*CVF Winter Conference on Applications of*

Computer Vision (*WACV*), Waikoloa, 2-7 January 2023, 4501-4510. https://doi.org/10.1109/wacv56688.2023.00448

[19] Shen, Z., Lin, C., Liao, K., Nie, L., Zheng, Z. and Zhao, Y. (2022) PanoFormer: Panorama Transformer for Indoor 360˚ Depth Estimation. In: Avidan, S., *et al*., Eds., *Computer Vision—ECCV* 2022, Springer, 195-211. https://doi.org/10.1007/978-3-031-19769-7_12

[20] Xiong, Y., Liao, R., Zhao, H., Hu, R., Bai, M., Yumer, E., *et al*. (2019) UPSNet: A Unified Panoptic Segmentation Network. 2019 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Long Beach, 15-20 June 2019, 8810-8818. https://doi.org/10.1109/cvpr.2019.00902

[21] Zhang, H., Wu, C., Zhang, Z., Zhu, Y., Lin, H., Zhang, Z., *et al*. (2022) ResNeSt: Split-Attention Networks. 2022 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition Workshops* (*CVPRW*), New Orleans, 19-20 June 2022, 2736-2746. https://doi.org/10.1109/cvprw56347.2022.00309

[22] Yu, C., Wang, J., Peng, C., Gao, C., Yu, G. and Sang, N. (2018) BiSeNet: Bilateral Segmentation Network for Real-Time Semantic Segmentation. In: Ferrari, V., *et al*., Eds., *Computer Vision—ECCV* 2018, Springer International Publishing, 334-349. https://doi.org/10.1007/978-3-030-01261-8_20

[23] Xie, E., Wang, W., Yu, Z., *et al*. (2021) SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers.

[24] Kirillov, A., Girshick, R., He, K. and Dollár, P. (2019) Panoptic Feature Pyramid Networks. 2019 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Long Beach, 15-20 June 2019, 6392-6401. https://doi.org/10.1109/cvpr.2019.00656

[25] He, K., Chen, X., Xie, S., Li, Y., Dollar, P. and Girshick, R. (2022) Masked Autoencoders Are Scalable Vision Learners. 2022 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), New Orleans, 18-24 June 2022, 15979-15988. https://doi.org/10.1109/cvpr52688.2022.01553

[26] Yu, W., Luo, M., Zhou, P., Si, C., Zhou, Y., Wang, X., *et al*. (2022) MetaFormer Is Actually What You Need for Vision. 2022 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), New Orleans, 18-24 June 2022, 10809-10819. https://doi.org/10.1109/cvpr52688.2022.01055

[27] Yu, C., Gao, C., Wang, J., Yu, G., Shen, C. and Sang, N. (2021) Bisenet V2: Bilateral Network with Guided Aggregation for Real-Time Semantic Segmentation. *International Journal of Computer Vision*, **129**, 3051-3068. https://doi.org/10.1007/s11263-021-01515-2

[28] Kirillov, A., Wu, Y., He, K. and Girshick, R. (2020) PointRend: Image Segmentation as Rendering. 2020 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 13-19 June 2020, 9796-9805. https://doi.org/10.1109/cvpr42600.2020.00982

[29] Dosovitskiy, A., Beyer, L., Kolesnikov, A., *et al*. (2020) An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale.

[30] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., *et al*. (2021) Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. 2021 *IEEE*/*CVF International Conference on Computer Vision* (*ICCV*), Montreal, 10-17 October 2021, 10012-10022. https://doi.org/10.1109/iccv48922.2021.00986

[31] Fan, M., Lai, S., Huang, J., Wei, X., Chai, Z., Luo, J., *et al*. (2021) Rethinking BiSeNet for Real-Time Semantic Segmentation. 2021 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Nashville, 20-25 June 2021, 9716-9725. https://doi.org/10.1109/cvpr46437.2021.00959

[32] Chu, X., Tian, Z., Wang, Y., Zhang, B., Ren, H., Wei, X., Xia, H. and Shen, C. (2021) Twins: Revisiting the Design of Spatial Attention in Vision Transformers. *Advances in Neural Information Processing Systems*, Vol. 34, 9355-9366.

[33] Liu, Z., Mao, H., Wu, C., Feichtenhofer, C., Darrell, T. and Xie, S. (2022) A Convnet for the 2020s. 2022 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (CVPR), New Orleans, 18-24 June 2022, 11976-11986. https://doi.org/10.1109/cvpr52688.2022.01167

[34] Guo, M.-H., Lu, C.-Z., Hou, Q., Liu, Z., Cheng, M.-M. and Hu, S.-M. (2022) Segnext: Rethinking Convolutional Attention Design for Semantic Segmentation. *Advances in Neural Information Processing Systems*, Vol. 35, 1140-1156.

[35] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 770-778. https://doi.org/10.1109/cvpr.2016.90

[36] Chang, A., Dai, A., Funkhouser, T., Halber, M., Niebner, M., Savva, M., *et al*. (2017) Matterport3D: Learning from RGB-D Data in Indoor Environments. 2017 *International Conference on 3D Vision* (3*DV*), Qingdao, 10-12 October 2017, 667-676. https://doi.org/10.1109/3dv.2017.00081

[37] Armeni, Sax, S., Zamir, A.R. and Savarese, S. (2017) Joint 2D-3D-Semantic Data for Indoor Scene Understanding.

[38] Dai, A., Chang, A.X., Savva, M., Halber, M., Funkhouser, T. and Niessner, M. (2017) ScanNet: Richly-Annotated 3D Reconstructions of Indoor Scenes. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*),

Honolulu, 21-26 July 2017, 5828-5839. https://doi.org/10.1109/cvpr.2017.261

[39] Song, S., Lichtenberg, S.P. and Xiao, J. (2015) SUN RGB-D: A RGB-D Scene Understanding Benchmark Suite. 2015 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Boston, 7-12 June 2015, 567-576. https://doi.org/10.1109/cvpr.2015.7298655