# SDGOCC: Semantic and Depth-Guided Bird's-Eye View Transformation for 3D Multimodal Occupancy Prediction

ZaiPeng Duan ChenXu Dang Xuzhong Hu Pei An Junfeng Ding Jie Zhan YunBiao Xu Jie Ma *
Huazhong University of Science and Technology

## Introduction

**What:**

**SDG-OCC** is a lightweight yet powerful LiDAR and image fusion-based occupancy prediction method. By combining semantic and depth-guided view transformation and fusion with occupancy-driven active distillation, it achieves SOTA performance while ensuring real-time speed.

**Why:**

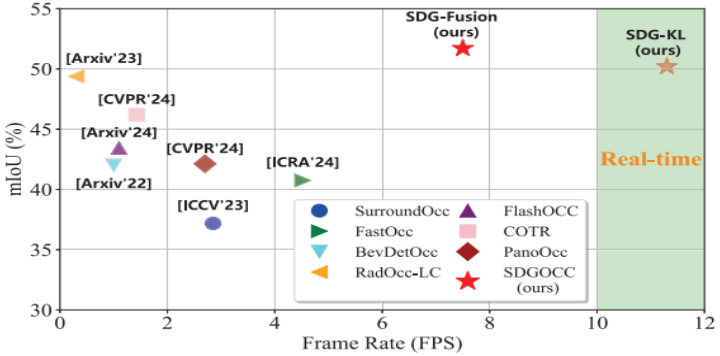
Existing methods:

- The lightweight method based on the LSS pipeline suffers from low feature utilization, leading to suboptimal performance.
- LiDAR information can enhance performance, but it introduces additional computational burden, leading to poor real-time efficiency.

**How:**

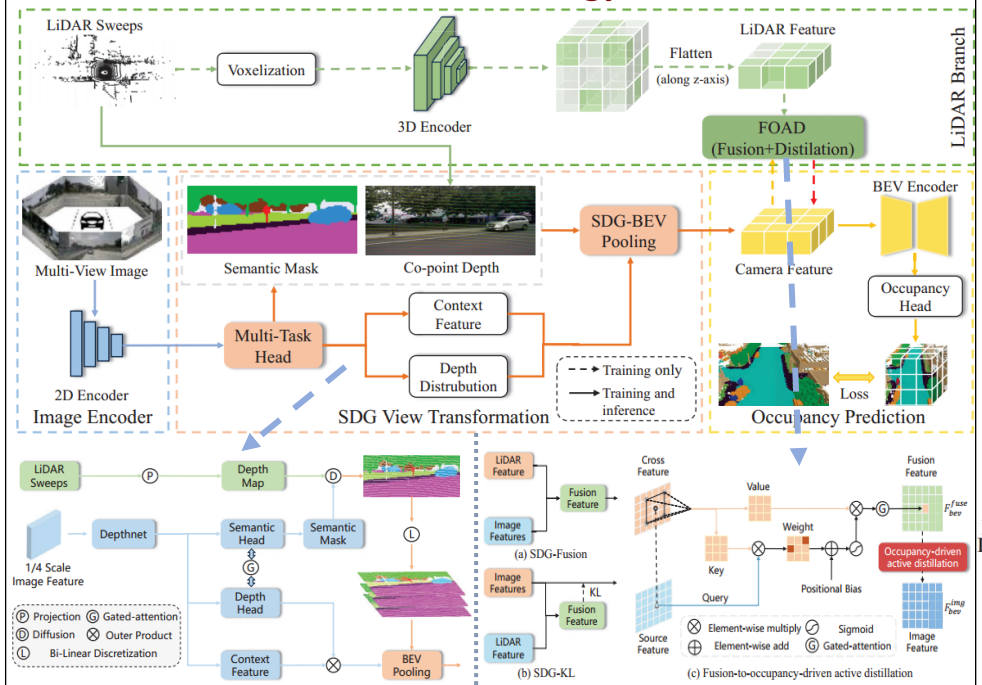
We propose:

- A view transformation method that leverages the geometric and semantic information of point clouds to guide 2D-3D transformation, significantly improving depth estimation accuracy and occupancy speed.
- An occupancy-driven active distillation module that integrates multimodal features and optimizes knowledge transfer based on LiDAR-identified regions.




Code is available at **https://github.com/DzpLab/SDGOCC.**

## Methodology

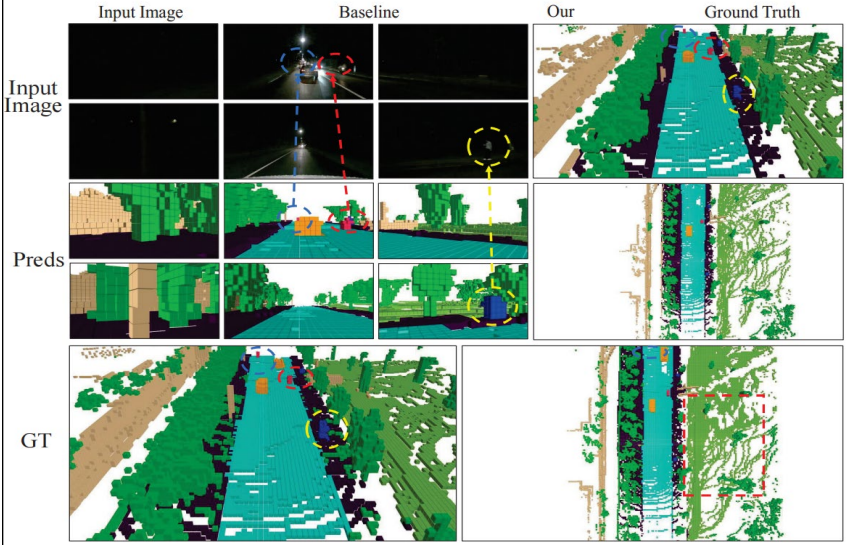



## Experiments

Effectiveness on Occ3D-nuScenes Val Dataset

| Method | Input | Backbone | Visible Mask | mIoU | others | barrier | bicycle | bus | car | const. veh. | motorcycle | pedestrian | traffic cone | trailer | truck | drive. suf. | other flat | sidewalk | terrain | manmade | vegetation | Time(ms) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TPVFormer [9] | C | R-50 | ✔ | 34.2 | 7.68 | 44.01 | 17.66 | 40.88 | 46.98 | 15.06 | 20.54 | 24.69 | 24.66 | 24.26 | 29.28 | 79.27 | 40.65 | 48.49 | 49.44 | 32.63 | 29.82 | 289.85 |
| SurroundOcc [35] | C | R-101 | ✔ | 37.1 | 8.97 | 46.33 | 17.08 | 46.54 | 52.01 | 20.05 | 21.47 | 23.52 | 18.67 | 31.51 | 37.56 | 81.91 | 41.64 | 50.76 | 53.93 | 42.91 | 37.16 | 303.03 |
| OccFormer [44] | C | R-50 | ✔ | 37.4 | 9.15 | 45.84 | 18.20 | 42.80 | 50.27 | 24.00 | 20.80 | 22.86 | 20.98 | 31.94 | 38.13 | 80.13 | 38.24 | 50.83 | 54.3 | 46.41 | 40.15 | - |
| VoxFormer [13] | C | R-101 | ✔ | 40.7 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| FBOcc [16] | C | R-50 | ✔ | 42.1 | 14.30 | 49.71 | 30.0 | 46.62 | 51.54 | 29.3 | 29.13 | 29.35 | 30.48 | 34.97 | 39.36 | 83.07 | 47.16 | 55.62 | 59.88 | 44.89 | 39.58 | - |
| PanoOcc [33] | C | R-101 | - | 42.13 | 11.67 | 50.48 | 29.64 | 49.44 | 55.52 | 23.29 | 33.26 | 30.55 | 30.99 | 34.43 | 42.57 | 83.31 | 44.23 | 54.40 | 56.04 | 45.94 | 40.40 | 322.58 |
| FastOcc [6] | C | R-101 | ✔ | 40.75 | 12.86 | 46.58 | 29.93 | 46.07 | 54.09 | 23.74 | 31.10 | 30.68 | 28.52 | 33.08 | 39.69 | 83.33 | 44.65 | 53.90 | 55.46 | 42.61 | 36.50 | 221.2 |
| BEVDet4D [8] | C | Swin-B | ✔ | 42.5 | 12.37 | 50.15 | 26.97 | 51.86 | 54.65 | 28.38 | 28.96 | 29.02 | 28.28 | 37.05 | 42.52 | 82.55 | 43.15 | 54.87 | 58.33 | 48.78 | 43.79 | 1000.0 |
| FlashOcc [42] | C | Swin-B | ✔ | 43.52 | 13.31 | 51.62 | 28.07 | 50.91 | 55.69 | 27.46 | 31.05 | 29.98 | 29.20 | 38.86 | 43.68 | 83.87 | 45.63 | 56.33 | 59.01 | 50.63 | 44.56 | 909.1 |
| COTR [21] | C | Swin-B | ✔ | 46.2 | **14.85** | 53.25 | **35.19** | 50.83 | 57.25 | 35.36 | 34.06 | 33.54 | **37.14** | 38.99 | 44.97 | 84.46 | **48.73** | 57.60 | 61.08 | 51.61 | 46.72 | 840.34 |
| HyDRa [36] | C+R | R-50 | - | 44.40 | - | - | - | 52.3 | - | - | 35.9 | 35.10 | - | - | 44.1 | - | - | - | - | - | - | - |
| OCCFusion [23] | C+L | R-101 | - | 46.79 | 11.65 | 47.81 | 32.07 | 57.27 | 57.51 | 31.80 | **40.11** | 47.35 | 33.74 | 45.81 | 50.35 | 78.79 | 37.17 | 44.36 | 53.36 | 63.18 | 63.20 | - |
| RadOcc-LC [43] | C+L | Swin-B | ✔ | 49.38 | 10.93 | **58.23** | 25.01 | 57.89 | 62.85 | 34.04 | 33.45 | 50.07 | 32.05 | 48.87 | 52.11 | 82.90 | 42.73 | 55.27 | 58.34 | 68.64 | 66.01 | 3333 |
| SDG-KL | C+L | R-50 | ✔ | 50.16 | 12.26 | 57.12 | 23.69 | 58.77 | 62.74 | 34.55 | 36.19 | 50.1 | 32.05 | 49.89 | 51.24 | 84.1 | 46.05 | 57.2 | 61.45 | 69.56 | 65.78 | **83** |
| SDG-Fusion | C+L | R-50 | ✔ | **51.66** | 13.21 | 57.77 | 24.3 | **60.33** | **64.28** | **36.21** | 39.44 | **52.36** | 35.80 | **50.91** | **53.65** | **84.56** | 47.45 | **58.00** | **61.61** | **70.67** | **67.65** | 133 |

## Visualization




**We anticipate that:**

- The proposed SDG-OCC provides deeper insights into 2D-3D visual transformation and multimodal occupancy prediction.