

S3DIS — это один из самых важных и часто используемых датасетов для семантической сегментации помещений.

S3DIS (Stanford Large-Scale 3D Indoor Spaces Dataset) — это масштабный датасет для 3D семантической сегментации внутренних помещений, созданный в Стэнфордском университете.

Ключевые характеристики:

- **Тип сцен:** Внутренние помещения (офисы, учебные аудитории, коридоры)
- **Масштаб:** 6 больших зданий, 271 комната
- **Объем:** ~215 миллионов точек
- **Аннотации:** 13 семантических классов
- **Сенсор:** Matterport Scanner (стационарный 3D сканер)
- **Особенность:** Полные 3D модели целых этажей

Семантические классы S3DIS:

```
0 - ceiling (потолок)
1 - floor (пол)
2 - wall (стена)
3 - beam (балка)
4 - column (колонна)
5 - window (окно)
6 - door (дверь)
7 - table (стол)
8 - chair (стул)
9 - sofa (диван)
10 - bookcase (книжный шкаф)
11 - board (доска)
12 - clutter (разное/мусор)
```

Структура каталогов S3DIS

Оригинальная структура:

```
s3dis/
└── Area_1/
    ├── conferenceRoom_1/
    │   ├── Annotation/
    │   │   ├── beam_1.ptx
    │   │   ├── ceiling_1.ptx
    │   │   ├── floor_1.ptx
    │   │   └── ...
    │   └── conferenceRoom_1.txt
    └── copyRoom_1/
```

```
|   └── office_1/  
|       └── ...  
├── Area_2/  
├── Area_3/  
├── Area_4/  
├── Area_5/  
└── Area_6/
```

Детальная структура комнаты:

```
Area_1/office_1/  
└── Annotation/          # Аннотированные объекты  
    ├── ceiling_1.ptx    # Точки потолка  
    ├── floor_1.ptx      # Точки пола  
    ├── wall_1.ptx       # Точки стен  
    ├── door_1.ptx       # Точки дверей  
    ├── table_1.ptx      # Точки столов  
    ├── chair_1.ptx      # Точки стульев  
    └── ...  
└── office_1.txt         # Полное точечное облако комнаты  
└── office_1.jpg         # RGB изображение (опционально)
```

Популярное фиксированное разделение:

```
train_areas = ['Area_1', 'Area_2', 'Area_3', 'Area_4', 'Area_6']  
test_area = ['Area_5']
```

Ключевые особенности S3DIS:

- Полные помещения:** Целые комнаты с полной геометрией
- Реалистичные условия:** Различные типы офисных помещений
- Сложная сегментация:** Много мелких объектов и "clutter"
- Стандартный бенчмарк:** Широко используется для сравнения методов
- Сбалансированные классы:** Хорошее распределение по категориям

Сравнение с другими датасетами:

Характеристика	S3DIS	ScanNet	Semantic3D
Тип сцен	Indoor помещения	Indoor помещения	Outdoor сцены
Масштаб	271 комната	1513 сканов	30 сцен
Точки	~215 млн	~2.5 млн видов	Миллиарды точек
Классы	13	20	8
Особенность	Полные комнаты	Последовательности	Высокая плотность