RGB versus Early-Fusion RGB-D Glass Segmentation: A Controlled Benchmark-to-Robot Evaluation on Pioneer P3-DX

EBUBECHUKWU AMASIATU[1], AND SALEM AMEEN[1]

[1] School of Science, Engineering and Environment, University of Salford, Manchester, United Kingdom
Corresponding author: Ebubechukwu Amasiatu (e-mail: e.amasiatu@edu.salford.ac.uk).

ABSTRACT Transparent glass is a persistent perception hazard for indoor mobile robots: RGB boundaries can be visually ambiguous, while commodity depth sensors often return missing or distorted measurements. This study isolates the effect of adding raw depth to a lightweight transformer segmenter. Two SegFormer-B0 models were trained on identical partitions of the 3,009-image RGB-D Glass Surface Detection dataset using the same optimization and checkpoint-selection protocol: a three-channel RGB model and a four-channel early-fusion RGB-D model. On the fixed 609-image test split, RGB outperformed RGB-D on seven of eight reported measures, reaching 0.9448 pixel accuracy, 0.8234 mean intersection over union, 0.7108 glass IoU, 0.8376 recall, and 0.8310 F1. RGB-D retained higher precision (0.8374 versus 0.8244). RGB also reduced mean inference latency by 10.7% (10.24 versus 11.47 ms) and increased throughput by 12.0% (97.63 versus 87.18 frames/s). The frozen checkpoints were then evaluated in 20 Pioneer P3-DX navigation runs. At the selected center-ratio threshold of 0.35, each model produced glass-triggered stops in both of its two glass runs (4/4 combined). Across all recorded thresholds, RGB triggered in 5/5 glass runs and RGB-D in 3/5; the independent distance safety stop terminated the remaining two RGB-D runs. Neither model produced a glass trigger in the ten non-glass runs. Under this controlled early-fusion design, raw depth did not improve unseen-test segmentation and reduced threshold robustness during robot deployment.

INDEX TERMS Autonomous mobile robots, glass surface detection, RGB-D sensing, robot perception, semantic segmentation, SegFormer, transparent obstacles.

I. INTRODUCTION

Autonomous mobile robots increasingly operate in built environments where transparent doors and partitions lie directly on navigable routes [1]. Glass remains difficult for both appearance and range sensing [2], [3]. Transmission, reflection, refraction, and weak material boundaries can make a barrier resemble free space [4], [5], while active depth measurements may be invalid or may describe surfaces behind the glass. A navigation system that accepts either observation without a glass-specific check can therefore commit to an unsafe trajectory.

RGB glass segmentation has progressed from contextual inference to reflection-aware, semantic-relation, and polarization-based models [6]-[10]. The Trans10K benchmark extended transparent-object segmentation to diverse real scenes [11]. These advances establish strong pixel-level recognition, but benchmark performance alone does not show whether a model delivers stable, low-latency evidence to a moving robot.

This study isolates one deployment question: does appending raw depth as a fourth input channel improve glass segmentation over RGB alone? The comparison holds the dataset, SegFormer-B0 backbone, split, optimization protocol, checkpoint rule, test metrics, computer, camera, robot, and stopping logic constant. The only architectural change is expansion of the first patch embedding from three to four channels. Frozen checkpoints are evaluated first on the fixed test split and then in the same robot pipeline.

The paper makes four contributions. First, it provides a paired, modality-controlled ablation of RGB against raw-depth early fusion. Second, it documents a validation-to-test ranking reversal that would be hidden by validation results alone. Third, it reports class-specific overlap, precision, recall, F1, latency, and throughput under common experimental conditions. Fourth, it connects those measurements to exact outcomes from 20 Pioneer P3-DX runs using an explicit decision region, threshold sweep, and independent safety layer. The findings provide controlled evidence about a common fusion choice rather than claiming that one sensor modality is universally superior.

II. RELATED WORK

A. RGB GLASS SEGMENTATION

RGB-only methods recover glass from cues that remain visible despite transparency. GDNet aggregates multi-scale context [6]; reflection-prior models combine contextual and reflection evidence [7]; and GlassSemNet models semantic relations between glass and surrounding architectural objects [8]. Polarization supplies additional physical cues but requires specialized imaging hardware [9], [10]. Trans10K broadened the target from glass surfaces to transparent objects in the wild [11]. SegFormer provides a deployment-oriented backbone because its hierarchical transformer encoder is paired with a lightweight multilayer-perceptron decoder [12].

B. RGB-D AND DEPTH-BASED PERCEPTION

Depth can complement appearance, but transparent material is itself a failure case for active range sensing. Early RGB-D work jointly inferred glass boundary and depth [13]. ClearGrasp, TransCG, and ClearPose developed learning-based geometry recovery and transparent-object benchmarks [14]-[16]. Contextual monocular depth for glass walls and broader specular-surface benchmarks further document the difficulty [17]-[19]. RGB-D GSD targets glass surfaces directly and uses cross-modal context mining and missing-depth-aware attention [20]. Together, this literature shows that depth is useful when its reliability and fusion mechanism are modeled explicitly; simply adding a raw channel is a separate empirical question.

C. OPERATIONAL VALIDATION GAP

Glass-aware mapping and boundary-guided recognition have been evaluated on mobile robots [21], [22], but many segmentation studies stop at benchmark metrics. Dataset bias can also change model rankings outside a development split [23]. For collision avoidance, a false negative permits continued motion toward a barrier, whereas a false positive normally causes a conservative stop. This asymmetry makes glass recall, latency, threshold stability, and behavior under a deterministic safety layer essential complements to mean IoU and precision.

III. MATERIALS AND METHODS

A. PAIRED EXPERIMENTAL DESIGN

Each modality passed through the same sequence: paired data preparation, baseline training, one-factor-at-a-time hyperparameter refinement, checkpoint selection by validation mean IoU, a single final evaluation on the fixed test split, and robot deployment. The test split did not influence optimization or checkpoint selection. Input dimensionality and the associated first-layer weights were the only model differences.

B. DATASET AND PREPROCESSING

RGB-D GSD contains 3,009 paired RGB images, depth maps, and pixel-level glass masks [20]. The official 609-image test partition was retained. From the 2,400 training images, 360 images (15%) were reserved for validation, leaving 2,040 for parameter optimization. RGB images, depth maps, and masks were matched by identifier and resized to 512×512 pixels. RGB channels used the pretrained SegFormer normalization statistics; depth was resized and normalized independently. Masks encoded glass and background.

C. INPUT REPRESENTATIONS AND SEGFORMER MODIFICATION

The RGB model received a $3 \times H \times W$ tensor. The RGB-D model received a $4 \times H \times W$ tensor formed by concatenating normalized depth with RGB. The first SegFormer-B0 patch-embedding convolution was expanded from three to four input channels. The pretrained RGB kernels were retained, and the new depth kernel was initialized with their channel-wise mean. The remaining MiT-B0 encoder and multilayer-perceptron decoder were identical [12].

TABLE I
CONTROLLED EXPERIMENTAL CONFIGURATION

Item	Configuration
Dataset / split	RGB-D GSD: 2,040 train, 360 validation, 609 test
Input	512 × 512; RGB or four-channel RGB-D
Architecture	SegFormer-B0; early fusion at first patch embedding
Optimization	AdamW; 50 epochs maximum; batch size 2
Learning rate / decay	6×10^{-5} / 0.02
Warmup / patience	3 / 12 epochs
Numerics	Mixed precision; gradient clipping 1.0
Selection / seed	Validation mean IoU / 42

D. TRAINING AND MODEL SELECTION

Both models started from pretrained SegFormer-B0 weights. Baseline runs established the reference configuration; learning rate, weight decay, and batch size were then changed one factor at a time under the same search procedure. The best checkpoint in each run was selected by validation mean IoU, and the strongest cumulative configuration was frozen before test evaluation. Training and inference used an ASUS Vivobook 16X K3605V with an Intel Core i7-13700H, 16 GB RAM, and an NVIDIA GeForce RTX 4050 GPU with 6 GB memory. The software stack comprised Python, PyTorch, Hugging Face Transformers and Datasets, OpenCV, NumPy, Pandas, Pillow, Matplotlib, and scikit-learn.

E. EVALUATION METRICS

For glass as the positive class, TP, TN, FP, and FN denote true-positive, true-negative, false-positive, and false-negative pixels. Pixel accuracy (PA) is

$$PA = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

For C classes, class intersection over union and its class mean are

$$IoU_c = \frac{TP_c}{TP_c+FP_c+FN_c} \tag{2}$$

$$mIoU = \frac{IoU_1+\cdots+IoU_C}{C} \tag{3}$$

Glass-class precision, recall, and F1 are

$$Precision = \frac{TP}{TP+FP} \tag{4}$$

$$Recall = \frac{TP}{TP+FN} \tag{5}$$

$$F_1 = \frac{2\times Precision\times Recall}{Precision+Recall} \tag{6}$$

Mean inference time was measured on the same laptop in milliseconds per frame. Throughput is

$$FPS = \frac{1000}{t_{inf}\ (ms)} \tag{7}$$

F. ROBOT PLATFORM AND DECISION RULE

Practical validation used a Pioneer P3-DX differential-drive robot, a first-generation Microsoft Kinect for Windows, and the evaluation laptop. Inference ran on the laptop, and stop or continue commands were transmitted through the robot-control framework. The collision decision region was the full-height central third of the predicted mask. The numerator counts pixels labelled as glass inside that region, and the denominator is the region's total pixel count. Their ratio is

$$r_c = \frac{N_g}{N_c} \tag{8}$$

A glass-triggered stop was issued when the center ratio met or exceeded the configured threshold. A separate distance-based safety stop remained active in every run. The recorded experiment used thresholds of 0.30, 0.35, 0.40, and 0.60; 0.35 was the selected operating point because both models produced a glass stop after a usable approach distance. The 20-run record contains ten glass and ten non-glass runs, balanced across the two models. For each run, threshold, scene class, distance travelled, runtime, and final stop mechanism were recorded.

IV. BENCHMARK RESULTS

A. VALIDATION AND UNSEEN-TEST PERFORMANCE

RGB-D reached the higher validation mean IoU and F1, with its best checkpoint at epoch 34; RGB reached its best checkpoint at epoch 41. On the fixed test split, the ranking reversed. RGB led on pixel accuracy, mean IoU, glass IoU, recall, F1, latency, and throughput. RGB-D led only on glass precision, by 0.0130 absolute.

TABLE II
VALIDATION PERFORMANCE OF THE OPTIMIZED MODELS

Model	Best epoch	Mean IoU	F1 score
RGB	41	0.8613	0.8746

Model	Best epoch	Mean IoU	F1 score
RGB-D	34	0.8677	0.8806

TABLE III
COMPARATIVE PERFORMANCE ON THE 609-IMAGE UNSEEN TEST SET

Metric	RGB	RGB-D	Absolute advantage
Pixel accuracy	0.9448	0.9434	RGB +0.0014
Mean IoU	0.8234	0.8164	RGB +0.0070
Glass IoU	0.7108	0.6980	RGB +0.0128
Glass precision	0.8244	0.8374	RGB-D +0.0130
Glass recall	0.8376	0.8075	RGB +0.0301
Glass F1	0.8310	0.8222	RGB +0.0088
Inference time	10.24 ms	11.47 ms	RGB 1.23 ms lower
Throughput	97.63 fps	87.18 fps	RGB +10.45 fps

B. TEST-SET EFFECTS AND COMPUTATIONAL COST

Relative to RGB-D, RGB increased glass IoU by 1.8%, recall by 3.7%, and F1 by 1.1%. RGB-D precision was 1.6% higher. The recall difference is the most safety-relevant segmentation effect because it corresponds to fewer missed glass pixels. The agreement among glass IoU, recall, F1, and the deployment outcomes is stronger evidence than any one aggregate score in isolation.

RGB reduced mean inference latency by 1.23 ms per frame (10.7%) and increased throughput by 10.45 frames/s (12.0%). Both models met real-time operation on the reported hardware, but the RGB margin leaves more of the control cycle for capture, preprocessing, communication, and braking. Because timing was paired on the same computer, the direction and magnitude of this difference are directly comparable.

V. PRACTICAL ROBOT DEPLOYMENT

A. THRESHOLD SELECTION

Table IV reports exact outcomes rather than qualitative stopping descriptions. At a threshold of 0.30, both models produced a glass stop in 1/1 run. At 0.35, both produced a glass stop in 2/2 runs. At 0.40 and 0.60, RGB produced a glass stop in 1/1 run at each threshold, whereas RGB-D reached the distance safety stop first in both runs. Across the complete glass set, the semantic trigger therefore activated in 5/5 RGB runs and 3/5 RGB-D runs.

TABLE IV
EXACT DEPLOYMENT OUTCOMES BY CENTER-RATIO THRESHOLD

Center-ratio threshold	Glass stop, RGB	Glass stop, RGB-D	False glass stop, non-glass (RGB / RGB-D)
0.30	1/1	1/1	0/1 / 0/1
0.35	2/2	2/2	0/2 / 0/2
0.40	1/1	0/1 (safety stop)	0/1 / 0/1
0.60	1/1	0/1 (safety stop)	0/1 / 0/1

B. PRIMARY GLASS AND NON-GLASS TRIALS

At the selected threshold of 0.35, each model was evaluated in two recorded glass runs. All four ended through the semantic glass trigger before contact (Table V). Travel and runtime are retained as trial traces; they are not used to rank the models because starting geometry differed between runs. Across ten non-glass runs at thresholds from 0.30 to 0.60, neither model issued a glass command; motion ended through the independent distance safety stop. The platform, laboratory scenes, and terminal records for representative runs at the 0.35 threshold are shown in Fig. 1.

TABLE V
MEASURED GLASS RUNS AT THE SELECTED CENTER-RATIO THRESHOLD OF 0.35

Recorded run	Model	Distance travelled	Runtime / trigger
Glass run A	RGB	2.01 m	13.68 s / glass
Glass run A	RGB-D	2.27 m	14.97 s / glass
Glass run B	RGB	1.79 m	12.28 s / glass
Glass run B	RGB-D	1.34 m	9.70 s / glass

C. THRESHOLD ROBUSTNESS AND SYSTEM BEHAVIOR

The 20-run record separates semantic stopping from fallback protection. RGB generated the intended glass trigger in every glass run; RGB-D generated it in three runs and relied on the distance safety layer at the two stricter thresholds. The safety layer prevented contact in both cases. Intermittent RGB-D frame-read conflicts were visible in the file-based Kinect bridge, but they were transient and did not change the recorded stopping outcomes; this is an integration property of the bridge rather than evidence about segmentation quality.

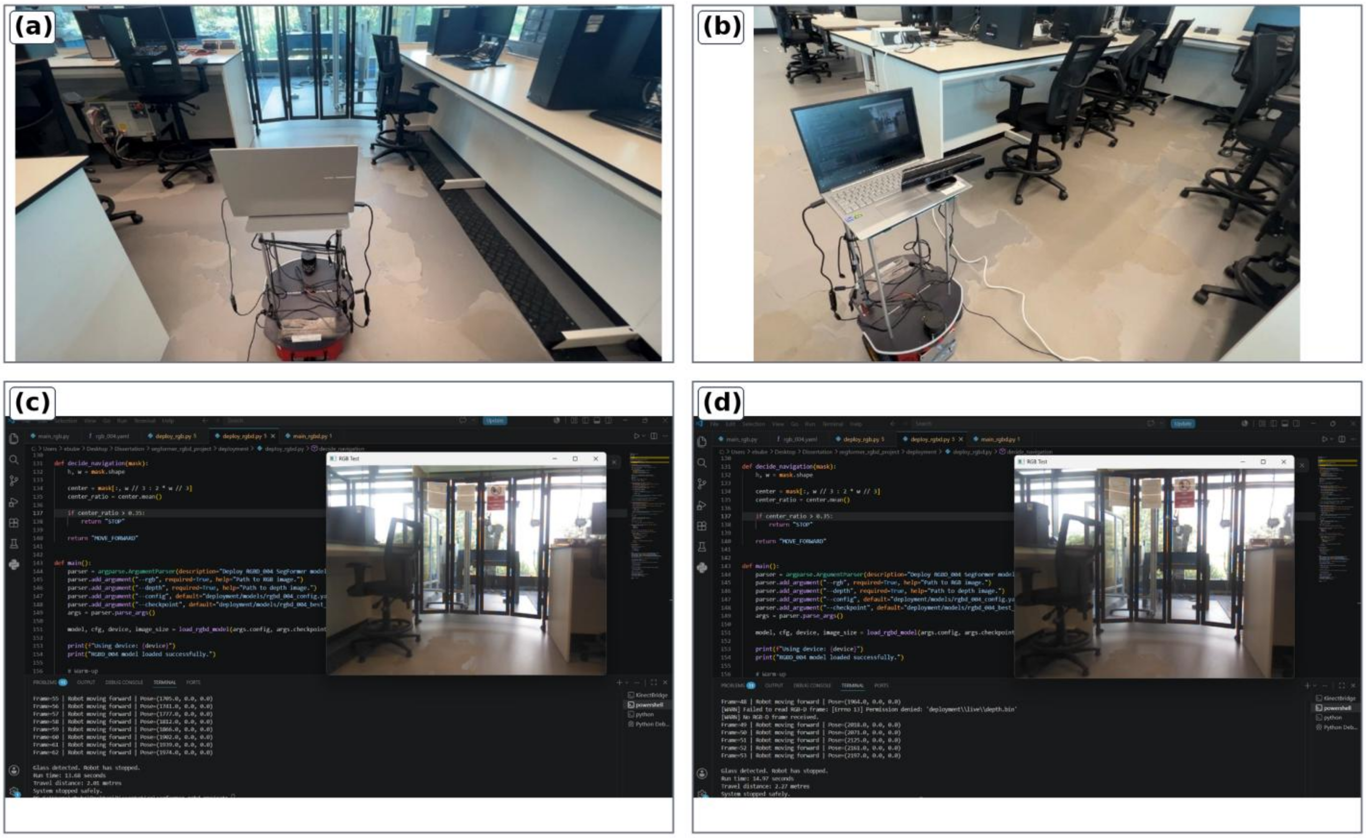

FIGURE 1. Laboratory deployment evidence: (a) and (b) Pioneer P3-DX, Kinect, laptop, and glass-partition scenes; (c) RGB glass-triggered stop at threshold 0.35 (13.68 s, 2.01 m); and (d) RGB-D glass-triggered stop at threshold 0.35 (14.97 s, 2.27 m).

VI. DISCUSSION

A. WHY RAW-DEPTH EARLY FUSION UNDERPERFORMED

Three mechanisms are consistent with the observed result. First, depth near glass is not a clean complementary signal: transmission, reflection, and invalid returns can encode the background or create holes [2], [3], [14], [17]-[20]. Second, first-layer fusion mixes RGB and depth before modality-specific features or confidence can be learned; the added kernel was initialized from RGB statistics rather than a depth-specific representation. Third, the validation-to-test reversal indicates that depth patterns useful on the validation subset were less stable on the fixed test split. The experiment establishes the performance outcome; it does not assign a unique causal weight to these mechanisms.

B. SAFETY-WEIGHTED INTERPRETATION

For collision prevention, a false negative has greater immediate physical cost than a conservative false positive. RGB's higher recall and its 5/5 semantic stops therefore support the same deployment choice. The two RGB-D safety-stop interventions also show why a learned segmentation trigger must not replace deterministic proximity protection. In this design, semantic glass perception extends obstacle awareness; the independent range rule remains the final protective layer.

C. BENCHMARK-TO-ROBOT TRANSLATION

The selected-threshold trials show that both frozen models can produce useful real-time robot behavior. The threshold sweep adds information that the test table cannot provide: models with close IoU values can differ in the spatial evidence delivered to a navigation rule. Reporting the decision region, threshold, fallback mechanism, and exact trigger counts makes this translation auditable.

D. DEPLOYMENT DECISION

For the evaluated SegFormer-B0 design and hardware, RGB is the preferred configuration: it leads on seven of eight test measures, reduces latency by 10.7%, and activates the semantic stop in 5/5 glass runs. Raw depth adds sensor and integration cost without an accuracy or robustness gain. Depth remains a credible input when paired with validity masks, confidence-aware completion, modality-specific encoders, or intermediate and late fusion; those designs are distinct from the early-fusion configuration tested here.

VII. SCOPE OF CLAIMS

The evidence supports a paired comparison of two fixed SegFormer-B0 configurations on RGB-D GSD and a 20-run indoor deployment. Training used one train/validation split and random seed 42; numerical gaps are therefore reported as observed effects rather than population-level significance. Robot counts apply to the tested laboratory, Pioneer P3-DX, Kinect, approach speed, decision region, and thresholds. These boundaries are retained because they materially determine the safety interpretation.

VIII. CONCLUSION

Under paired data, architecture, optimization, hardware, and decision logic, RGB outperformed four-channel early-fusion RGB-D on seven of eight test measures. RGB achieved higher mean IoU, glass IoU, recall, and F1 while reducing latency by 10.7% and increasing throughput by 12.0%. The robot experiment reproduced this ordering at the decision level: RGB generated a glass trigger in 5/5 glass runs, RGB-D in 3/5, and neither model generated a false glass trigger in ten non-glass runs. The central result is specific and actionable: appending raw depth at the first SegFormer embedding did not improve glass segmentation and reduced threshold robustness. For this platform and fusion design, RGB is the stronger default; depth should be included only through a reliability-aware representation.

IX. DATA, CODE, AND MATERIALS AVAILABILITY

The RGB-D GSD benchmark is publicly available from the RGB-D GSD dataset repository. The complete reproducibility package includes the source code, environment specification, trained RGB and RGB-D checkpoints, benchmark outputs, the complete 20-run record, and deployment videos. Project repository: RGB-and-RGB-D-Glass-Surface-Detection-For-Safe-Pioneer-Robot-Navigation.

X. ACKNOWLEDGMENT

The authors thank the technical staff of the School of Science, Engineering and Environment, University of Salford, for access to the robotics laboratory and Pioneer platform. OpenAI ChatGPT was used to assist with language editing, structural refinement, and document formatting of author-prepared scientific content. It was not used to produce experimental data, source code, equations, figures, or numerical results. The authors verified the references, analyses, and conclusions and accept full responsibility for the final manuscript.

XI. CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] G.-Z. Yang et al., "The grand challenges of Science Robotics," Sci. Robot., vol. 3, no. 14, Art. no. eaar7650, 2018, doi: 10.1126/scirobotics.aar7650.

[2] S. Ameen and H. S. Baskaran, "Beyond RGB-D: Perception of glass, mirrors, and see-through scenes for robotics," engrXiv, preprint, May 2026, doi: 10.31224/7048.

[3] J. Jiang, G. Cao, J. Deng, T.-T. Do, and S. Luo, "Robotic perception of transparent objects: A review," IEEE Trans. Artif. Intell., vol. 5, no. 6, pp. 2547-2567, Jun. 2024, doi: 10.1109/TAI.2023.3326120.

[4] S. A. Shafer, "Using color to separate reflection components," Color Res. Appl., vol. 10, no. 4, pp. 210-218, 1985, doi: 10.1002/col.5080100409.

[5] G. J. Klinker, S. A. Shafer, and T. Kanade, "A physical approach to color image understanding," Int. J. Comput. Vis., vol. 4, no. 1, pp. 7-38, 1990, doi: 10.1007/BF00137441.

[6] H. Mei et al., "Don't hit me! Glass detection in real-world scenes," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2020, pp. 3684-3693, doi: 10.1109/CVPR42600.2020.00374.

[7] J. Lin, Z. He, and R. W. H. Lau, "Rich context aggregation with reflection prior for glass surface detection," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2021, pp. 13415-13424, doi: 10.1109/CVPR46437.2021.01321.

[8] J. Lin, Y.-H. Yeung, and R. W. H. Lau, "Exploiting semantic relations for glass surface detection," in Adv. Neural Inf. Process. Syst., vol. 35, 2022, pp. 22490-22504.

[9] H. Mei et al., "Glass segmentation using intensity and spectral polarization cues," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2022, pp. 12612-12621, doi: 10.1109/CVPR52688.2022.01229.

[10] A. Kalra, V. Taamazyan, S. K. Rao, K. Venkataraman, R. Raskar, and A. Kadambi, "Deep polarization cues for transparent object segmentation," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2020, pp. 8602-8611, doi: 10.1109/CVPR42600.2020.00863.

[11] E. Xie, W. Wang, W. Wang, M. Ding, C. Shen, and P. Luo, "Segmenting transparent objects in the wild," in Computer Vision - ECCV 2020, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds. Cham, Switzerland: Springer, 2020, pp. 696-711, doi: 10.1007/978-3-030-58601-0_41.

[12] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: Simple and efficient design for semantic segmentation with transformers," in Adv. Neural Inf. Process. Syst., vol. 34, 2021, pp. 12077-12090.

[13] T. Wang, X. He, and N. Barnes, "Glass object localization by joint inference of boundary and depth," in Proc. 21st Int. Conf. Pattern Recognit. (ICPR), 2012, pp. 3783-3786.

[14] S. S. Sajjan, M. Moore, M. Pan, G. Nagaraja, J. Lee, A. Zeng, and S. Song, "ClearGrasp: 3D shape estimation of transparent objects for manipulation," in Proc. IEEE Int. Conf. Robot. Autom. (ICRA), 2020, pp. 3634-3642, doi: 10.1109/ICRA40945.2020.9197518.

[15] H. Fang, H.-S. Fang, S. Xu, and C. Lu, "TransCG: A large-scale real-world dataset for transparent object depth completion and a grasping baseline," IEEE Robot. Autom. Lett., vol. 7, no. 3, pp. 7383-7390, Jul. 2022, doi: 10.1109/LRA.2022.3183256.

[16] X. Chen, H. Zhang, Z. Yu, A. Opipari, and O. C. Jenkins, "ClearPose: Large-scale transparent object dataset and benchmark," in Computer Vision - ECCV 2022. Cham, Switzerland: Springer, 2022, pp. 381-396, doi: 10.1007/978-3-031-20074-8_22.

[17] Y. Liang, B. Deng, W. Liu, J. Qin, and S. He, "Monocular depth estimation for glass walls with context: A new dataset and method," IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 12, pp. 15081-15097, Dec. 2023, doi: 10.1109/TPAMI.2023.3308551.

[18] A. Costanzino, P. Z. Ramirez, M. Poggi, F. Tosi, S. Mattoccia, and L. Di Stefano, "Learning depth estimation for transparent and mirror surfaces," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2023, pp. 9210-9221, doi: 10.1109/ICCV51070.2023.00848.

[19] P. Z. Ramirez, A. Costanzino, F. Tosi, M. Poggi, S. Salti, S. Mattoccia, and L. Di Stefano, "Booster: A benchmark for depth from images of specular and transparent surfaces," IEEE Trans. Pattern Anal. Mach. Intell., vol. 46, no. 1, pp. 85-102, Jan. 2024, doi: 10.1109/TPAMI.2023.3323858.

[20] J. Lin, Y.-H. Yeung, S. Ye, and R. W. H. Lau, "Leveraging RGB-D data with cross-modal context mining for glass surface detection," Proc. AAAI Conf. Artif. Intell., vol. 39, no. 5, pp. 5254-5261, 2025, doi: 10.1609/aaai.v39i5.32558.

[21] R. Singh and K. S. Nagla, "Multi-data sensor fusion framework to detect transparent object for the efficient mobile robot mapping," Int. J. Intell. Unmanned Syst., vol. 7, no. 1, pp. 2-18, 2019, doi: 10.1108/IJIUS-05-2018-0013.

[22] Y. Tao, H. Gao, Y. Wen, L. Duan, and J. Lan, "Glass recognition and map optimization method for mobile robot based on boundary guidance," Chin. J. Mech. Eng., vol. 36, Art. no. 74, 2023, doi: 10.1186/s10033-023-00902-9.

[23] A. Torralba and A. A. Efros, "Unbiased look at dataset bias," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2011, pp. 1521-1528, doi: 10.1109/CVPR.2011.5995347.

EBUBECHUKWU AMASIATU is an M.Sc. candidate in Robotics and Automation with the School of Science, Engineering and Environment, University of Salford, Manchester, United Kingdom. His work focuses on mobile robotics, autonomous navigation, computer vision, embedded systems, and the practical integration of learning-based perception with robot safety. He developed and deployed the RGB and RGB-D glass-detection systems evaluated in this article on a Pioneer P3-DX mobile robot.

SALEM AMEEN received the Ph.D. degree from the University of Salford, Manchester, United Kingdom, in 2017. He is a Lecturer with the School of Science, Engineering and Environment, University of Salford. His research interests include artificial intelligence, mobile robotics, computer vision, efficient deep learning, transformer architectures, multimodal robot perception, autonomous systems, and the perception of transparent and reflective surfaces.