基于视觉失真的玻璃表面检测方法

谭鑫, 齐福霖, 王楠, 张志忠*, 谢源, 马利庄

(华东师范大学计算机科学与技术学院　上海　200062)
(zzzhang@cs.ecnu.edu.cn)

摘　要: 针对玻璃性质造成的玻璃表面检测困难问题, 提出一种基于视觉失真线索的玻璃表面检测方法. 首先, 采用视觉失真感知模块对主干网络特征进行处理, 利用玻璃会引起覆盖图像区域视觉失真的现象, 引导主干网络对玻璃表面进行初步定位, 获取初始玻璃特征图; 其次, 采用结构细化模块以利用图像中玻璃个数信息对初始玻璃特征图逐级细化, 以获得玻璃表面的精细化边缘. 与 17 种方法进行实验的结果表明, 该方法的玻璃表面检测效果在 4 个基准数据集上交并比提升 0.83%~4.73%, F 值提升 1.40%~6.60%, 平均绝对误差降低 0.72%~2.60%, 平衡错误率降低 0.58%~2.66%.

关键词: 玻璃表面检测; 视觉失真; 显著性检测
中图法分类号: TP391.41　　**DOI:** 10.3724/SP.J.1089.2023-00342

Glass Surface Detection Method Based on Visual Distortion

Tan Xin, Qi Fulin, Wang Nan, Zhang Zhizhong*, Xie Yuan, and Ma Lizhuang

(*School of Computer Science and Technology, East China Normal University, Shanghai*　200062)

Abstract: A glass surface detection method based on visual distortion clues is proposed to address the challenges in detecting glass surfaces caused by the characteristics of glass materials. Firstly, the backbone features are processed by a visual distortion aware module. By leveraging the phenomenon of visual distortion in the image regions covered by glass, the backbone network is guided to perform an initial localization of the glass surface, thereby obtaining an initial glass feature map. Subsequently, a structural refinement module is employed to progressively refine the initial glass feature map by utilizing the information about the number of glass objects in the image, resulting in finely-detailed edges of the glass surface. The experimental results conducted with 17 different methods indicate that the proposed method achieves improvements in IoU ranging from 0.83% to 4.73%, improvements in F-measure ranging from 1.40% to 6.60%, reductions in MAE ranging from 0.72% to 2.60%, and decreases in BER ranging from 0.58% to 2.66% across four benchmark datasets.

Key words: glass surface detection; visual distortion; salient object detection

玻璃表面检测是语义分割领域中一项较新的、极具挑战性的任务. 受其物理性质影响, 玻璃表面在实际场景中通常缺乏纹理; 同时, 玻璃复杂的动态外观受光线和物质的相互作用(即反射、折射和

收稿日期: 2023-06-17; 修回日期: 2023-09-28. **基金项目**: 上海市青年科技英才扬帆计划(23YF1410500); 国家自然科学基金(62302167, 62222602). **谭鑫**(1996—), 男, 博士, 副研究员, 硕士生导师, CCF 会员, 主要研究方向为计算机视觉、模式识别; **齐福霖**(2000—), 女, 硕士研究生, 主要研究方向为计算机视觉; **王楠**(1993—), 女, 博士, 主要研究方向为计算机视觉、深度学习; **张志忠**(1991—), 男, 博士, 副教授, 硕士生导师, CCF 会员, 论文通信作者, 主要研究方向为机器学习、计算机视觉; **谢源**(1987—), 男, 博士, 教授, 博士生导师, CCF 会员, 主要研究方向为机器学习、计算机视觉; **马利庄**(1963—), 男, 博士, 教授, 博士生导师, CCF 杰出会员, 主要研究方向为计算机图形图像、计算机视觉、数字媒体、智能信息处理等.

透射)、物体形状以及背景环境等多种因素的影响[1]，因此，对玻璃的检测存在一定的困难. 玻璃在日常生活中随处可见，其存在可能对视觉任务如深度估计[2]和实例分割等[3]造成灾难性影响. 因此，在复杂的现实场景中准确地检测玻璃表面，是一项极具挑战性和实用意义工作.

近年来，随着卷积神经网络(convolutional neural network, CNN)的蓬勃发展、大规模玻璃表面数据集的出现[4-7]，以及人们对复杂场景中环境解析的客观需求提升，越来越多的人致力于研究玻璃表面检测问题. Mei 等[4]通过扩大感受野提取丰富的上下文信息，用于检测场景中尺寸多样的玻璃表面. Yu 等[7]对不同层次的特征进行融合，利用高层语义信息和低层信息进行玻璃表面检测. 但是，由于在设计过程中忽视与玻璃性质结合，这些方法在复杂场景中可能无法获取充足的上下文信息，导致方法失效. He 等[8]利用边界信息进行玻璃表面检测，然而，考虑可提供边界信息的参照物体在单幅 RGB 输入图像中并不总是存在的，该方法可能会失效. Mei 等[1]结合玻璃偏振性，利用 RGB 图像和偏振图像提供的强度信息和偏振信息进行玻璃表面检测. Lin 等[5]在进一步提取和融合上下文信息的基础上，利用玻璃反射性质，以玻璃反射图作为额外输入对预测结果图进行细化. 尽管这些方法在检测过程中尝试结合玻璃性质，但它们只有在满足额外训练输入的要求时才较为有效，同时造成不必要的人力和财力开销. 因此，玻璃表面检测仍然是一个待解决的问题，具有极高的研究价值. 为了在玻璃表面检测任务中取得更好的效果，本文将玻璃性质与玻璃表面检测方法相结合，实现了在无需额外训练输入的前提下进行有效的检测.

经过对玻璃性质的调研[9-10]，本文注意到玻璃的光学效应会导致被覆盖区域失真，如出现亮度退化、色彩偏差和模糊等现象. 因此，本文将该信息作为研究重点. 当透过玻璃观察原始场景时，玻璃的光学效应使原始场景的纹理可能会被放大或者倾斜，导致玻璃覆盖区域的纹理扭曲，进而使图像中该区域失真. 玻璃区域与非玻璃区域的失真程度存在差异，这一差异可为主干网络准确地识别玻璃区域提供重要的指导. 考虑现实中难以直接观测到玻璃导致的区域失真，本文提出了一种基于特征域的可学习模型，通过提取视觉失真特征引导主干网络在无需额外训练数据的情况下精确地检测玻璃表面. 为避免复杂场景中遮挡物对玻璃表面检测造成影响，本文通过分类任务获取玻璃表面个数信息辅助玻璃表面检测，便于在玻璃表面尺寸过小或过大，甚至不存在或为全玻璃等极端情况下，对玻璃表面进行有效的检测.

本文提出一种基于视觉失真的玻璃表面检测方法，旨在利用玻璃表面光学效应造成的图像失真，在无额外训练输入的情况下有效地检测玻璃表面. 首先，本文引入了视觉失真感知模块(visual distortion aware module, VDA)，利用玻璃区域和非玻璃区域之间的失真程度差异，引导主干网络对玻璃表面进行初步定位；其次，提出了结构细化模块(structural refinement module, SRM)，通过分类任务提取玻璃表面个数信息，提高在极端场景下的检测精度，并对预测的玻璃特征图进行迭代细化；最后，本文在广泛使用的 RGB 玻璃表面数据集上进行了实验，证明所提方法的有效性.

1　相关工作

1.1　显著性检测

显著性检测的目的是在图像中检测出突出的视觉物体. 近年来，基于深度学习的显著性检测方法均广泛利用了显著性特征. Qin 等[11]建议在预测的细化架构中利用边界信息. Pang 等[12]对相邻层的特征进行整合，以获得更丰富的多尺度特征. Siris 等[13]通过感知上下文学习方法提取语义场景上下文. 何伟等[14]采用空间注意力和通道注意力机制，聚合深层特征和浅层特征进行显著性检测. 邓慧萍等[15]利用跨层特征融合模块，有效地融合相邻层特征进行显著性检测. 然而，由于玻璃表面在图像中借用周围场景或物体的视觉外观(如颜色、纹理等)并可能多次出现，与具有一致视觉外观且位于图像中心位置的突出视觉物体不同，因此直接应用显著性检测方法检测玻璃表面可能无效.

1.2　镜面检测

镜面检测指识别或检测场景中平面镜的过程. 平面镜内的虚像会使视觉系统混淆. 例如，Yang 等[16]通过学习镜面内呈现内容和镜面外环境内容的不连续性，对场景中的镜面进行检测. Lin 等[17]利用镜面内呈现内容和镜面外被反射内容之间的相似性进行镜面检测. Guan 等[18]提出利用镜面与周围物体的语义联系信息检测镜面. Tan 等[19]利用镜面本身的视觉手性(由 Lin 等[20]提出，描述镜面反射引起的视觉数据的统计变化)引导网络进行检测. 在实际场景中，镜面往往具有自身的边界并对周

围环境中的物体进行反射, 明显的边界划分和场景中内容的对称性为镜面检测提供了有效的引导. 然而, 玻璃本身并不具备边界, 并且其反射作用弱于镜面, 因此上述镜面检测方法可能无法有效地应用于玻璃表面检测任务中.

1.3 透明物体检测

透明物体检测的目的是检测现实场景中形状、尺寸各异的透明物体, 该技术有助于机器人和机器手在室内外场景中进行高效交互. 早期的方法是利用额外的传感器或标注获取补充检测信息解决该问题, 例如, 光场信息[21]、深度信息[22]和极化信息[23]等. 近年来, Xie 等[6]提出了一种利用透明物体边界信息进行检测的方法, 并建立了一个大规模透明物体数据集 Trans10K. 该数据集根据透明物体的尺寸、用途等将透明物体分为“things”和“stuff” 2 个类别. 在实际检测过程中, stuff 类别相对于 things 类别的物体更难检测[7]. 因此, 在玻璃表面检测过程中, 上述透明物体检测方法可能无法达到较好的检测效果.

1.4 玻璃表面检测

在玻璃表面检测中, 现有的方法通常利用上下文信息[4,7]、依赖边界信息[8], 或引入额外的标注信息[1,5]进行检测.

Mei 等[4]在语义分割网络中引入了大范围上下文特征聚合模块, 通过获取更大的感受野来获得丰富的、不同层次的上下文特征, 从而提高对尺寸不一的玻璃制品检测的准确率. Yu 等[7]通过判别性增强模块和基于聚焦和探索的融合模块, 有效地弥合了不同层次特征之间的差距, 充分利用深层特征中的语义信息进行目标定位, 同时在浅层特征中保留底层信息, 以细节化地检测玻璃, 在常规场景下可利用充足的上下文信息进行准确检测. 然而, 这些方法在设计过程中未充分考虑玻璃性质, 可能因环境影响而无法获取充足的上下文信息, 导致检测失效.

He 等[8]采用增强边界学习进行检测, 通过细化差分模块提取细致的边界信息, 并采用基于点的图卷积网络模块进行边界细化, 在常规场景中可利用明显的参照物边界信息, 实现玻璃表面的检测. 然而, 由于单幅 RGB 输入图像并不总是包括提供边界信息的参照物体(如门框、窗框和警示贴纸等), 这些方法可能无法取得理想的检测效果.

Lin 等[5]在增加玻璃表面反射图像输入的前提下, 采用了基于反射的细化模块, 利用了玻璃的反射性质, 更好地区分了玻璃区域和非玻璃区域, 在反射信息充足的场景中可较好地检测玻璃表面. Mei 等[1]利用了玻璃的偏振性质, 在训练过程中使用了偏振相机(LUCID PHX050S)捕获的偏振图像和 RGB 图像, 并将其作为输入, 充分地利用了强度信息和补充的偏振信息, 可在 RGB 图像质量有限时进行玻璃表面检测. 尽管这些方法都试图利用玻璃的性质, 但它们仅在满足额外标注的前提下有效, 并且获取额外训练输入增加了不必要的人力和财力成本.

尽管现有方法在特定场景(具有丰富的上下文信息、明确的参照物边界、额外的标注信息可获取)下取得了较好的检测结果, 但由于它们受适用场景的限制, 其泛用性受到了影响, 无法在复杂的现实场景中进行准确的检测并获得理想的结果. 由于玻璃性质以及现有方法的有限泛用性, 因此玻璃表面检测仍然是一个具有研究价值和挑战性的问题.

与先前的方法不同, 本文着眼于解决玻璃表面检测中的难题, 即在现实世界中, 玻璃表面的环境条件可能复杂多变, 并且常常缺乏额外的辅助信息. 为了应对这一挑战, 本文通过深入研究玻璃性质, 在无额外信息输入的情况下, 降低算法对环境上下文信息和参照物信息的依赖性, 并提高算法在极端场景下的性能, 进而实现在多种复杂场景下有效地检测玻璃表面. 本文关注实际场景中玻璃表面造成的局部失真现象, 通过在特征空间中提取玻璃区域和非玻璃区域失真程度的差异信息, 无须参照物体或额外标注信息实现对玻璃表面的检测.

2 本文方法

2.1 概 述

本文提出一种基于视觉失真的玻璃表面检测网络, 由 VDA 和 SRM 构成. 该网络从单幅输入图像中提取玻璃表面光学效应导致的图像失真程度差异, 并利用其来引导主干网络检测玻璃区域. 在检测到玻璃区域后, 网络利用玻璃表面个数信息逐级细化预测掩模, 从而获得具有精细边界的玻璃掩模. 本文方法工作流程如图 1 所示. 首先, 通过主干网络 SwinB[24]从输入图像中提取基础特征 $F_g^i(i=5,4,3,2,1)$ 作为 VDA 的输入. VDA 通过提取和融合不同分辨率下的视觉失真特征, 生成视觉失真感知特征 $F_{\mathrm{VDA}}^i(i=5,4,3,2,1)$. 为了充分地利用网络不同层的视觉失真感知特征, 本文采用轻量解码

器融合 $F^i_{\mathrm{VDA}}(i=5,4,3)$ 生成初始玻璃特征图 g^4. 其次，SRM 以玻璃特征图 $g^{i+1}(i=3,2,1)$ 和视觉失真感知特征 $F^i_{\mathrm{VDA}}(i=3,2,1)$ 作为输入，通过预测玻璃表面个数的类别 $k^i(i=3,2,1)$ 提取图像中玻璃表面个数信息，用于细化输出的玻璃特征图 $g^i(i=3,2,1)$. 最终，通过对玻璃特征图 g^1 进行二分上采样恢复图像分辨率，再通过 Sigmoid 函数获得输出的玻璃掩模.

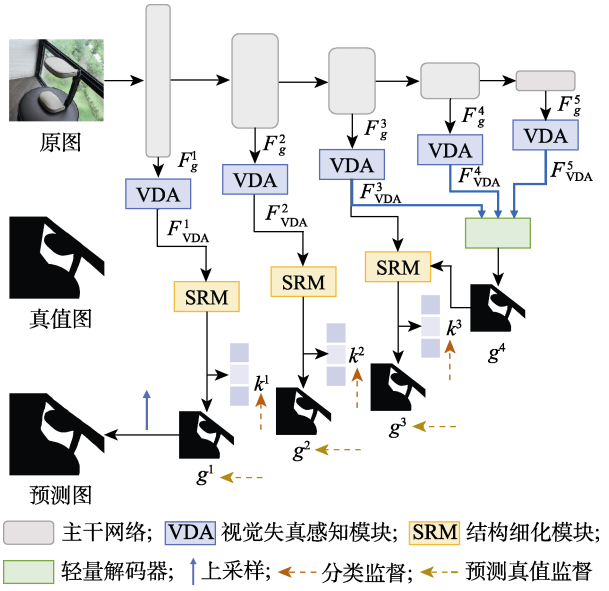


图 1　本文网络的整体框架

2.2　VDA

考虑玻璃的光学效应，原始背景透过玻璃时会被放大或倾斜，导致玻璃区域和非玻璃区域间失真程度不同. 这种现象可以为主干网络进行玻璃表面检测提供引导. 因此，本文提出 VDA，关注玻璃造成的图像失真现象，利用并行多分辨率卷积流提取高分辨率下的精确细节信息以及低分辨率下的丰富上下文信息，据此对比玻璃区域和非玻璃区域之间的失真程度差异，利用失真程度差异引导主干网络对玻璃表面进行定位和范围估计. VDA 以主干网络提取的基础特征 F_g 为输入，生成视觉感知特征 F_{VDA}. VDA 符号说明如表 1 所示，结构如图 2 所示，以主干网络单层基础特征为例说明 VDA 功能.

具体而言，VDA 首先对输入特征图进行预处理，采用卷积核大小为 3，步长为 1 的卷积运算实现通道数降维，以降低模型计算量并实现跨通道信息融合. 在此基础上，采用并行卷积运算(卷积核大小为 3，步长为 2)，调整特征图的空间维度，扩大感受野范围，获取不同分辨率下玻璃特征 $F_{g_n}(n=1,2,4)$，用以学习长程失真程度差异，其计算过程为

$$\begin{cases} F_{g_1} = \mathrm{conv}_{3\times3,1}(F_g) \\ F_{g_2} = \mathrm{conv}_{3\times3,1}(F_g)\downarrow \\ F_{g_4} = \mathrm{conv}_{3\times3,1}(F_g)\downarrow\downarrow \end{cases} \tag{1}$$

其中，F_g 表示主干网络当前层提取的基础特征；$\mathrm{conv}_{k\times k,s}$ 表示卷积核大小为 k；步长为 s 的卷积运算；$\downarrow$ 和 $\downarrow\downarrow$ 分别表示进行单次和 2 次卷积核大小为 3，步长为 2 的卷积操作.

表 1　VDA 符号说明

符号表示	符号说明
F_g	主干网络当前层提取的基础特征
F_{g_n}	不同分辨率下的玻璃特征
F_{s_n}	不同分辨率下的残差细化特征
F_{VD_n}	不同分辨率下视觉失真特征
F_{VDA}	当前层视觉失真感知特征

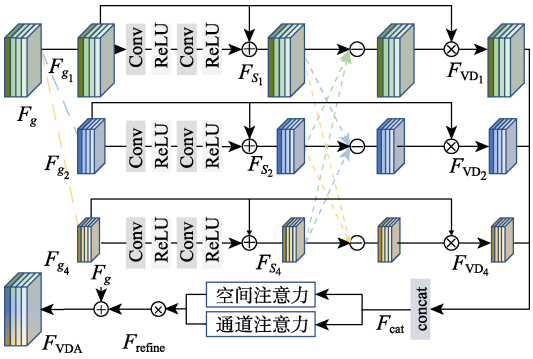


图 2　VDA 示意图

本文继而采用残差连接结构处理不同分辨率输入特征，实现特征细化. 残差连接结构由连续的 2 个卷积层构成，卷积核大小为 3，步长为 1，以不同分辨率下玻璃特征 $F_{g_n}(n=1,2,4)$ 作为输入，通过逐像素相加运算生成不同分辨率下的残差细化特征 $F_{s_n}(n=1,2,4)$，计算过程为

$$F_{s_n} = \left(\mathrm{conv}_{3\times3,1}\circ\mathrm{conv}_{3\times3,1}(F_{g_n})\right)\oplus F_{g_n} \tag{2}$$

其中，$\mathrm{conv}_{k\times k,s}$ 表示卷积核大小为 k，步长为 s 的卷积运算；$\oplus$ 表示逐像素相加操作.

考虑现实世界中玻璃所导致的视觉失真现象难以被人类直接感知，本文通过在特征域中对比玻璃区域与非玻璃区域失真程度差异，学习玻璃区域造成的图像失真信息. 采用卷积核大小为 3、

步长为 2 的卷积操作, 以及卷积核大小为 4, 步长为 2 的反卷积操作调整特征图 $F_{s_n}(n=1,2,4)$ 尺寸, 在不同分辨率下通过逐像素相减运算提取失真程度差异, 再通过逐像素相乘运算将其嵌入到不同分辨率下玻璃特征 $F_{g_n}(n=1,2,4)$ 中, 生成不同分辨率下视觉失真特征 $F_{\mathrm{VD}_n}(n=1,2,4)$, 以引导主干网络有效地检测场景中尺寸不一的玻璃表面, 其计算过程为

$$\begin{cases} F_{\mathrm{VD}_1}=\left(F_{s_1}\ominus\left(F_{s_2}\uparrow\right)\ominus\left(F_{s_2}\Uparrow\right)\right)\otimes F_{g_1} \\ F_{\mathrm{VD}_2}=\left(F_{s_2}\ominus\left(F_{s_1}\downarrow\right)\ominus\left(F_{s_4}\uparrow\right)\right)\otimes F_{g_2} \\ F_{\mathrm{VD}_4}=\left(F_{s_4}\ominus\left(F_{s_1}\Downarrow\right)\ominus\left(F_{s_2}\downarrow\right)\right)\otimes F_{g_4} \end{cases} \tag{3}$$

其中, $\uparrow$ 和 $\Uparrow$ 分别表示进行单次和连续 2 次卷积核大小为 4, 步长为 2 的反卷积操作; $\ominus$ 表示逐像素相减操作; $\otimes$ 表示逐像素相乘操作.

为避免复杂场景中冗余背景信息对玻璃表面检测造成干扰, 本文对不同分辨率下视觉感知特征 $F_{\mathrm{VD}_n}(n=1,2,4)$ 进行拼接, 生成拼接特征 F_{cat}. 然后采用通道注意力机制和空间注意力机制[13]进行特征细化, 以自适应增强与失真程度差异相关的通道特征; 为有效地利用空间信息, 通过逐像素相乘运算对通道增强特征和空间增强特征进行融合, 生成细化特征 F_{refine}. 最后, 采用逐像素相加运算将失真程度差异嵌入到主干网络基础特征 F_g 中, 引导主干网络进行有效的玻璃表面检测, 并生成视觉失真感知特征 F_{VDA}. 其计算过程为

$$\begin{cases} F_{\mathrm{cat}}=\mathrm{concat}\left(F_{\mathrm{VD}_1},\left(F_{\mathrm{VD}_2}\uparrow\right),\left(F_{\mathrm{VD}_4}\Uparrow\right)\right) \\ F_{\mathrm{refine}}=A_{\mathrm{channel}}\left(F_{\mathrm{cat}}\right)\otimes A_{\mathrm{spatial}}\left(F_{\mathrm{cat}}\right) \\ F_{\mathrm{VDA}}=F_{\mathrm{refine}}\oplus F_g \end{cases} \tag{4}$$

其中, concat 表示特征图拼接; A_{channel} 表示通道注意力机制, A_{spatial} 表示空间注意力机制; F_{refine} 表示经过通道注意力和空间注意力机制处理后的细化特征.

为了充分地利用不同层次视觉失真预测玻璃实际位置和范围, 本文采用一个轻量解码器, 对第 3~5 层主干网络层经 VDA 处理后的视觉感知特征 $F_{\mathrm{VDA}}^{i}(i=3,4,5)$ 进行融合, 以生成初始预测玻璃特征图 g_4. 该轻量解码器由 3 个用于缩减通道数的并行 3×3 卷积层, 2 个用于获取详细信息的串行 3×3 卷积层, 以及一个用于生成初始玻璃特征图 g_4 的 1×1 卷积层构成. 该解码器的目标是在尽可能不增加计算复杂度的前提下, 利用不同层次的失真程度差异进行特征融合, 以生成准确的位置和范围预测的初始玻璃特征图. 为了进一步改进最终预测结果的精度, 本文还利用浅层视觉失真感知特征中的细节信息, 对预测结果进行迭代细化. 在这个过程中, 考虑深层特征具有更丰富的语义信息, 对后 3 个阶段的视觉失真感知信息进行融合, 以利用多层次的失真程度差异. 另外, 由于浅层特征通常捕获图像的低级特征且对噪声等信息敏感, 可能会对检测过程造成干扰, 因此本文采用解码器的形式对后 3 层提取的特征进行融合, 并预测初始的玻璃特征图 g_4.

2.3 SRM

已有工作[6,8]利用玻璃表面周围参照物(如门框、窗框等)提供的边界信息对预测玻璃特征图进行结构细化, 通过增加边界真值图作为额外训练输入, 显式地预测玻璃边界图, 并将边界信息引入玻璃掩模输出过程. 然而, 显式输出玻璃边界图一方面增加了额外的预处理过程, 不精确的边界预测图可能导致原玻璃预测掩模出现混淆, 网络关注部分偏移, 影响实际输出效果. 另一方面, 这些工作依赖参照物体提供的边界信息, 忽视了现实场景中可能存在玻璃表面不存在的情况, 导致模型倾向于关注场景中边界内容, 而非玻璃表面本身, 这在一定程度上影响了模型的泛化能力.

与前述方法不同, SRM 以分类子任务的形式利用玻璃表面个数信息, 提高玻璃预测掩模的精度. SRM 在分类分支中引入玻璃表面个数进行监督, 以缓解极端情形下检测困难的问题(如图像中不存在玻璃表面或图像中全为玻璃表面). 在细化分支中, 通过迭代重用玻璃特征图抑制视觉失真感知特征 F_{VDA}^{i} 中非玻璃区域造成的特征干扰, 输出具有精确定位和范围的玻璃掩模. SRM 符号说明如表 2 所示, 结构如图 3 所示.

表 2 SRM 符号说明

符号表示	符号说明
F_{VDA}^{i}	第 i 层获取的视觉失真感知特征
g^{i+1}	第 $i+1$ 层获取的玻璃特征图
g^{i}	第 i 层 SRM 处理后获得的玻璃特征图
k^{i}	第 i 层预测图像中玻璃表面个数类别

SRM 首先对视觉感知特征 F_{VDA}^{i} 进行预处理, 采用并行膨胀卷积运算构造特征金字塔(膨胀率分别设定为 1, 2, 3, 4, 且卷积核大小为 3), 以捕捉上下文信息, 并进行通道降维. 本文对提取的特征进

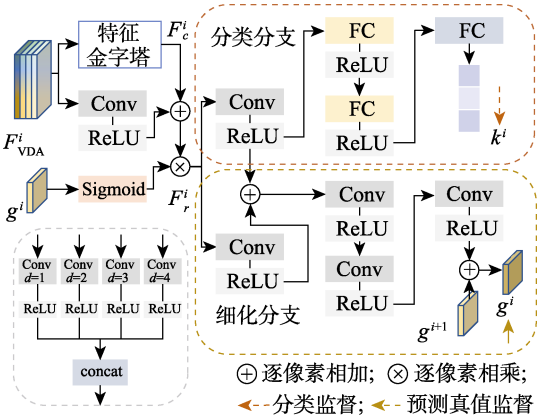


图 3　SRM 示意图

行拼接，通过逐像素相加的方式与降维处理后的视觉失真感知特征进行融合，调整细节信息的影响，生成融合特征 F_c^i.

考虑预测玻璃特征图有助于抑制非玻璃区域造成的特征干扰[16]，本文先对预测玻璃特征图 g^{i+1} 进行上采样，将其调整至融合特征 F_c^i 尺寸后采用 Sigmoid 函数处理. 通过逐像素相乘操作，利用预测玻璃特征图引导融合特征 F_c^i，以关注玻璃区域并生成细化特征 F_r^i. 其计算过程为

$$\begin{cases} F_c^i = \mathrm{concat}\left(\mathrm{conv}_{3\times3}^1\left(F_{\mathrm{VDA}}^i\right),\cdots,\mathrm{conv}_{3\times3}^4\left(F_{\mathrm{VDA}}^i\right)\right) \\ F_r^i = \left(\mathrm{conv}_{1\times1}^1\left(F_{\mathrm{VDA}}^i\right)\oplus F_c^i\right)\otimes\sigma\left(\mathrm{upsample}\left(g^{i+1}\right)\right) \end{cases} \tag{5}$$

其中，F_{VDA}^i 表示第 i 层获取的视觉失真感知特征；concat 表示特征图拼接；F_c^i 表示特征图拼接后获得的融合特征，$\mathrm{conv}_{k\times k}^r$ 表示卷积核大小为 k，膨胀率为 r 的卷积运算；g^{i+1} 表示第 $i+1$ 层获取的玻璃特征图；σ 表示 Sigmoid 函数；upsample 表示二分上采样.

然后，以双分支处理细化特征 F_r^i. 在分类分支中，首先进行卷积操作(卷积核大小为 3，步长为 1)提取包含玻璃表面个数的类别信息，通过全连接层和 ReLU 激活函数预测玻璃表面计数类别 k^i. 本文分别用 0, 1, 2 表示不含玻璃表面的图像(个数等于 0)、仅含有单个玻璃表面的图像(个数等于 1)，以及常规情况下的图像(个数大于等于 2)，引导网络在预测玻璃特征图时关注极端情况. 在细化分支中，首先进行卷积操作(卷积核大小为 3，步长为 1)提取细节信息，再通过逐像素相加操作融合个数类别信息，采用串行卷积层进行特征提取和通道数调整. 最后，采用逐像素相加操作融合预测玻璃特征图 g^{i+1} 信息，得到结构细化处理后的玻璃特征图 g^i.

2.4　损失函数

本文采用二元交叉熵损失(binary cross entropy loss, BCE)函数和交并比损失(intersection over union loss, IoU)函数的混合损失函数，计算预测玻璃掩模和真实值之间的差异. 其中，BCE 函数在像素级计算预测掩模和真实值之间的差异，IoU 函数关注预测掩模中玻璃表面区域[11]. 本文采用交叉熵损失(cross entropy loss, CE)函数计算预测个数类别与真实值之间的差异. 混合损失函数计算过程为

$$\mathcal{L} = \sum_{i=1}^{4}\omega_{\mathrm{BCE}}\mathcal{L}_{\mathrm{BCE}}^i + \sum_{i=1}^{4}\omega_{\mathrm{IoU}}\mathcal{L}_{\mathrm{IoU}}^i + \sum_{i=1}^{3}\omega_{\mathrm{CE}}\mathcal{L}_{\mathrm{CE}}^i \tag{6}$$

其中，$\mathcal{L}_{\mathrm{BCE}}^i$ 和 $\mathcal{L}_{\mathrm{IoU}}^i$ 分别表示由于监督不同层输出掩模的 BCE 和 IoU；$\mathcal{L}_{\mathrm{CE}}^i$ 表示由于监督不同层预测个数类别的 CE. 本文将各损失函数权重分别设置为 $\omega_{\mathrm{BCE}}=1$，$\omega_{\mathrm{IoU}}=1$，$\omega_{\mathrm{CE}}=0.01$.

4　实　验

4.1　实验设置

4.1.1　实验细节

本文方法采用 Pytorch 框架[25]实现，使用单张 NVIDIA GeForce RTX 2080 Ti 进行训练. 在训练过程中，选择 SwinB[24]作为主干网络，并对其他参数进行随机初始化. 对输入图像进行裁剪，设定图像尺寸为 384×384 像素，采用随机缩放、翻转、旋转进行数据增广；采用 Adam 优化器[26]进行训练，设定初始学习率为 10^{-5}；采用“poly”策略[27]调整学习率，并将乘数设定为 $\left(1-I_{\mathrm{current}}/I_{\mathrm{max}}\right)^{0.9}$，其中，$I_{\mathrm{current}}$ 和 I_{max} 分别表示当前迭代次数和总迭代次数. 本文设定训练 epoch=240，批尺寸为 4. 在测试过程中，同样对输入图像进行裁剪，并调整其尺寸为 384×384 像素；经网络处理后恢复原输入尺寸，对网络输出结果不使用如条件随机场(conditional random field, CRF)[28]方法进行后处理操作.

4.1.2　实验数据集

本文在所有广泛使用的玻璃表面数据集上进行实验，包括 GDD 数据集[4]，GSD 数据集[5]，Trans10K 数据集[6]和 HSO 数据集[7]，以证明本文方法的有效性. GDD 数据集[4]包括从室内外场景中收集的 2 980 幅训练图像，936 幅测试图像. GSD 数据集[5]包括从真实场景中采集的 3 285 幅训练图像和 813 幅测试图像. Trans10K 数据集[6]包括 5 000 幅训练

图像, 1 000 幅验证图像, 4 428 幅测试图像, 该数据集将透明物体划分为 things 和 stuff 这 2 个类别. 参照 PGSNet[7], 本文从 Trans10K 数据集中选取 2 455 幅训练图像和 1 771 幅仅包括 stuff 类别的图像进行模型评估. HSO 数据集[7]包括从 4 个广泛使用的数据集中收集的 3 070 幅训练图像和 1 782 幅测试图像, 来源数据集包括 Matterport3D 数据集[29], SUN RGB-D 数据集[30], ScanNet 数据集[31]和 2D3DS 数据集[32].

4.1.3 评估指标

本文采用 4 个广泛使用的评估指标进行效果评估, 包括 IoU、平均绝对误差(mean absolute error, MAE), F 值(F-measure, F_β)和平衡错误率(balance error rate, BER). IoU 在语义分割中广泛使用, 用于计算预测玻璃表面区域和真实玻璃表面区域的重叠度, 其计算公式为

$$\mathrm{IoU}=\frac{\left|M_{\mathrm{P}}\cap M_{\mathrm{G}}\right|}{\left|M_{\mathrm{P}}\cup M_{\mathrm{G}}\right|} \tag{7}$$

其中, M_{P} 表示预测掩码, M_{G} 表示标注掩码. $\left|M_{\mathrm{P}}\cap M_{\mathrm{G}}\right|$ 表示预测掩码和真实标注掩码间交集区域的像素数目; $\left|M_{\mathrm{P}}\cup M_{\mathrm{G}}\right|$ 表示预测掩码和真实标注掩码间并集区域的像素数目. MAE 计算公式为

$$\mathrm{MAE}=\frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}\left|M_{\mathrm{P}}(i,j)-M_{\mathrm{G}}(i,j)\right| \tag{8}$$

其中, H 和 W 分别表示输入图像的高度和宽度; M_{P} 表示预测掩码; M_{G} 表示真实标注掩码. F_β 计算准确率(precision, P)和召回率(recall, R)的调和平均数, 其计算公式为

$$F_\beta=\frac{\left(1+\beta^2\right)P\times R}{\beta^2P+R} \tag{9}$$

准确率和召回率的计算公式为

$$\begin{cases}P=\dfrac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}\\ R=\dfrac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}\end{cases} \tag{10}$$

其中, TP(true positive)表示预测为玻璃表面且实际为玻璃表面的像素数目; FP(false positive)表示预测为玻璃表面但实际为非玻璃表面的像素数目; FN(false negative)表示预测为非玻璃表面但实际为玻璃表面的像素数目. 参考 PGSNet[7]和 EBLNet[8], 本文设定 $\beta=0.3$. BER 计算玻璃区域和非玻璃区域的平均误差, 其计算方式为

$$\mathrm{BER}=1-0.5\times\left(\frac{N_{\mathrm{TP}}}{N_{\mathrm{P}}}+\frac{N_{\mathrm{TN}}}{N_{\mathrm{N}}}\right) \tag{11}$$

其中, N_{TP}, N_{TN}, N_{P}, N_{N} 分别表示真正例像素数、真负例像素数、玻璃区域和非玻璃区域像素数.

4.2 实验结果及分析

4.2.1 与其他方法比较

为验证本文法的有效性, 将其与相关方法进行对比实验. 包括从语义分割中选取 PSPNet[33], DANet[34], CCNet[35]和 FaPN[36]; 从阴影分割中选取 BDRAR[37]; 从显著性检测中选取 BASNet[11]和 MINet[12]; 从伪装物体检测中选取 SINet[38]; 从医学图像检测中选取 PraNet[39]; 从镜面检测中选取 MirrorNet[16]和 PMDNet[17]; 从透明物体检测中选取 TransLab[6]和 Trans2Seg[40]; 从玻璃表面检测中选取 GDNet[4], GSDNet[5], PGSNet[7]和 EBLNet[8].

4.2.2 定量实验结果

本文与多领域检测方法比较, 以证明本文方法有效性, 其中, IoU 和 F_β 数值越高越好, MAE 和 BER 数值越低越好. 定量实验结果对比如表 3 和表 4 所示, 在无后处理操作的情况下, 本文方法在 4 个广泛使用的玻璃表面检测数据集上关于 4 个标准评估指标均取得提升.

4.2.3 定性实验结果

为进一步验证本文方法在现实场景中的有效性, 在 GDD 数据集[4]上与 7 种方法进行可视化对比. 如图 4 所示, 图中从左至右依次为用于阴影检测的 BDRAR[37]、用于显著性检测的 MINet[12]、用于伪装物体检测的 SINet[38]、用于透明物体检测的 TransLab[6]、用于玻璃表面检测的 GDNet[4], GSDNet[5]和 EBLNet[8], 以及本文方法的可视化结果和真值. 对比实验结果表明, 现有前沿水平方法容易受到场景中复杂信息的影响导致玻璃表面漏检, 例如, 地面反光(第 1 行)和地面材质差异(第 2 行). 与先前不同, 本文方法可以在这些情况下进行有效的检测, 并生成具有精细边界的玻璃表面掩模. 此外, 现有方法大多依赖场景中参考物体的边界信息进行玻璃表面检测, 导致出现检测混淆, 进而造成过检(第 3 行)或漏检(第 4~6 行). 与之不同的是, 本文方法基于玻璃性质, 利用失真程度差异, 可实现玻璃表面初步定位和范围估计. 此外, SRM 的应用, 使得其对于小面积玻璃表面仍能进行有效的检测, 并具有平滑、精细的边界(第 7 行). 综上所述, 本文方法利用失真程度差异和玻璃表面个数信息, 可以在复杂的现实场景中对玻璃表

表 3　GDD 和 GSD 测试集上定量实验结果对比

方法	GDD[4] (CVPR'20)				GSD[5] (CVPR'21)			
	IoU↑	F_β↑	MAE↓	BER↓	IoU↑	F_β↑	MAE↓	BER↓
PSPNet[33]	0.840 6	0.867	0.084	0.087 9	0.702 6	0.834	0.110	0.106 6
BDRAR*[37]	0.800 1	0.847	0.098	0.098 7	0.759 1	0.860	0.081	0.086 1
BASNet[11]	0.828 8	0.896	0.094	0.087 0	0.697 9	0.808	0.106	0.135 4
MINet[12]	0.820 3	0.847	0.092	0.085 5	0.772 9	0.879	0.077	0.095 4
SINet[38]	0.832 7	0.912	0.101	0.083 5	0.770 4	0.875	0.077	0.092 5
TransLab[6]	0.816 4	0.849	0.097	0.097 0	0.780 5	0.828	0.069	0.091 9
Trans2Seg[40]	0.844 1	0.872	0.078	0.073 6	0.796 5	0.839	0.069	0.082 1
MirrorNet*[16]	0.850 7	0.903	0.083	0.076 7	0.742 0	0.828	0.090	0.107 6
PMDNet*[17]	0.870 0	0.930	0.067	0.061 7	0.817 0	0.890	0.061	0.067 4
GDNet[4]	0.876 3	0.937	0.063	0.056 2	0.825 1	0.857	0.058	0.064 1
GSDNet*[5]	0.880 7	0.932	0.059	0.057 1	0.836 4	0.903	0.055	0.061 2
EBLNet[8]	0.882 2	0.935	0.056	0.053 8	0.817 0	0.878	0.059	0.067 5
PGSNet[7]	0.878 1	0.901	0.062	0.055 6	0.836 5	0.868	0.054	0.062 5
本文	**0.918 3**	**0.951**	**0.039**	**0.039 2**	**0.864 9**	**0.920**	**0.040**	**0.054 0**

注. *表示使用 CRF[28]进行后处理；粗体表示最优值.

表 4　Trans10K-stuff 和 HSO 测试集上定量实验结果对比

方法	Trans10K-stuff[6] (ECCV'20)				HSO[7] (IEEE TIP'22)			
	IoU↑	F_β↑	MAE↓	BER↓	IoU↑	F_β↑	MAE↓	BER↓
PSPNet[33]	0.878 9	0.907	0.045	0.054 6	0.776 0	0.814	0.095	0.105 7
BDRAR*[37]	0.850 0	0.870	0.061	0.060 4	0.753 2	0.802	0.101	0.111 3
DANet[34]	0.881 8	0.907	0.045	0.052 8	0.776 9	0.817	0.091	0.106 0
CCNet[35]	0.882 0	0.906	0.044	0.051 5	0.781 7	0.820	0.092	0.103 4
PraNet[39]	0.871 5	0.881	0.058	0.053 1	0.719 3	0.756	0.128	0.131 1
FaPN[36]	0.890 9	0.913	0.042	0.048 0	0.780 5	0.835	0.089	0.095 1
TransLab[6]	0.871 0	0.897	0.051	0.054 4	0.743 2	0.781	0.123	0.120 0
Trans2Seg[40]	0.749 8	0.767	0.124	0.107 3	0.779 8	0.817	0.095	0.096 5
MirrorNet*[16]	0.883 0	0.907	0.047	0.049 5	0.788 2	0.820	0.102	0.099 3
GDNet[4]	0.886 8	0.907	0.046	0.047 2	0.787 3	0.817	0.097	0.093 2
GSDNet*[5]	0.896 7	0.917	0.042	0.045 2	0.788 6	0.818	0.103	0.097 9
PGSNet[7]	0.897 9	0.917	0.042	0.043 9	0.800 6	0.836	0.089	0.090 8
本文	**0.906 2**	**0.944**	**0.034**	**0.038 1**	**0.847 9**	**0.902**	**0.063**	**0.064 2**

注. *表示使用 CRF[28]进行后处理；粗体表示最优值.

面进行有效的检测，而不需要额外的训练数据.

4.2.4　性能评估

为验证本文方法的性能，将其与相关方法关于浮点运算(floating point operations, FLOPs)和每秒传输帧数(frames per second, FPS)进行比较. FLOPs 是一种用于衡量计算机程序或算法中浮点运算数量的度量单位，表示在执行计算任务期间所执行的浮点运算的总数，主要用于评估和比较不同模型或算法的计算复杂度和效率, FLOPs 数值越低越好，更低的 FLOPs 表示模型计算复杂度越低，所需的计算资源越少；FPS 是用于衡量图像或视频处理系统性能的指标，表示模型每秒处理的帧数，FPS 数值越高越好，更高的 FPS 表示模型能够处理更多的帧数，提供更流畅的图像输出.

对比方法包括用于镜面检测的 PMDNet[17]；用于玻璃表面检测的 GDNet[4], GSDNet[5]和 PGSNet[7]. 为了公平比较，本文将输入图像尺寸统一设置为 384×384 像素，并采用同一设备在 GSD 数据集[5]



图 4 GDD 测试集上定性实验结果对比

上进行 FPS 测试, 以避免硬件设备对测试结果造成影响. 对比结果如表 5 所示, 本文方法 FLOPs 为 86.67G, FPS 为 24.00, 均取得最好水平, 验证了其在具有较高检测准确率的同时兼顾了模型性能.

表 5 不同方法在 GSD 测试集上模型性能对比

方法	FLOPs↓	FPS↑
PMDNet[17]	101.54G	10.59
GDNet[4]	231.36G	16.79
GSDNet[5]	92.64G	21.92
PGSNet[7]	96.14G	21.00
本文	**86.67G**	**24.00**

注. 粗体表示最优值.

4.3 消融实验

本文首先在主干网络[24]上添加轻量解码器作为基准方法, 在 GSD 数据集[5]上进行了 3 组消融实验, 分别验证了 VDA 和 SRM 的有效性, 实验结果如表 6 所示. 在基准方法上分别添加 VDA, SRM, 以及 VDA 和 SRM 后, IoU 和 F_β 均有提升, MAE 和 BER 均有下降, 由此可见本文提出模块的有效性, 且二者在玻璃表面检测任务中相互促进, 有助于提升对玻璃表面的检测准确率.

表 6 GSD 测试集上消融实验结果

方法	IoU↑	F_β↑	MAE↓	BER↓
基准方法	0.829 9	0.896	0.050	0.067 3
基准方法+VDA	0.847 8	0.908	0.043	0.063 0
基准方法+SRM	0.847 6	0.906	0.047	0.063 0
本文	**0.864 9**	**0.920**	**0.040**	**0.054 0**

注. 粗体表示最优值.

4.4 泛化性评估

受限于现有玻璃表面数据集较少包含不含玻璃表面的情况, 本文选用显著性检测中常用的数据集进行验证, 包括 ECSSD 数据集[41]和 DUTS-TE 数据集[42], 并规定输出检测图像中玻璃表面像素在总像素中占比小于 3%可视为输出图像中不包含玻璃表面. SRM 预测玻璃表面个数类别, 本文方法在 GDD[4]和 GSD[5]数据集上分类准确率分别为 79.48%和 80.31%, 在 Trans10K-stuff[6]和 HSO[7]数据集上分类准确率分别为 81.64%和 73.34%. 本文与以 GSD 数据集[5]为训练数据集的代表方法对比, 当输入图像中不含玻璃表面时, 检测准确率对比实验结果如表 7 所示, 本文方法在无玻璃表面场景下具有较好的泛化性.

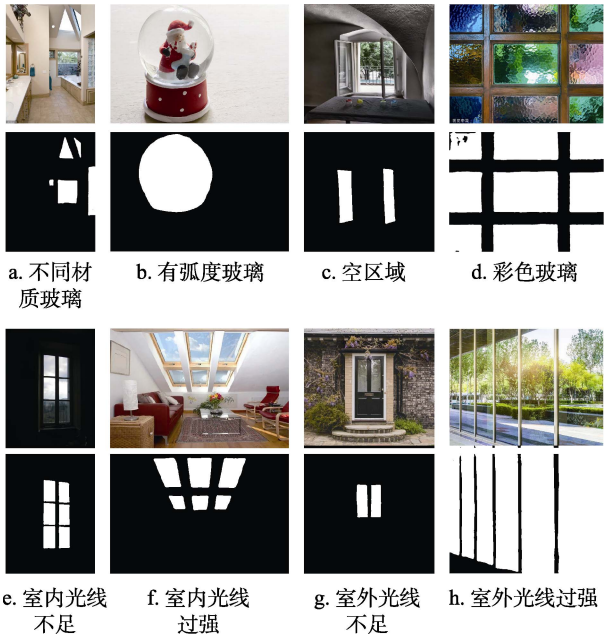
为了进一步证明本文方法具有优异的鲁棒性, 在多种复杂场景下进行了一系列测试和验证, 实

表 7　GSD 测试集上无玻璃场景下分类准确率比较　%

方法	ECSSD[41]	DUTS-TE[42]
GSDNet[5]	10.67	12.47
本文	**76.60**	**77.68**

注. 粗体表示最优值.

验结果如图 5 所示，其中上方为输入图像，下方为本文方法的预测结果. (1) 图 5a 展示了同一场景下不同材质的玻璃表面，图中中部及上方为普通玻璃，而右侧为压花玻璃，本文方法能够准确地检测出这些不同材质的玻璃. (2) 为了考察算法对具有一定弧度的玻璃表面的适应能力，选择了球形玻璃作为示例，结果如图 5b 所示，本文方法能够准确地预测出球形玻璃的位置及范围. (3) 图 5c 展示了一幅图像中包含空区域的情况. 空区域是指没有玻璃覆盖物的图像区域，本文方法能够有效地将这些空区域与玻璃表面进行区分，从而准确地检测玻璃表面. (4) 本文还考虑了玻璃表面可能存在其他化学物质的情况，图 5d 展示了一幅包括彩色玻璃的图像，采用本文方法能够准确地得到相应的预测结果，而不受到玻璃颜色的影响. (5) 本文还对图像拍摄条件的不同情况进行了测试，图 5e 中，考虑了室内环境光线不足的情况，而在图 5f 中是室内环境光线过强；还考虑了图 5g 所示室外环境光线不足和图 5h 所示室外环境光线过强的情况，以验证本文方法在各种不同光照条件下的稳定性和可靠性. 定性实验结果表明，本文方法在多种复杂场景下具有较好的鲁棒性，能够准确地预测玻璃表面的位置和范围.



a. 不同材质玻璃　b. 有弧度玻璃　c. 空区域　d. 彩色玻璃

e. 室内光线不足　f. 室内光线过强　g. 室外光线不足　h. 室外光线过强

图 5　复杂场景下本文方法鲁棒性定性分析

4.5　局限性分析

本文方法在处理复杂遮挡物边界细节时存在一些限制，需要进一步改进. 由于本文未对遮挡物边界细节进行重点分析，在面对复杂且过于精细的遮挡物时，其预测结果可能无法很好地区分遮挡物边界细节，仍有提升空间. 具体体现如图 6 所示，如在玻璃表面前方存在花瓣和花茎等边界细节较为精细和复杂的遮挡物的情况下，本文的预测结果可能存在一些微小区域的漏检或过检. 为了进一步提升本文方法的性能和准确性，计划通过增强信息提取能力和构建复杂多样化的数据集来解决无玻璃表面场景下的检测问题，并利用更多维度信息提高算法的系统性和鲁棒性.



a. 原图　b. 本文方法　c. 真值图

图 6　本文方法局限性分析

5　结　语

本文提出了一种基于视觉失真的玻璃表面检测方法，旨在利用玻璃性质实现在无额外训练输入的前提下在复杂场景中实现对形式多样的玻璃表面的有效检测，该网络一方面通过 VDA 提取不同分辨率的视觉失真特征，引导主干网络进行玻璃表面检测；另一方面通过 SRM 利用玻璃表面个数信息对预测玻璃掩模进行迭代细化. 本文方法在广泛使用的玻璃表面检测数据集上，包括 GDD 数据集、GSD 数据集、Trans10K-stuff 数据集和 HSO 数据集进行实验，4 个标准评估指标均取得提升. 此外，本文在提升性能的同时注重模型效率，在 FLOPs 和 FPS 均取得较好结果. 本文关注输入图像中不包括玻璃表面的特殊情况，并在该场景下进行了算法泛化性评估，取得了良好的结果.

参考文献(References):

[1] Mei H Y, Dong B, Dong W, *et al*. Glass segmentation using intensity and spectral polarization cues[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 12612-12621

[2] Kusupati U, Cheng S, Chen R, *et al*. Normal assisted stereo depth estimation[C] //Proceedings of the IEEE/CVF Confer-

ence on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 2186-2196

[3] Tian X, Xu K, Yang X, *et al*. Learning to detect instance-level salient objects using complementary image labels[J]. International Journal of Computer Vision, 2022, 130(3): 729-746

[4] Mei H Y, Yang X, Wang Y, *et al*. Don't hit me! Glass detection in real-world scenes[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 3684-3693

[5] Lin J Y, He Z B, Lau R W H. Rich context aggregation with reflection prior for glass surface detection[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 13410-13419

[6] Xie E Z, Wang W J, Wang W H, *et al*. Segmenting transparent objects in the wild[C] //Proceedings of the 16th European Conference on Computer Vision. Heidelberg: Springer, 2020: 696-711

[7] Yu L T, Mei H Y, Dong W, *et al*. Progressive glass segmentation[J]. IEEE Transactions on Image Processing, 2022, 31: 2920-2933

[8] He H, Li X T, Cheng G L, *et al*. Enhanced boundary learning for glass-like object segmentation[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 15839-15848

[9] Murase H. Surface shape reconstruction of an undulating transparent object[C] //Proceedings of the 3rd International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 1990: 313-317

[10] McHenry K, Ponce J, Forsyth D. Finding glass[C] //Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2005: 973-979

[11] Qin X B, Zhang Z C, Huang C Y, *et al*. BASNet: boundary-aware salient object detection[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 7471-7481

[12] Pang Y W, Zhao X Q, Zhang L H, *et al*. Multi-scale interactive network for salient object detection[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 9410-9419

[13] Siris A, Jiao J B, Tam G K L, *et al*. Scene context-aware salient object detection[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 4136-4146

[14] He Wei, Pan Chen. The salient object detection based on attention-guided network[J]. Journal of Image and Graphics, 2022, 27(4): 1176-1190(in Chinese)
(何伟, 潘晨. 注意力引导网络的显著性目标检测[J]. 中国图象图形学报, 2022, 27(4): 1176-1190)

[15] Deng Huiping, Cao Zhaoyang, Xiang Sen, *et al*. Saliency detection based on context-aware cross-layer feature fusion for light field images[J]. Journal of Electronics & Information Technology, 2023, 45(12): 4489-4498(in Chinese)
(邓慧萍, 曹召洋, 向森, 等. 基于上下文感知跨层特征融合的光场图像显著性检测[J]. 电子与信息学报, 2023, 45(12): 4489-4498)

[16] Yang X, Mei H Y, Xu K, *et al*. Where is my mirror?[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2019: 8808-8817

[17] Lin J Y, Wang G D, Lau R W H. Progressive mirror detection[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 3694-3702

[18] Guan H K, Lin J Y, Lau R W H. Learning semantic associations for mirror detection[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 5931-5940

[19] Tan X, Lin J Y, Xu K, *et al*. Mirror detection with the visual chirality cue[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(3): 3492-3504

[20] Lin Z Q, Sun J, Davis A, *et al*. Visual chirality[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 12292-12300

[21] Xu Y C, Nagahara H, Shimada A, *et al*. TransCut: transparent object segmentation from a light-field image[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2015: 3442-3450

[22] Zhu Y Y, Qiu J X, Ren B. Transfusion: a novel slam method focused on transparent objects[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 5999-6008

[23] Kalra A, Taamazyan V, Rao S K, *et al*. Deep polarization cues for transparent object segmentation[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 8599-8608

[24] Liu Z, Lin Y T, Cao Y, *et al*. Swin Transformer: hierarchical vision transformer using shifted windows[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 9992-10002

[25] Paszke A, Gross S, Massa F, *et al*. PyTorch: an imperative style, high-performance deep learning library[C] //Proceedings of the 33rd International Conference on Neural Information Processing Systems. New York: ACM Press, 2019: Article No.721

[26] Kingma D P, Ba J. Adam: a method for stochastic optimization[OL]. [2023-06-17]. https://arxiv.org/abs/1412.6980

[27] Chen L C, Papandreou G, Schroff F, *et al*. Rethinking atrous convolution for semantic image segmentation[OL]. [2023-06-17]. https://arxiv.org/abs/1706.05587

[28] Krähenbühl P, Koltun V. Efficient inference in fully connected CRFs with Gaussian edge potentials[C] //Proceedings of the 25th International Conference on Neural Information Processing Systems. New York: ACM Press, 2011: 109-1174

[29] Chang A, Dai A, Funkhouser T, *et al*. Matterport3D: learning from RGB-D data in indoor environments[C] //Proceedings of the International Conference on 3D Vision. Los Alamitos: IEEE Computer Society Press, 2017: 667-676

[30] Song S R, Lichtenberg S P, Xiao J X. SUN RGB-D: a RGB-D

scene understanding benchmark suite[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2015: 567-576

[31] Dai A, Chang A X, Savva M, *et al*. ScanNet: richly-annotated 3D reconstructions of indoor scenes[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 2432-2443

[32] Armeni I, Sax S, Zamir A R, *et al*. Joint 2D-3D-semantic data for indoor scene understanding[OL]. [2023-06-17]. https://arxiv.org/abs/1702.01105

[33] Zhao H S, Shi J P, Qi X J, *et al*. Pyramid scene parsing network[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 6230-6239

[34] Fu J, Liu J, Tian H J, *et al*. Dual attention network for scene segmentation[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 3141-3149

[35] Huang Z L, Wang X G, Huang L C, *et al*. CCNet: criss-cross attention for semantic segmentation[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2019: 603-612

[36] Huang S H, Lu Z C, Cheng R, *et al*. FaPN: feature-aligned pyramid network for dense image prediction[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 844-853

[37] Zhu L, Deng Z J, Hu X W, *et al*. Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection[C] //Proceedings of the 15th European Conference on Computer Vision. Heidelberg: Springer, 2018: 122-137

[38] Fan D P, Ji G P, Sun G L, *et al*. Camouflaged object detection[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 2774-2784

[39] Fan D P, Ji G P, Zhou T, *et al*. PraNet: parallel reverse attention network for polyp segmentation[C] //Proceedings of the 23rd International Conference on Medical Image Computing and Computer Assisted Intervention. Heidelberg: Springer, 2020: 263-273

[40] Xie E Z, Wang W J, Wang W H, *et al*. Segmenting transparent objects in the wild with transformer[OL]. [2023-06-17]. https://arxiv.org/abs/2101.0846

[41] Shi J P, Yan Q, Xu L, *et al*. Hierarchical image saliency detection on extended CSSD[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(4): 717-729

[42] Wang L J, Lu H C, Wang Y F, *et al*. Learning to detect salient objects with image-level supervision[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 3796-3805