



CSU101: An Educational Dataset for Introductory Computer Vision

Ethan Seefried, Changsoo Jung, Videep Venkatesha, Trevor Chartier, Caleb Christian, Jack Fitzgerald, Mariah Bradford, Sifatul Anindho, Matthew Sturgeon, Nathaniel Blanchard
Colorado State University



DENVER
COLORADO

Motivation + Dataset

The Teaching Gap

Current CV teaching datasets are useful but fragmented. Students often learn classification, detection, and 3D reconstruction using unrelated datasets, which increases cognitive overhead and makes the learning path feel disconnected.

CSU101 Core Idea

CSU101 uses familiar campus scenes to scaffold an introductory CV course. Students start with building classification, move to campus object detection, and finish with 3D Gaussian Splatting using the same underlying video data.

Dataset at a Glance



Figure 1. CSU101 at a glance. Dataset scale and supported tasks: classification, detection, and 3D reconstruction.

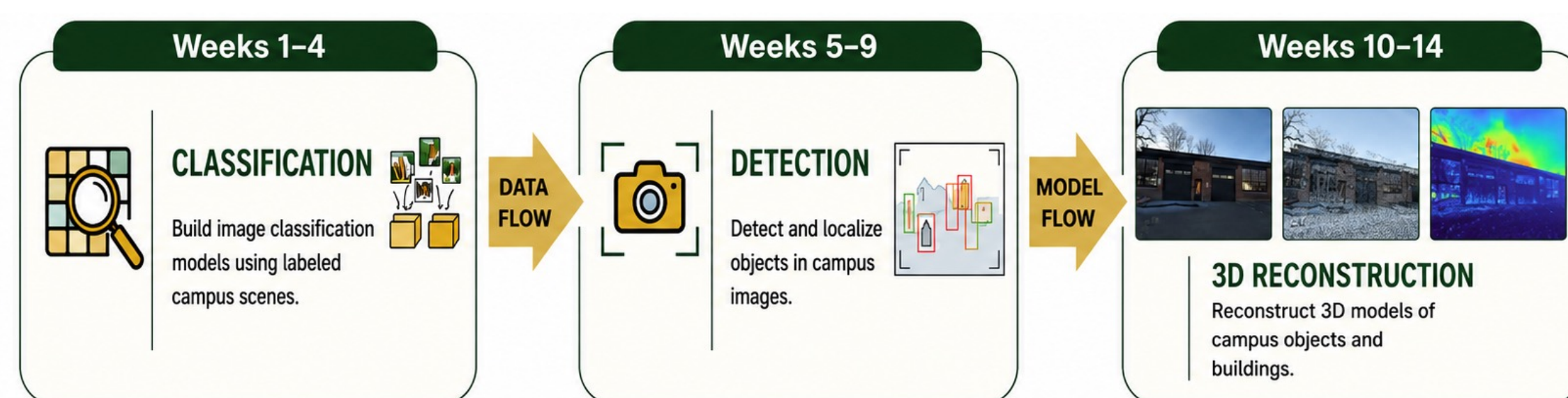


Figure 2. CSU101 curriculum progression.

CSU101 Pipeline

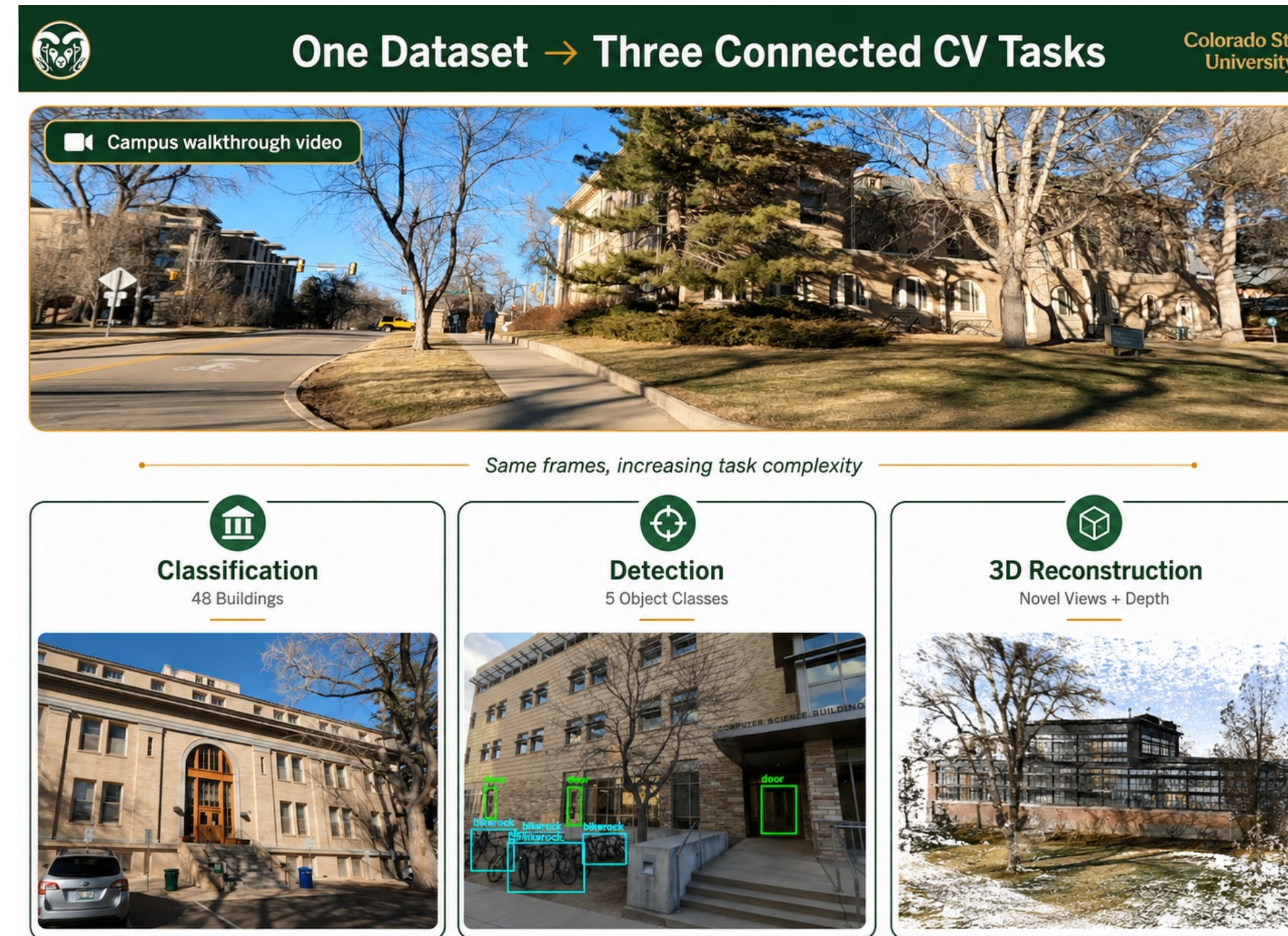


Figure 3: CSU101 Pipeline

Dataset Curation

Dataset Collection

- Campus walkthrough videos captured using a **GoPro Hero 10**
- Videos recorded at **24 FPS and 1080p resolution**
- Each building was filmed by walking around its perimeter when possible
- Dataset covers **48 university buildings**
- Privacy protections: **dorms excluded** and **visible faces blurred**

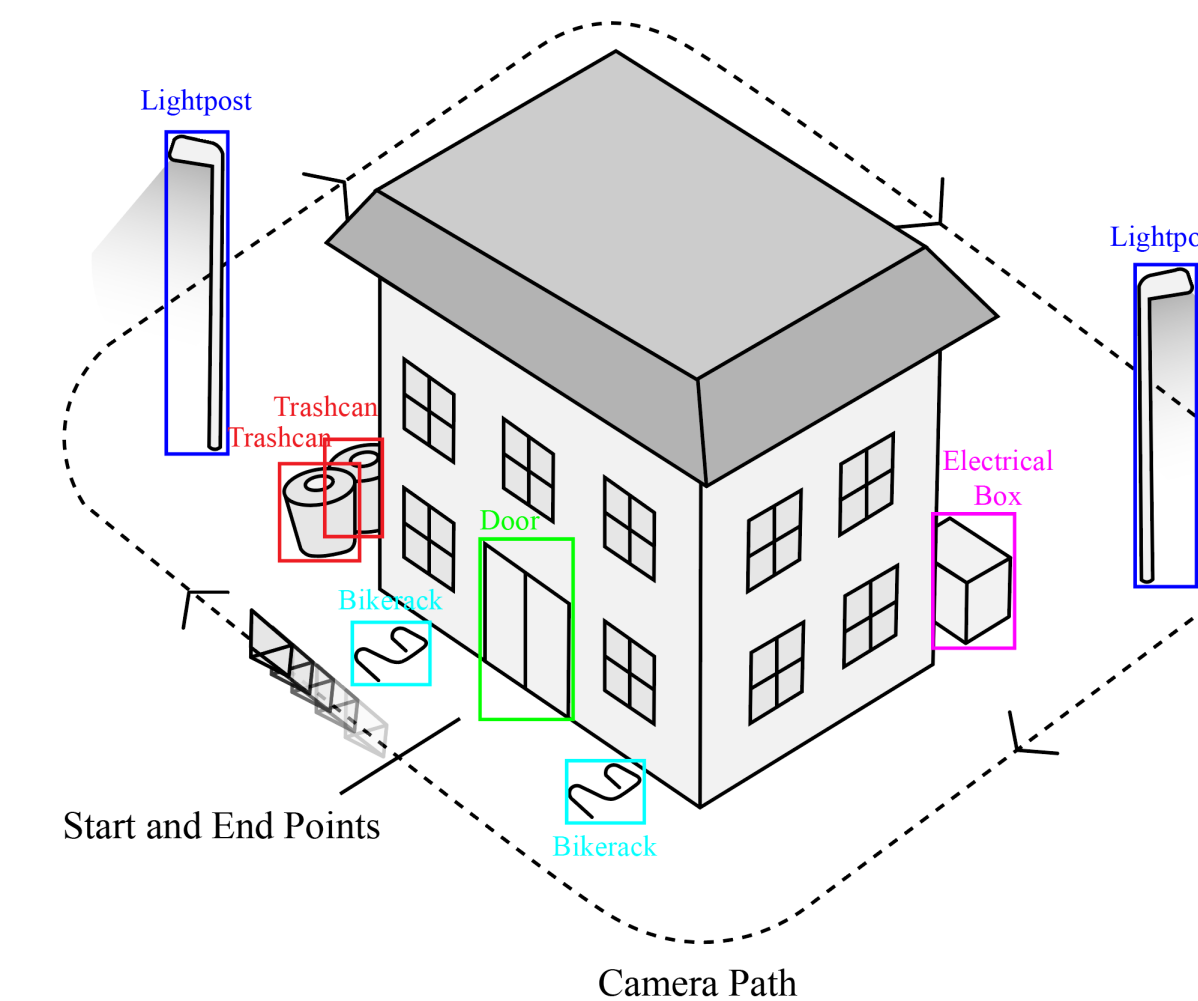


Figure 4: Example of dataset Collection and Annotation

Object Distributions

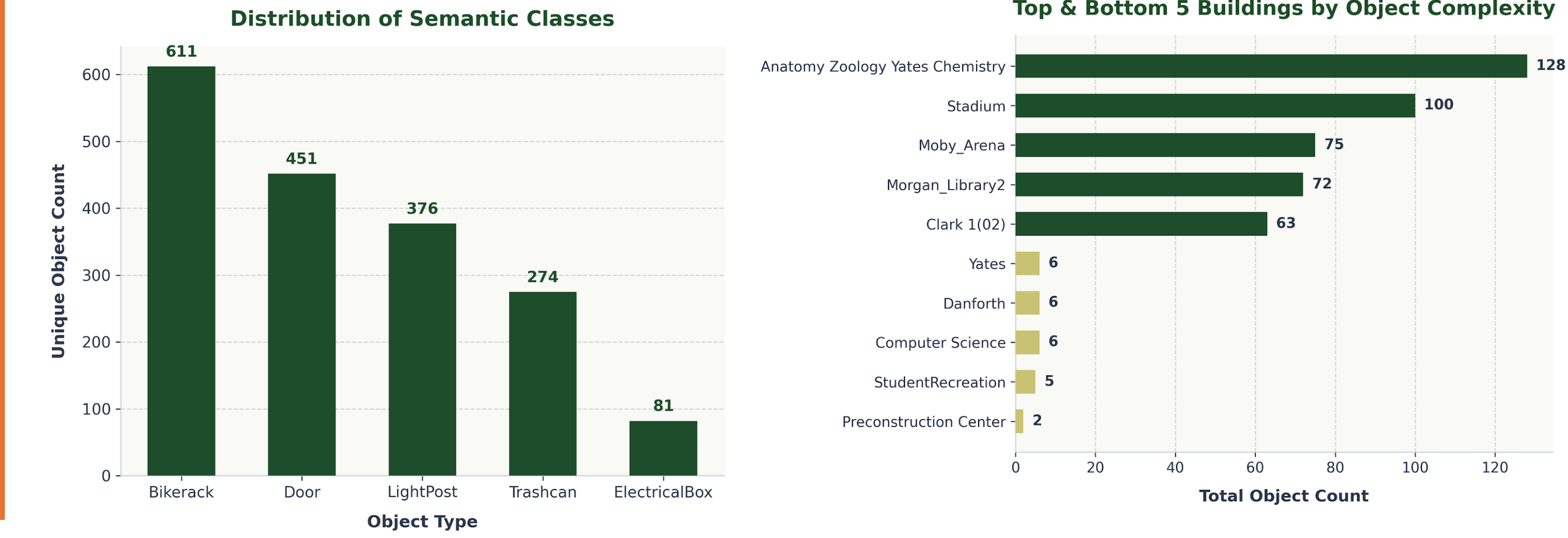


Figure 5: Object Counts within CSU101

Figure 6: Object Counts per Building

Results

Baseline Models

Classification

- VGG19, EfficientNet, ResNet50, ResNeXt50, ViT, Swin

Detection

- Faster R-CNN, FCOS, RetinaNet, SSD, SSDlite

3D Reconstruction

- Splatfacto / Gaussian Splatting

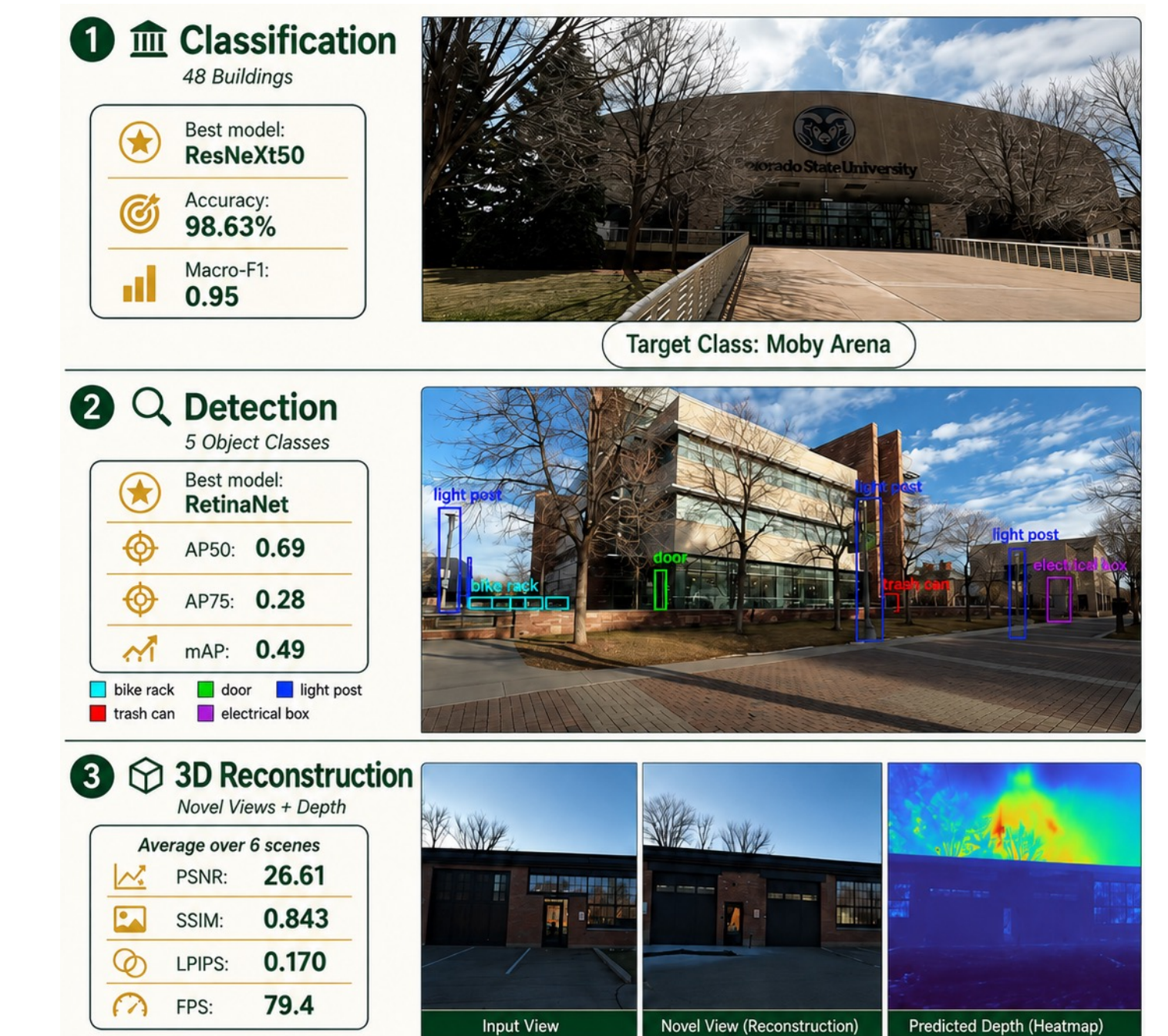


Figure 7: Baseline Results on CSU101

Conclusions

CSU101 unifies classification, object detection, and 3D reconstruction in one classroom-ready dataset. Familiar campus scenes, labels, baselines, and notebooks support a coherent path from 2D recognition to 3D understanding.