Learning Semantic Associations for Mirror Detection

香港城市大學
City University of Hong Kong

Huankang Guan, Jiaying Lin, Rynson W.H. Lau

CVPR JUNE 19-24 2022 NEW ORLEANS • LOUISIANA

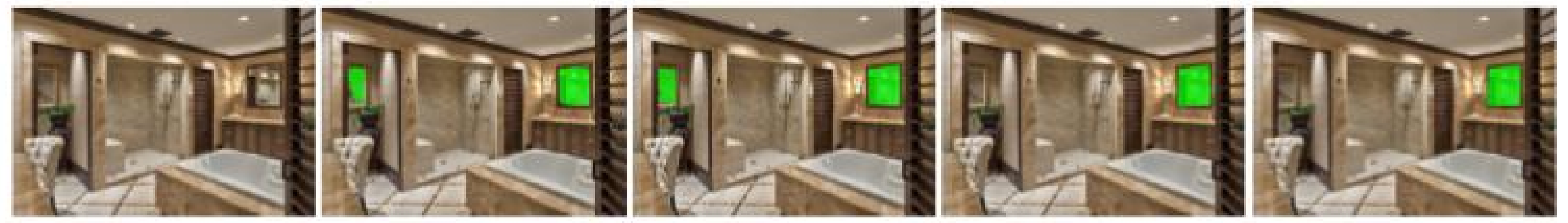
Introduction

The presence of mirrors can affect many vision tasks, such as depth predition and object detection, as mirrors show deceptive picture by reflecting the surroundings. Due to the absence of a consistent visual pattern, mirror detection is challenging. In this work, we propose to explore the semantic associations between the different scene objects for a more reliable mirror detection.

Motivation

We observe that humans tend to place mirrors in relation to certain objects for specific functional purposes. For example, we usually put a mirror above the sink in the washroom to allow people to check their look after washing faces or hands. This aligns with studies in cognitive neuroscience that the surrounding objects can provide a complementary and effective source of contextual information, helping the visual system to locate the target objects.

Example of a **mirror** above the **sink** in the **bathroom**:

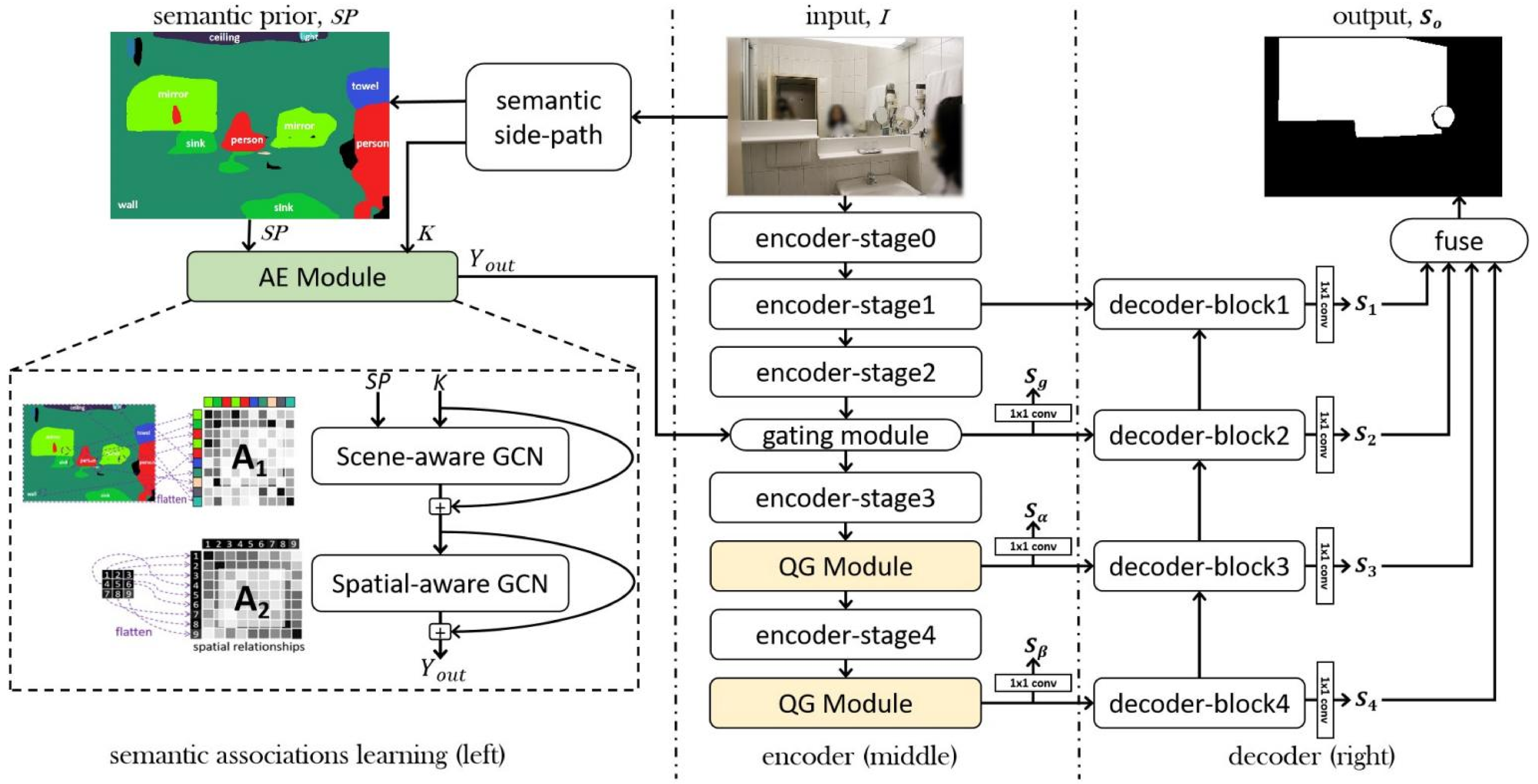


RGB Input MirrorNet PMD-Net Ours GT

Methodology

Semantic Side-Path is designed to performs pixel-level semantic segmentation task to discover the surrounding objects and capture the class-specific knowledge.

The Proposed model



AE Module consists of a scene-aware GCN and a spatial-aware GCN.

QG Module can facilitate diffusion and aggregation of semantic association knowledge by introducing self-attention and reversed attention mechanism.

The adjacient matrix of the scene-aware GCN (A1):

$$\mathbf{SP} \in \mathbb{R}^{H \times W \times C} \mapsto \mathbf{SP} \in \mathbb{R}^{HW \times C},$$
$$\mathbf{K} \in \mathbb{R}^{H \times W \times D} \mapsto \mathbf{K} \in \mathbb{R}^{HW \times D},$$
$$\Sigma = W_1 \, \mathbf{K} \, W_2,$$
$$A_1 = \mathbf{SP} \, \Sigma \, \mathbf{SP}^T,$$

The adjacient matrix of the spatial-aware GCN (A2):

$$(A_2)_{ij} = distance(loc_i, loc_j).$$

Ablation Study:

Method	encoder	decoder	ssp	AEM	QGM1	QGM2	Spat. GCNs	freeze ssp	f-measure↑	Accuracy↑
ID1	ResNeXt101	✓							0.8129	96.52
ID2	ResNeXt101	✓	✓					✓	0.8244	96.51
ID3	ResNeXt101	✓	✓	✓			✓	✓	0.8284	96.91
ID4	ResNeXt101	✓	✓	✓	✓		✓	✓	0.8368	96.72
ID5	ResNeXt101	✓	✓		✓	✓	✓	✓	0.8398	96.75
ID6	ResNeXt101	✓	✓	✓	✓	✓		✓	0.8297	96.67
ID7	ResNeXt101	✓	✓	✓	✓	✓	✓		0.8373	96.78
Ours	ResNeXt101	✓	✓	✓	✓	✓	✓	✓	0.8437	96.82

Quantitative Comparison

Table 1. Quantitative comparison between our method and ten related methods. We report max f-measure, IoU, accuracy and MAE. Best results are marked in red, and the second best results are marked in blue.

Method	PMD Dataset [24]				MSD Dataset [47]			
	f-measure↑	IoU↑	Accuracy↑	MAE↓	f-measure↑	IoU↑	Accuracy↑	MAE↓
GCPANet [11]	0.7548	58.01	95.59	0.04428	0.8477	74.76	93.11	0.06929
EGNet [55]	0.7987	60.17	96.34	0.03676	0.8238	66.68	91.54	0.08479
BDRAR [59]	0.7433	58.43	95.66	0.04346	0.8619	75.37	93.50	0.06510
DSC [17]	0.7548	59.81	95.65	0.04372	0.8479	75.36	92.82	0.07206
CPNet [48]	0.7342	56.36	94.85	0.05175	0.8314	69.86	92.44	0.07603
GloRe [10]	0.7743	61.25	95.61	0.04411	0.8600	76.10	93.07	0.06957
PSPNet [54]	0.8057	60.44	96.13	0.03920	0.8459	67.99	92.19	0.07875
DeepLabv3+ [9]	0.8096	64.08	96.43	0.04001	0.8750	77.48	94.13	0.05932
MirrorNet [47]	0.7775	62.50	96.27	0.04101	0.8597	77.41	92.75	0.07257
PMD-Net [24]	0.8276	62.40	96.80	0.08782	0.8691	76.88	93.94	0.06130
Ours	0.8437	66.84	96.82	0.04935	0.8887	79.85	94.63	0.05421

Code is Available

Table 4. **Comparison on the RGBD dataset [30].** F^{ω}_{β}: weighted f-measure score [29]. **BER**: Balanced Error Rate. †: using depth maps during both training and inference. *: results are reported by [30]. Best performances are in red.

Method	IoU↑	F^{ω}_{β} ↑	MAE↓	BER↓
MirrorNet [47]	68.37*	0.723*	0.062*	8.66*
PMD [24]	72.27*	0.775*	0.054*	10.71*
PDNet [30]	73.57*	0.783*	0.053*	9.26*
Ours	74.99	0.800	0.048	10.56
PDNet† [30]	77.77*	0.825*	0.042*	7.77*
Ours†	78.43	0.834	0.041	8.16

Visual Comparison

RGB Input MirrorNet PMD-Net Ours GT

Contributions

- We present a novel mirror detection approach that learns the semantic associations between mirrors and their surrounding objects for reliable mirror detection.
- We propose a novel Associations Exploration (AE) Module to infer scene object associations with fully connected graph models, and a novel Quadruple Graph (QG) Module to facilitate the diffusion and aggregation of semantic association knowledge using graph convolutions. We verify the effectiveness of the proposed modules with comprehensive studies.
- Extensive experiments show that our method outperforms existing state-of-the-art methods both quantitatively and qualitatively on the two popular mirror detection benchmarks, PMD and MSD.

Learning Semantic Associations for Mirror Detection

Huankang Guan, Jiaying Lin, Rynson W.H. Lau

香港城市大學 City University of Hong Kong

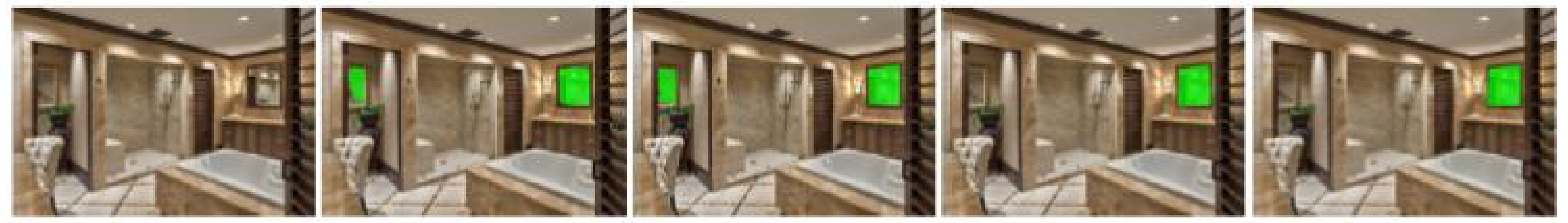
Introduction

The presence of mirrors can affect many vision tasks, such as depth predition and object detection, as mirrors show deceptive picture by reflecting the surroundings. Due to the absence of a consistent visual pattern, mirror detection is challenging. In this work, we propose to explore the semantic associations between the different scene objects for a more reliable mirror detection.

Motivation

We observe that humans tend to place mirrors in relation to certain objects for specific functional purposes. For example, we usually put a mirror above the sink in the washroom to allow people to check their look after washing faces or hands. This aligns with studies in cognitive neuroscience that the surrounding objects can provide a complementary and effective source of contextual information, helping the visual system to locate the target objects.

Example of a **mirror** above the **sink** in the **bathroom**:

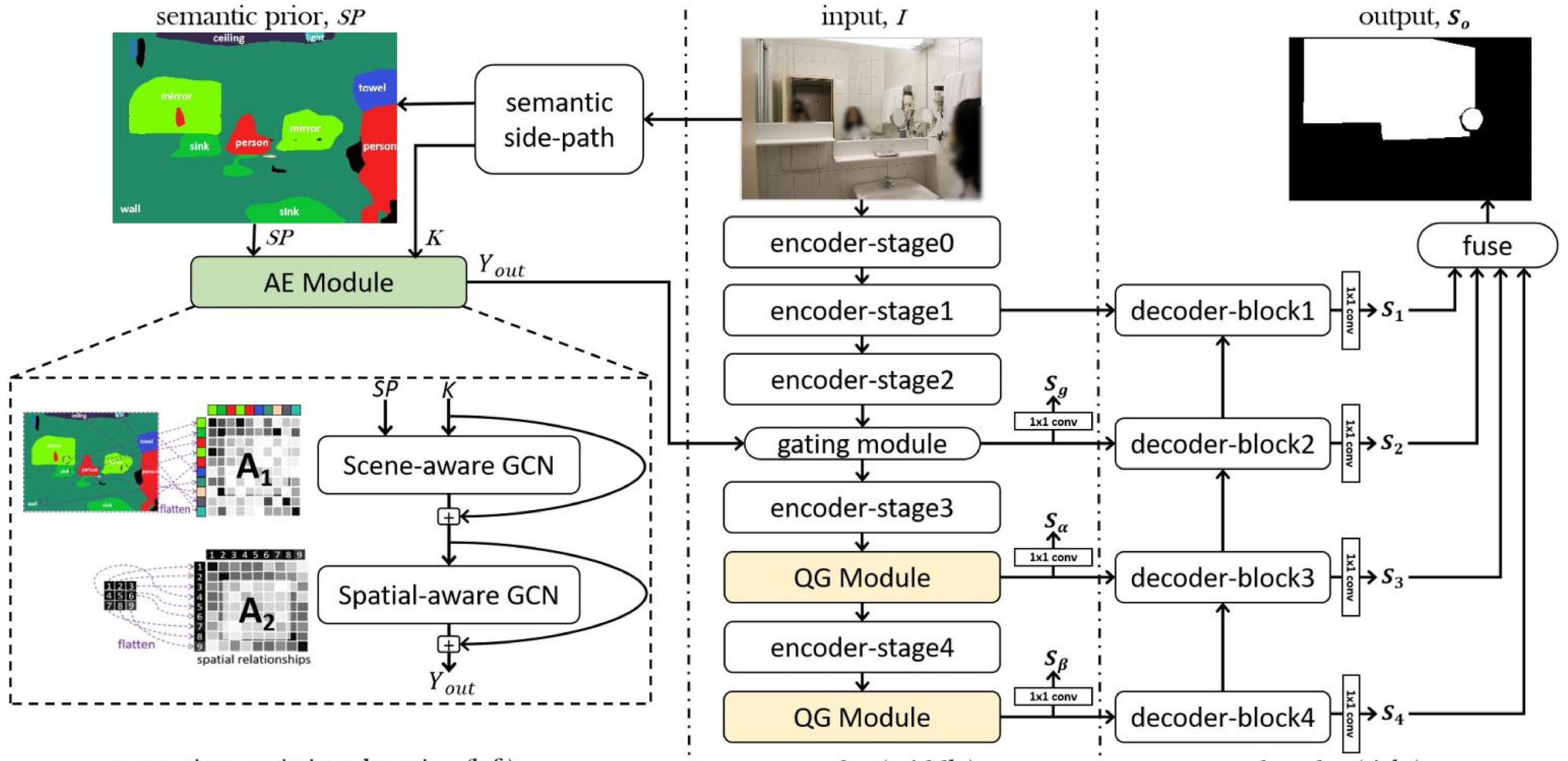


RGB Input MirrorNet PMD-Net Ours GT

Methodology

Semantic Side-Path is designed to performs pixel-level semantic segmentation task to discover the surrounding objects and capture the class-specific knowledge.

The Proposed model



AE Module consists of a scene-aware GCN and a spatial-aware GCN.

QG Module can facilitate diffusion and aggregation of semantic association knowledge by introducing self-attention and reversed attention mechanism.

The adjacent matrix of the scene-aware GCN (A1):

$$\mathbf{SP} \in \mathbb{R}^{H\times W\times C} \mapsto \mathbf{SP} \in \mathbb{R}^{HW\times C},$$
$$\mathbf{K} \in \mathbb{R}^{H\times W\times D} \mapsto \mathbf{K} \in \mathbb{R}^{HW\times D},$$
$$\Sigma = W_1\,\mathbf{K}\,W_2,$$
$$A_1 = \mathbf{SP}\,\Sigma\,\mathbf{SP}^T,$$

The adjacent matrix of the spatial-aware GCN (A2):

$$(A_2)_{ij} = distance(loc_i, loc_j).$$

Ablation Study:

Method	encoder	decoder	ssp	AEM	QGM1	QGM2	Spat. GCNs	freeze ssp	f-measure↑	Accuracy↑
ID1	ResNeXt101	✓							0.8129	96.52
ID2	ResNeXt101	✓	✓					✓	0.8244	96.51
ID3	ResNeXt101	✓	✓	✓			✓	✓	0.8284	96.91
ID4	ResNeXt101	✓	✓	✓	✓		✓	✓	0.8368	96.72
ID5	ResNeXt101	✓	✓		✓	✓	✓	✓	0.8398	96.75
ID6	ResNeXt101	✓	✓	✓	✓	✓		✓	0.8297	96.67
ID7	ResNeXt101	✓	✓	✓	✓	✓	✓		0.8373	96.78
Ours	ResNeXt101	✓	✓	✓	✓	✓	✓	✓	0.8437	96.82

Quantitative Comparison

Table 1. Quantitative comparison between our method and ten related methods. We report max f-measure, IoU, accuracy and MAE. Best results are marked in red, and the second best results are marked in blue.

Method	PMD Dataset [24]				MSD Dataset [47]			
	f-measure↑	IoU↑	Accuracy↑	MAE↓	f-measure↑	IoU↑	Accuracy↑	MAE↓
GCPANet [11]	0.7548	58.01	95.59	0.04428	0.8477	74.76	93.11	0.06929
EGNet [55]	0.7987	60.17	96.34	0.03676	0.8238	66.68	91.54	0.08479
BDRAR [59]	0.7433	58.43	95.66	0.04346	0.8619	75.37	93.50	0.06510
DSC [17]	0.7548	59.81	95.65	0.04372	0.8479	75.36	92.82	0.07206
CPNet [48]	0.7342	56.36	94.85	0.05175	0.8314	69.86	92.44	0.07603
GloRe [10]	0.7743	61.25	95.61	0.04411	0.8600	76.10	93.07	0.06957
PSPNet [54]	0.8057	60.44	96.13	0.03920	0.8459	67.99	92.19	0.07875
DeepLabv3+ [9]	0.8096	64.08	96.43	0.04001	0.8750	77.48	94.13	0.05932
MirrorNet [47]	0.7775	62.50	96.27	0.04101	0.8597	77.41	92.75	0.07257
PMD-Net [24]	0.8276	62.40	96.80	0.08782	0.8691	76.88	93.94	0.06130
Ours	0.8437	66.84	96.82	0.04935	0.8887	79.85	94.63	0.05421

Code is Available

Table 4. **Comparison on the RGBD dataset [30].** F_β^ω: weighted f-measure score [29]. **BER**: Balanced Error Rate. †: using depth maps during both training and inference. *: results are reported by [30]. Best performances are in red.

Method	IoU↑	F_β^ω ↑	MAE↓	BER↓
MirrorNet [47]	68.37*	0.723*	0.062*	8.66*
PMD [24]	72.27*	0.775*	0.054*	10.71*
PDNet [30]	73.57*	0.783*	0.053*	9.26*
Ours	74.99	0.800	0.048	10.56
PDNet† [30]	77.77*	0.825*	0.042*	7.77*
Ours†	78.43	0.834	0.041	8.16

Contributions

- We present a novel mirror detection approach that learns the semantic associations between mirrors and their surrounding objects for reliable mirror detection.
- We propose a novel Associations Exploration (AE) Module to infer scene object associations with fully connected graph models, and a novel Quadruple Graph (QG) Module to facilitate the diffusion and aggregation of semantic association knowledge using graph convolutions. We verify the effectiveness of the proposed modules with comprehensive studies.
- Extensive experiments show that our method outperforms existing state-of-the-art methods both quantitatively and qualitatively on the two popular mirror detection benchmarks, PMD and MSD.

Conclusions

- Our experiments show that while the proposed AE module can help learn semantic associations effectively, the proposed QG module can help detect mirrors accurately with the learned semantic associations.
- Experiement results reveal the effectiveness of semantic associations for mirror detection.