COMS W 4995 Deep Learning

for

Computer Vision

Team Members:

Jiakai Xu (ax2155)
Xingchen Sha (xs2472)
Xinjie Zhang (xz3236)

Introduction to Glass Surface Detection

Glass surfaces, including glass windows / walls / doors, pervade our everyday lives.

Significantly impacts various

computer vision tasks:

- Depth estimation,
- 3D scene understanding
- Vision-language navigation
- Image captioning
- ……

Challenges in Detecting Glass Surfaces

1. Smooth Surface Reflection.

- Exists high-resolution reflections as mirrors
- Scenery with semantic context is reflected.
- False-positive predictions of glass surface presence within mirror regions.

Challenges in Detecting Glass Surfaces

2. Diverse Glass Patterns

- Variability in transparency, ranging from fully transparent to semi-opaque
- Influence of environment creating vastly different appearances
- Non-Uniform structures such as scratches, smudges, or coatings

Related Works

- Matterport3D: Learning from RGB-D Data in Indoor Environments
- Glass Segmentation using Intensity and Spectral Polarization Cues
- SegFormer: simple and efficient design for semantic segmentation with transformers
- Context-Aware Domain Adaptation in Semantic Segmentation

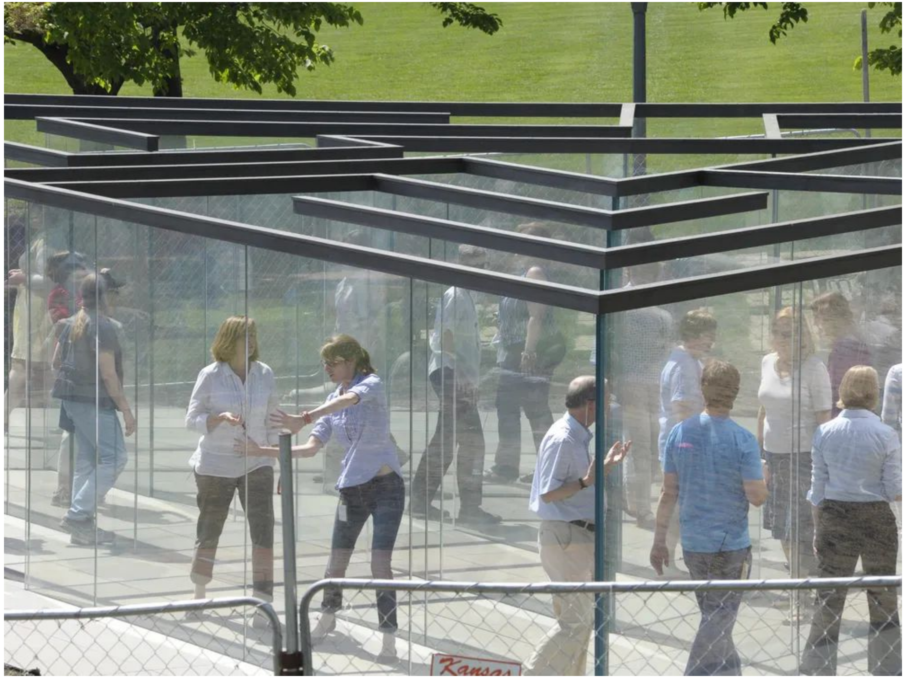
Angel Chang et al., "Matterport3D: Learning from RGB-D Data in Indoor Environments," *In Proceedings of 2017 International Conference on 3D Vision*, Qingdao, China, 2017, pp. 667-676, doi: 10.1109/3DV.2017.00081

Enze Xie et al., "SegFormer: simple and efficient design for semantic segmentation with transformers," *In Proceedings of the 35th International Conference on Neural Information Processing Systems*. Curran Associates Inc., Red Hook, NY, USA, Article 924, pp. 12077–12090.

Haiyang Mei et al., "Glass Segmentation using Intensity and Spectral Polarization Cues," *In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, LA, USA, 2022, pp. 12612-12621, doi: 10.1109/CVPR52688.2022.01229.

Jinyu Yang et al., "Context-Aware Domain Adaptation in Semantic Segmentation," *In Proceedings of the* 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2021, pp. 514-524, doi: 10.1109/WACV48630.2021.00056.

Motivations & Intuitions



https://www.smithsonianmag.com/smart-news/visit-glass-labyrinth-kansas-city-180951600/

Motivations & Intuitions

Exploring Semantic Relationships Between Glass Surfaces and Environment

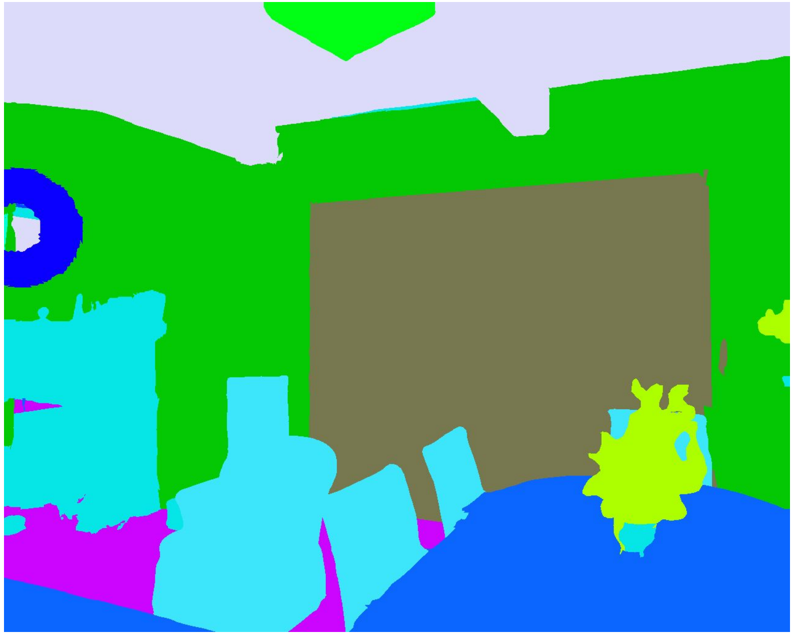
For instance, glass windows are more likely to co-occur with walls and curtains due to their strong semantic relevance, whereas their association with objects like cars and trees is comparatively weaker.

How to leverage this observation to design an efficient glass detection model?

Contextual Learning

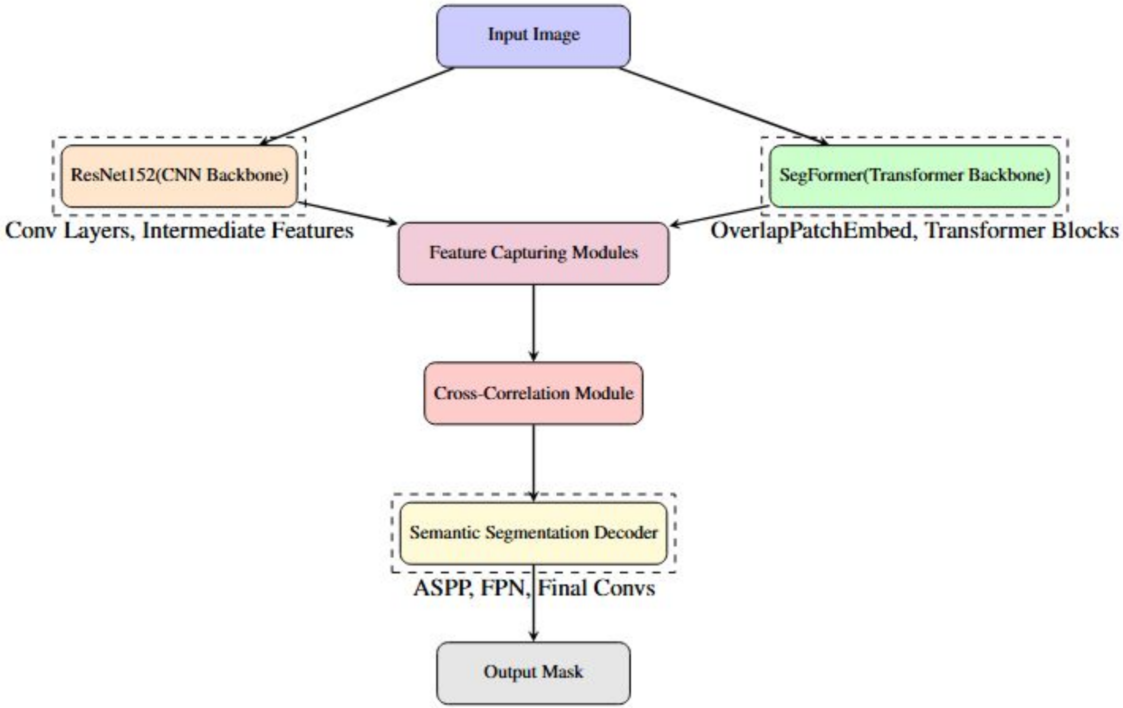
Learn the contextual correlations among objects.

Original Image → **Semantic Context**



GlassMindNet Model Architecture

3 MODEL ARCHITECTURE



GlassMindNet: Model Architecture Part 1: Backbone Architecture

1. Segmentation Transformer Backbone (SegFormer)

- **Purpose:** Extracts high-level semantic features from the input image.
- **OverlapPatchEmbed:** Converts the input image into overlapping patch embeddings.
- **Transformer Blocks:** Consist of multi-head attention mechanisms and MLPs for capturing long-range dependencies.
- **Feature Outputs:** Provides multi-scale feature maps for different stages of processing.

2. ResNet152 Backbone:

- **Purpose:** Captures low-level spatial features essential for precise boundary detection.
- **Convolutional Layers:** Standard ResNet152 layers adapted for feature extraction.
- **Layer Outputs:** Intermediate feature maps from different ResNet stages.

GlassMindNet: Model Architecture Part 2: Feature Fusion Modules

1. Feature Capturing Modules

- **Purpose:** Processes and aligns features from both backbones to ensure compatibility and enhance feature richness.
- **Projection Layers:** Align feature dimensions between SegFormer and ResNet152 outputs.
- **Attention Mechanisms:** Emphasize relevant features and suppress irrelevant ones.

2. Cross-Correlation Module

- **Purpose:** Integrates semantic and spatial features by modeling their relationships, enhancing the network's ability to discern glass surfaces amidst complex backgrounds
- **Multi-Head Attention:** Facilitates the interaction between semantic queries and spatial keys/values.
- **Normalization Layers:** Ensure stable and efficient training dynamics.
- **Feed-Forward Networks:** Further process the integrated features for downstream tasks.

GlassMindNet Model Architecture Part 3: Semantic Segmentation Decoder

1. **Purpose:**

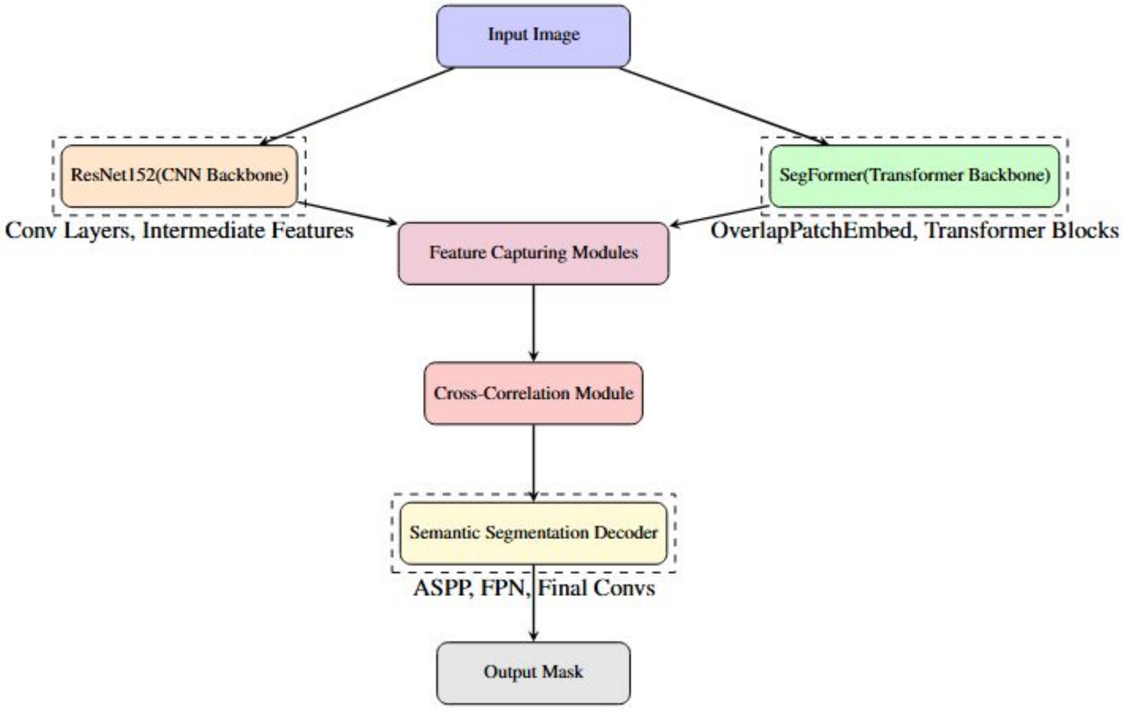
   - Transforms the integrated features into precise segmentation masks indicating glass surfaces.

2. **Components:**

   - **Atrous Spatial Pyramid Pooling (SPP):** Captures multi-scale context by applying parallel atrous convolutions with different dilation rates.
   - **Feature Pyramid Networks (FPN):** Combines features from different scales to refine segmentation outputs.
   - **Final Convolutional Layers:** Produce the binary mask indicating glass regions

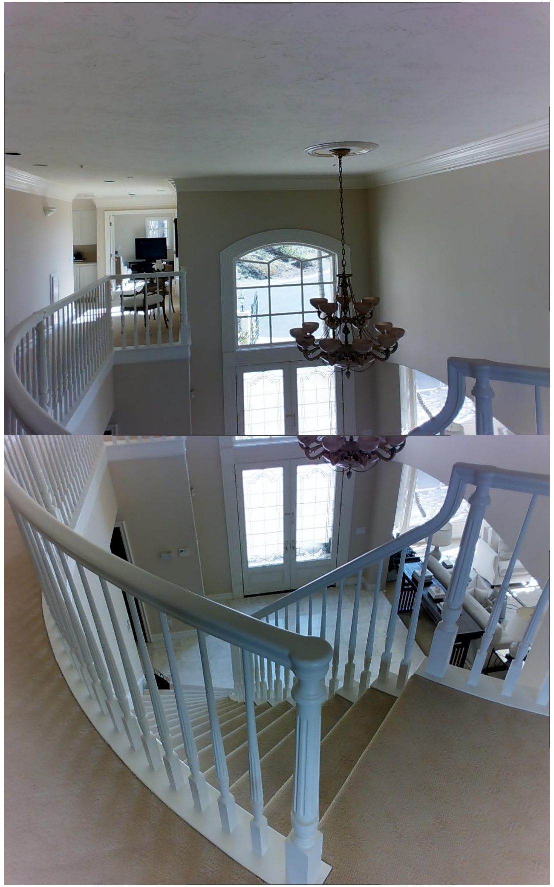
GlassMindNet Model Architecture

3 Model Architecture

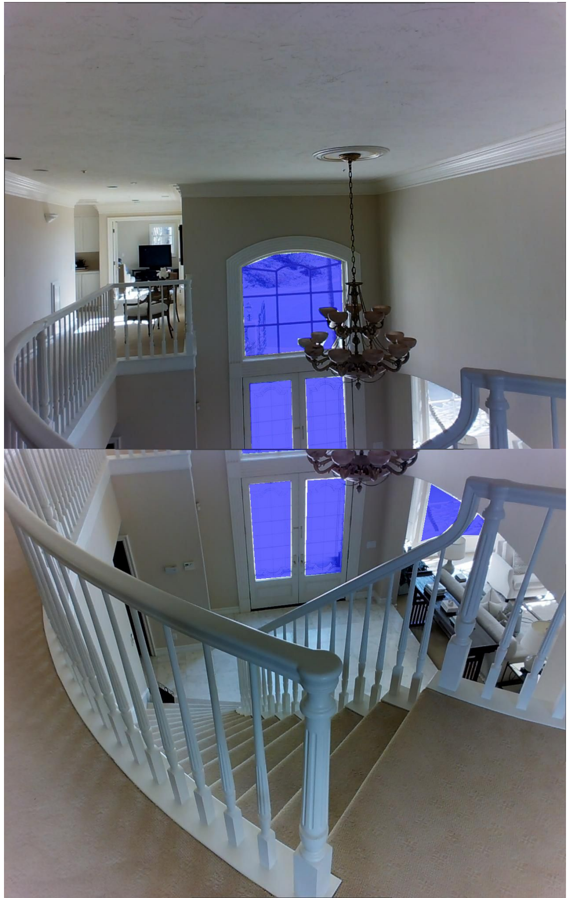


Evaluation & Visualization

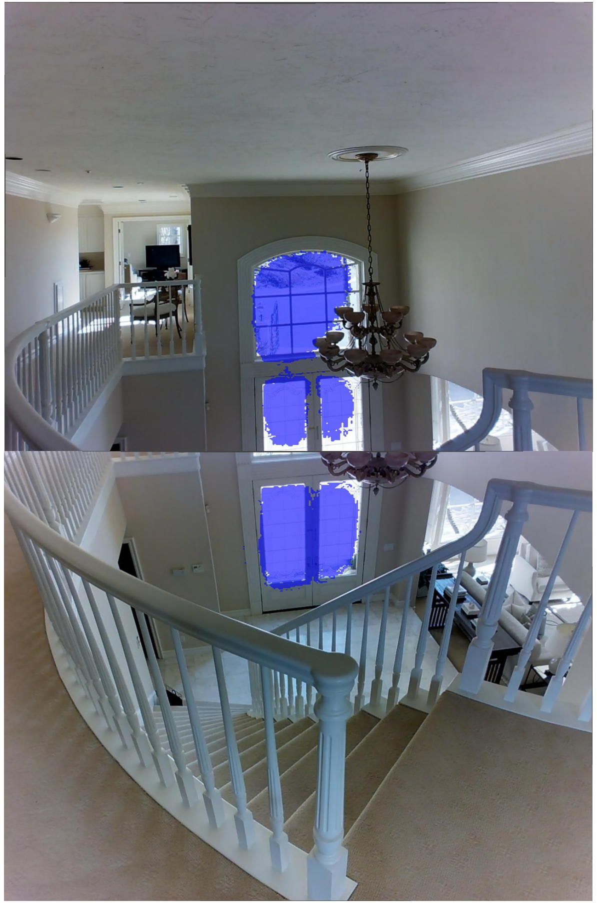
Original Image



Ground Truth



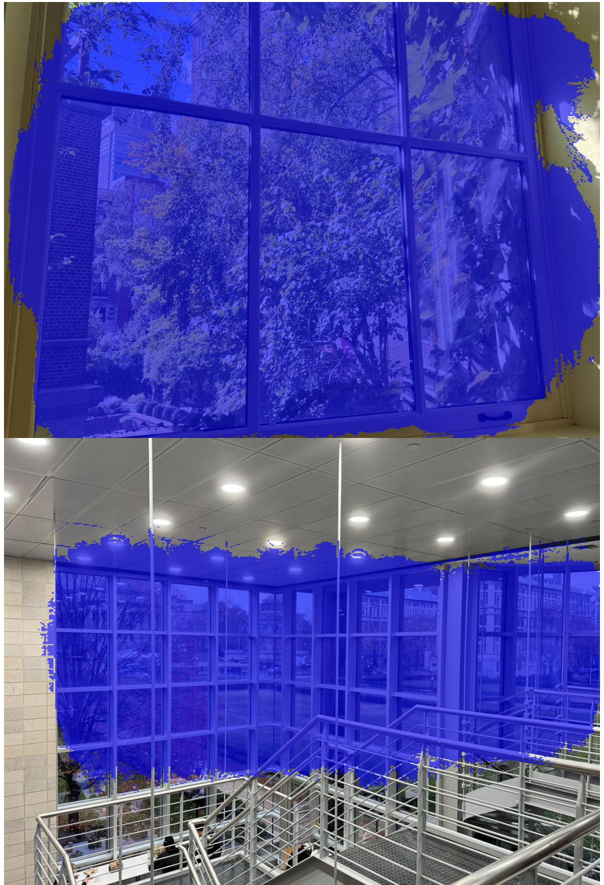
Prediction



Evaluation & Visualization

Original Image

Prediction



Evaluation & Visualization

Intersection over union (IoU), Mean Absolute Error (MAE), maximum F-measure (Fβ), and balance error rate (BER).

Method	Venue	IoU↑	F_β ˆ	MAE↓	BER↓
PSPNet *Zhaoet al.* (2017)	CVPR 2017	0.560	0.679	0.093	13.40
DeepLabV3+ *Chenet al.* (2018)	CVPR 2018	0.557	0.671	0.100	13.11
PSANet *Zhaoet al.* (2018)	ECCV 2018	0.550	0.656	0.104	12.61
DANet *Fuet al.* (2019)	CVPR 2019	0.543	0.673	0.098	13.84
SCA-SOD *Siriset al.* (2021)	ICCV 2021	0.558	0.689	0.087	15.03
SETR *Zhenget al.* (2021)	CVPR 2021	0.567	0.679	0.086	13.25
Segmenter *Strudelet al.* (2021)	ICCV 2021	0.536	0.645	0.140	14.02
Swin *Liuet al.* (2021)	ICCV 2021	0.596	0.702	0.082	11.34
ViT *Yuanet al.* (2021)	ICLR 2021	0.562	0.693	0.087	14.72
SegFormer *Xieet al.* (2021*b*)	NeurIPS 2021	0.547	0.683	0.094	15.15
Twins *Chuet al.* (2021)	NeurIPS 2021	0.590	0.703	0.084	12.43
GDNet *Linet al.* (2021)	CVPR 2020	0.529	0.642	0.101	18.17
GlassNet *Meiet al.* (2020)	CVPR 2021	0.721	0.821	0.061	10.02
GlassMindNet (Ours)	-	**0.8917**	**0.9479**	**0.0372**	**6.41**

Table 1: Comparison of **GlassMindNet** with state-of-the-art methods on the glass surface detection task.

Limitations & Future Work

How to avoid these mistakes?

Answers:

Better Backbone Models (e.g., ViT)

Improve the Robustness (Batch size)

Thank you!

Contact us:
ax2155@columbia.edu,
xs2472@columbia.edu,
xz3236@columbia.edu