Sketch/Image-Based 3D Scene Retrieval: Benchmark, Algorithm, Evaluation

Juefei Yuan[1], Hameed Abdul-Rashid[1], Bo Li[1], Yijuan Lu[2]

[1]University of Southern Mississippi, [2]Texas State University

THE UNIVERSITY OF SOUTHERN MISSISSIPPI

TEXAS STATE UNIVERSITY
The rising STAR of Texas

Outline

- **Introduction**
- Related Work
- Benchmark
- Method
- Evaluation
- Conclusions and Future Work

Introduction

- **2D Scene Sketch/Image-Based 3D Scene Retrieval**
  - Focuses on retrieving relevant 3D scene models
  - Using scene sketches/image(s) as input
- **Motivation**
  - Vast applications: 3D scene reconstruction, autonomous driving cars, 3D geometry video retrieval, and 3D AR/VR Entertainment
- **Challenges**
  - 2D sketches/images lack 3D scene information
  - Semantic gap: iconic 2D scene sketches or realistic 2D scene images and accurate 3D scene models

Introduction

- **2D Scene Sketch/Image-Based 3D Scene Retrieval**
  - Brand new research topic in sketch/image-based 3D object retrieval:
    - ✓ A query sketch/image contains several objects
    - ✓ Objects may overlap with each other
    - ✓ Relative context configurations among the objects

- We build the **Scene_SBR_IBR benchmark**
  - To promote this challenging research direction
  - Most comprehensive and largest 2D scene sketch/image-based 3D scene retrieval benchmark

Outline

- Introduction
- **Related Work**
- Benchmark
- Method
- Evaluation
- Conclusions and Future Work

Related Work

- **3D Scene Retrieval**
  - Fisher and Hanrahan [1] proposed context-based 3D model retrieval
    - ✓ 3D Bounding box
    - ✓ Dimensionality & context Information
  - Xu et. al [2] proposed Sketch2Scene, a system for automatic 2D sketch-based 3D scene composition
    - ✓ Functional & spatial relationships
    - ✓ Using structural groups

[1] M. Fisher and P. Hanrahan. Context-based search for 3D models. ACM Trans. Graph., 29:182:1–182:10, 2011.

[2] B. Li and et al. A comparison of 3D shape retrieval methods based on a large-scale benchmark supporting multimodal queries. Computer Vision and Image Understanding, 131:1– 27, 2015.

Related Work

- **2D/3D Scene Datasets**
  - Xiao et. al built Scene UNderstanding (SUN) datasets
    - ✓ 130,519 images across 899 scene categories [3]
    - ✓ Expanded to 908 classes [4]
  - Xiao et. al created SUN3D [5]
    - ✓ RGB-D video database with camera pose and object labels

[3] J. Xiao and et al. SUN database: Large-scale scene recognition from abbey to zoo. In CVPR, pages 3485–3492. IEEE Computer Society, 2010.

[4] J. Xiao and et al. SUN database: Exploring a large collection of scene categories. International Journal of Computer Vision, 119(1):3–22, 2016.

[5] J. Xiao and et al. SUN3D: A database of big spaces reconstructed using SfM and object labels. In ICCV, pages 1625–1632, 2013.

Related Work

- **2D/3D Scene Datasets (Cont.)**
  - Song et. al constructed SUNCG [6]
    - ✓ 46,622 synthetized 3D scenes with 2,644 objects
    - ✓ 84 scene categories
  - Zhou et. al compiled Places [7]
    - ✓ 10,624,928 images
    - ✓ 434 scene categories.

[6] S. Song and et al. Semantic scene completion from a single depth image. In CVPR, pages 190–198. IEEE Computer Society, 2017.

[7] B. Zhou and et al. Places: A 10 million image database for scene recognition. IEEE Trans. Pattern Anal. Mach. Intell., 40(6):1452–1464, 2018

Outline

- Introduction
- Related Work
- **Benchmark**
- Method
- Evaluation
- Conclusions and Future Work

Scene_SBR_IBR Benchmark Overview

- **Overview**
  - We have substantially extended the SceneSBR and SceneIBR with 20 additional classes [8, 9]
- **Motivation**
  - Results of SceneSBR and SceneIBR called for a more comprehensive dataset that can support both types of retrieval
- **Building process**
  - Voting method amongst three individuals
  - Scene labels chosen from Places88 [7]
  - Data collected from Flickr, Google Images and 3D Warehouse

[7] B. Zhou and et al. Places: A 10 million image database for scene recognition. IEEE Trans. Pattern Anal. Mach. Intell., 40(6):1452–1464, 2018

[8] J. Yuan and et al. SHREC'18 track: 2D scene sketch-based 3D scene retrieval. In 3DOR, pages 1–8, 2018

[9] H. Abdul-Rashid and et al. SHREC'18 track: 2D scene image-based 3D scene retrieval. In 3DOR, pages 1–8, 2018.

Scene_SBR_IBR Benchmark

- **2D Scene Sketch Query Dataset (Subset 1)**
  - 750 2D scene sketches
  - 30 classes, each with 25 sketches
- **2D Scene Image Query Dataset (Subset 2)**
  - 30,000 scene images
  - 30 classes, each with 1,000 images
- **3D Scene Model Target Dataset (Subset 3)**
  - 3,000 3D scene models
  - 30 classes, each with 100 models

2D Scene Sketch Query Dataset

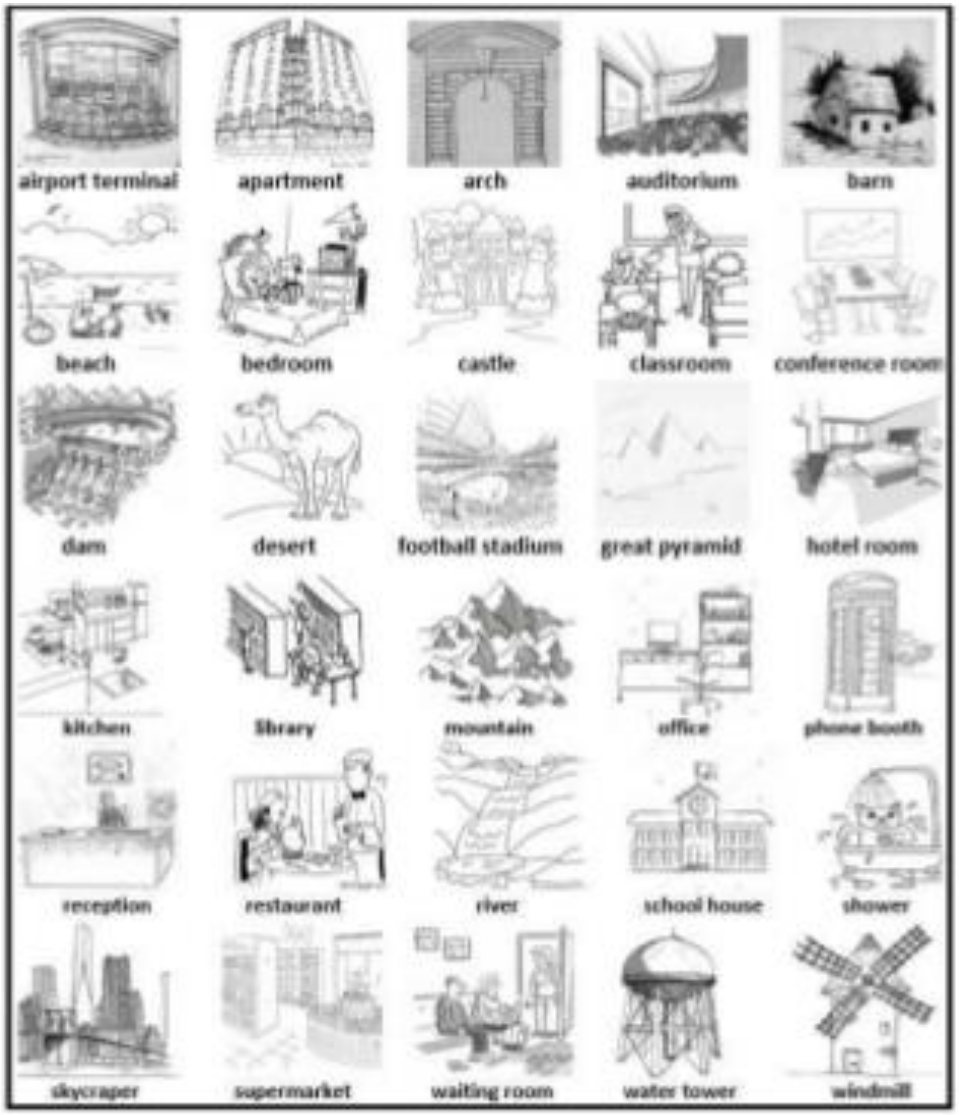

Fig. 1 Example 2D scene query sketches (1 per class)

2D Scene Image Query Dataset

Fig. 2 Example 2D scene query images (1 per class)

3D Scene Model Target Dataset

Fig. 3 Example 3D target scene models (1 per class)

Scene_SBR_IBR Benchmark (Cont.)

- **Supporting both modalities**
  - Scene_SBR: Subsets 1 & 3 (sketch-based retrieval)
  - Scene_IBR: Subsets 2 & 3 (image-based retrieval)
- **To evaluate learning-based 3D scene retrieval**

Table 1. Training and testing dataset information of our **Scene_SBR_IBR** benchmark.

Datasets	Sketches	Images	Models
Training (per class)	18	700	70
Testing (per class)	7	300	30
Total (per class)	25	1,000	100
Total (all 30 classes)	750	30,000	3,000

Evaluation

- **Seven** commonly adopted performance metrics in 3D model retrieval techniques [10]:
  - Precision-Recall plot (PR)
  - Nearest Neighbor (NN)
  - First Tier (FT)
  - Second Tier (ST)
  - E-Measures (E)
  - Discounted Cumulated Gain (DCG)
  - Average Precision (AP)
- We also have developed the code to compute them
  - http://orca.st.usm.edu/~bli/Scene_SBR_IBR/data.html

[10] B. Li and et al. A comparison of 3D shape retrieval methods based on a large-scale benchmark supporting multimodal queries. Computer Vision and Image Understanding, 131:1–27, 2015.

Outline

- Introduction
- Related Work
- Benchmark
- **Method**
- Evaluation
- Conclusions and Future Work

Our Retrieval Algorithm VMV-VGG

VMV-VGG Architecture

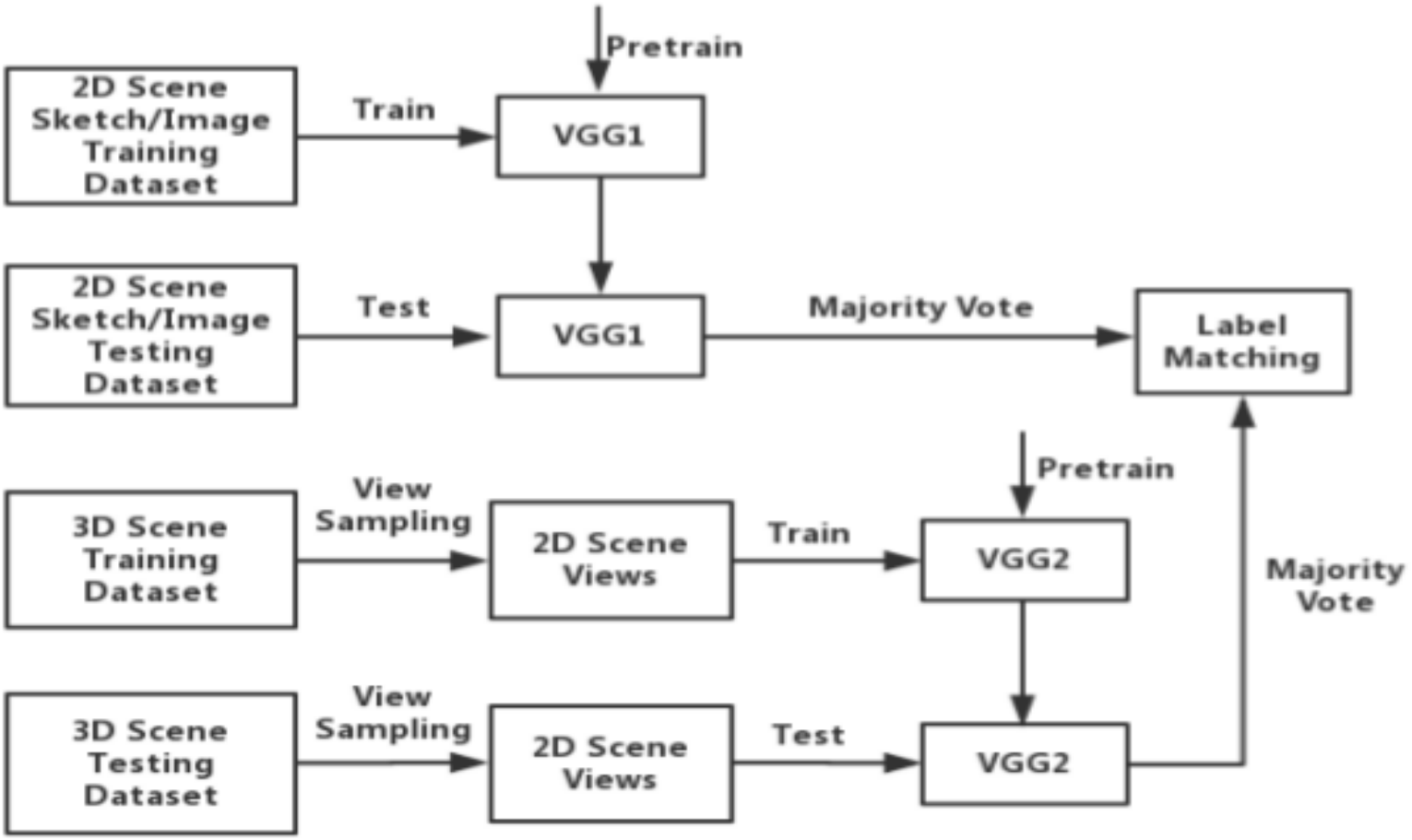

Fig. 4 Our VMV-VGG architecture

VMV-VGG Algorithm

- **VMV-VGG six steps**
  - **(1)** Scene view sampling (Qmacro script)
  - **(2)** Data Augmentation
    - ✓ Random rotations, reflections, or translations
  - **(3)** Pre-training and training on VGG1 and VGG2
  - **(4)** Fine-tuning on scene sketches/images/views
  - **(5)** Sketch/image/view classification
  - **(6)** Majority vote-based label matching

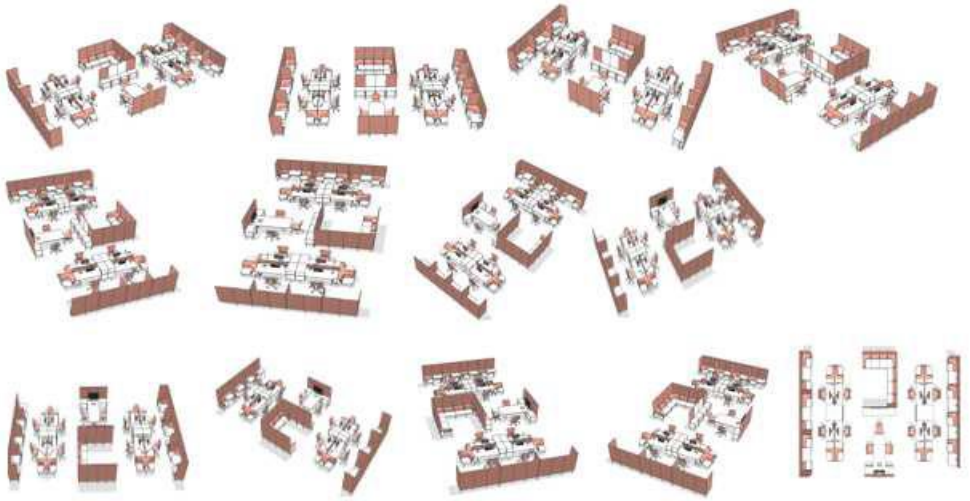

Fig. 5 A set of 13 sample views of an office scene model

Outline

- Introduction
- Related Work
- Benchmark
- Method
- **Evaluation**
- Conclusions and Future Work

Evaluation Overview

- **Evaluation purpose**:
  - Provide the baseline performance for sketch/image-based 3D scene retrieval on our benchmark **Scene_SBR_IBR**
  - Examine the benchmark's comprehensiveness and difficulty level
- **Evaluation content**:
  - Run our VMV-VGG algorithm on the two sub-level benchmarks
    - ✓ Scene_SBR
    - ✓ Scene_IBR

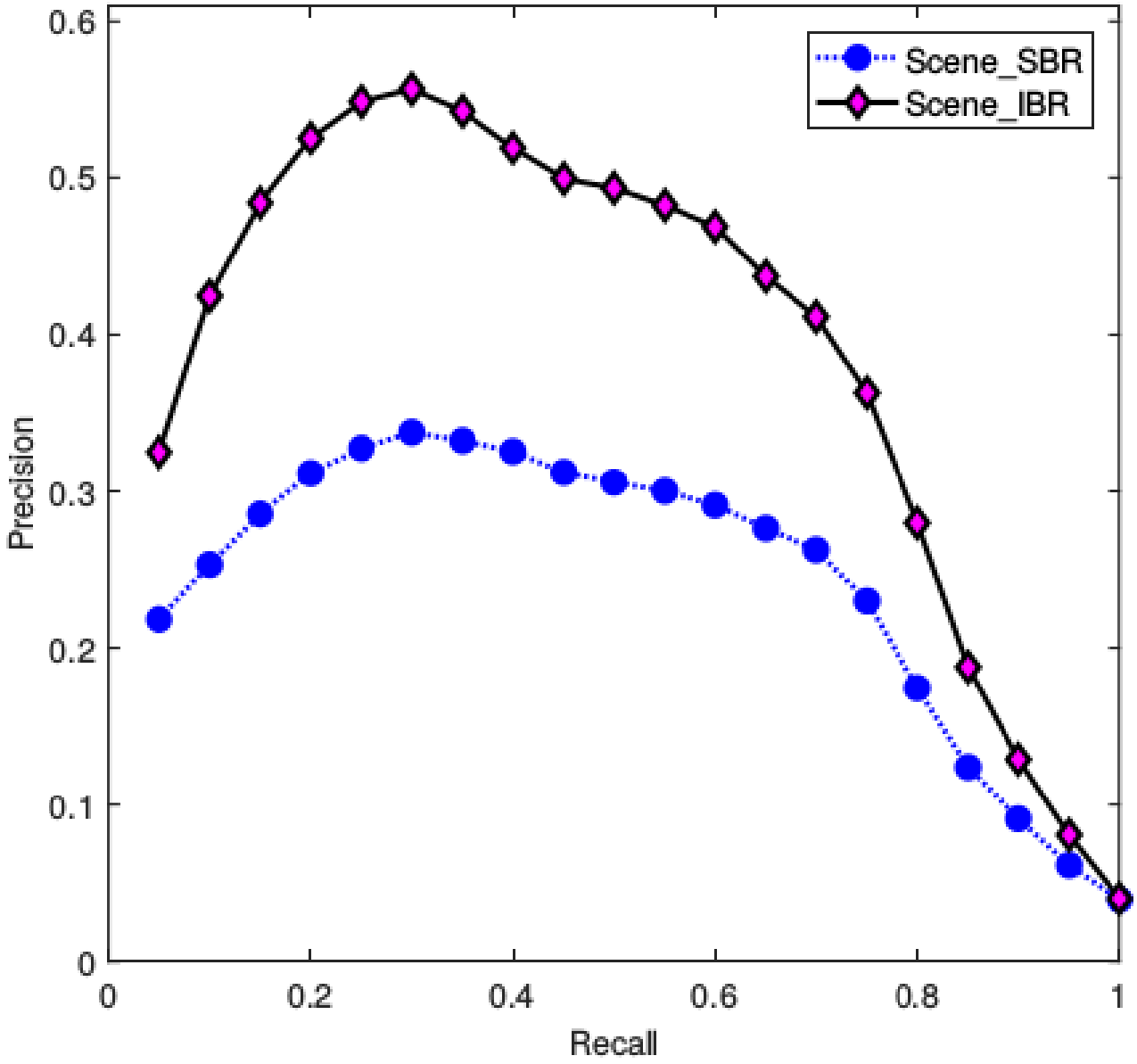

Fig. 6. Precision-Recall diagram performance of the proposed VMV-VGG on our Scene SBR IBR benchmark.

Results: Performance Metrics

- Overall performance dropped significantly if compared with SHREC'18 tracks, due to substantial increase in
  - Comprehensiveness and challenge level
  - Much more scene categories in Scene_SBR_IBR

Table 2. Performance metrics of our VMV-VGG on our **Scene_SBR_IBR** benchmark.

Benchmark	NN	FT	ST	E	DCG	AP
Scene_SBR	0.081	0.281	0.369	0.280	0.533	0.244
Scene_IBR	0.122	0.458	0.573	0.452	0.644	0.392

Table 3. Performance on the SHREC'18 Scene Sketch-Based 3D Scene Retrieval Track Benchmark.

Participant	Method	NN	FT	ST	E	DCG	AP
Testing dataset							
Li	MMD-VGG	0.771	0.630	**0.835**	0.633	0.856	0.685

Table 4. Performance on the SHREC'18 Scene Image-Based 3D Scene Retrieval Track Benchmark.

Participant	Method	NN	FT	ST	E	DCG	AP
Testing dataset							
Li	MMD-VGG	0.910	0.750	0.899	0.750	0.929	0.8032

Outline

- Introduction
- Related Work
- Benchmark
- Method
- Evaluation
- **Conclusions and Future Work**

Conclusions & Future Work

- **Conclusions**
  - **Objective:** To foster this challenging and interesting research direction: *Scene Sketch/Image-Based 3D Scene Retrieval*
  - **Dataset:** Build ***the current largest*** 2D scene sketch/image 3D scene retrieval benchmark
  - **Method:** Baseline performance has been provided by VMV-VGG
  - **Evaluation:** Performed a *comparative evaluation* on the accuracy
  - **Impact:** Provided *the largest and most comprehensive common evaluation platform* for sketch/image-based 3D scene retrieval
- **Future work**
  - Build a large-scale and/or multimodal 2D scene-based 3D scene retrieval benchmark
  - Semantics-driven 2D scene image-based 3D scene retrieval

References

- [1] M. Fisher and P. Hanrahan. Context-based search for 3D models. ACM Trans. Graph., 29:182:1–182:10, 2011.
- [2] B. Li and et al. A comparison of 3D shape retrieval methods based on a large-scale benchmark supporting multimodal queries. Computer Vision and Image Understanding, 131:1– 27, 2015.
- [3] J. Xiao and et al. SUN database: Large-scale scene recognition from abbey to zoo. In CVPR, pages 3485–3492. IEEE Computer Society, 2010.
- [4] J. Xiao and et al. SUN database: Exploring a large collection of scene categories. International Journal of Computer Vision, 119(1):3–22, 2016.
- [5] J. Xiao and et al. SUN3D: A database of big spaces reconstructed using SfM and object labels. In ICCV, pages 1625–1632, 2013.
- [6] S. Song and et al. Semantic scene completion from a single depth image. In CVPR, pages 190–198. IEEE Computer Society, 2017.
- [7] B. Zhou and et al. Places: A 10 million image database for scene recognition. IEEE Trans. Pattern Anal. Mach. Intell., 40(6):1452–1464, 2018
- [8] J. Yuan and et al. SHREC'18 track: 2D scene sketch-based 3D scene retrieval. In 3DOR, pages 1–8, 2018
- [9] H. Abdul-Rashid and et al. SHREC'18 track: 2D scene image-based 3D scene retrieval. In 3DOR, pages 1–8, 2018.
- [10] B. Li and et al. A comparison of 3D shape retrieval methods based on a large-scale benchmark supporting multimodal queries. Computer Vision and Image Understanding, 131:1–27, 2015.

Thank you!

Q&A?