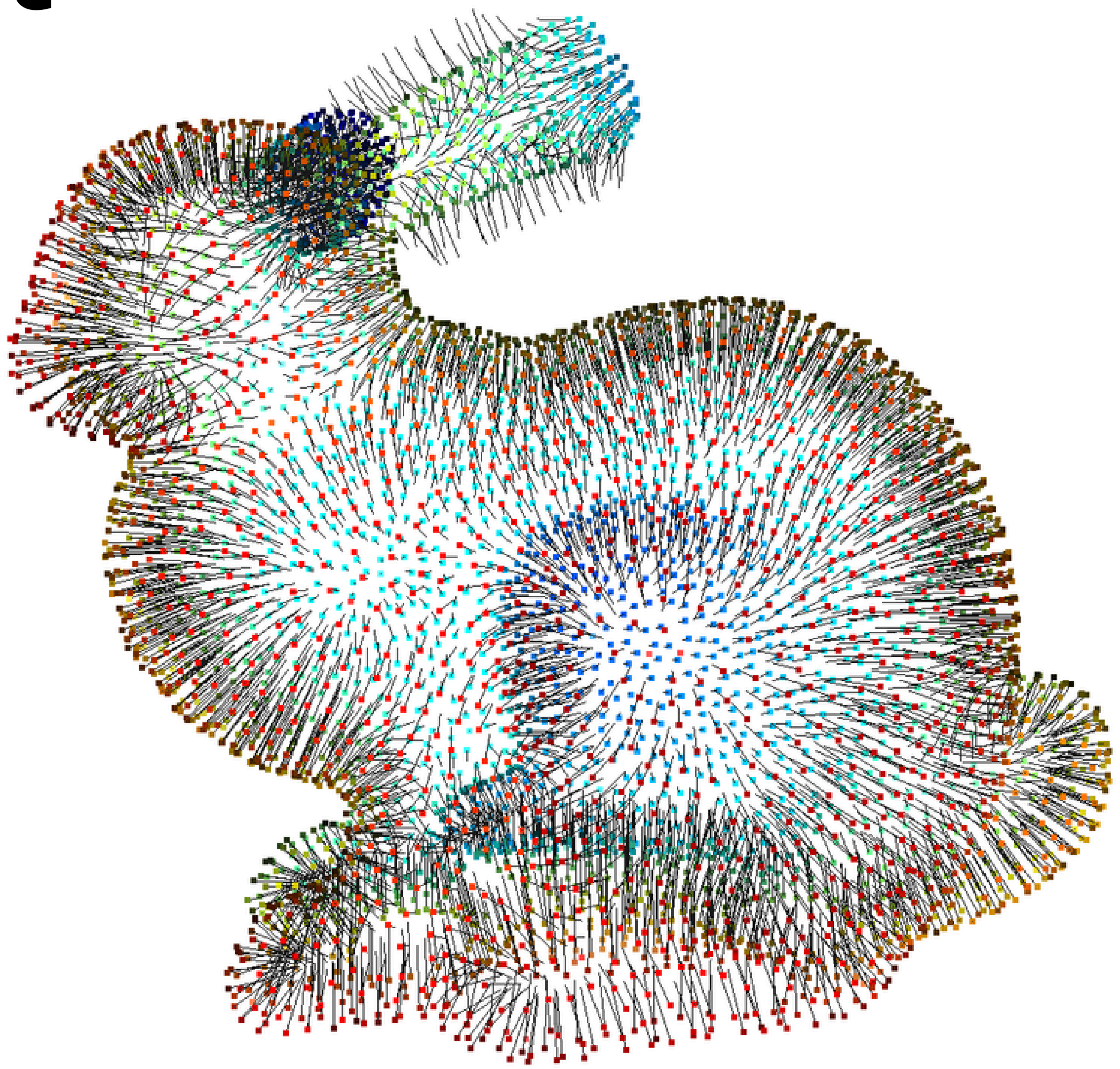
SAPIENZA
UNIVERSITÀ DI ROMA

Lightweight Convolutional Occupancy Networks for Efficient Virtual Scene Generation



Computer Vision A.A. 2024-2025 - Project 5

Leonardo Sandri -2137374

Presentation Outline

Content:

Challenge: Fast, Accurate 3D Scenes

The Problem:

- Generating accurate 3D representations of scenes is critical in each field that requires simulations. However, a major challenge is the trade-off between reconstruction quality and inference speed, which limits real-time applicability.

Project Goal:

- Implement and optimize a Lightweight Convolutional Occupancy Network (TransONet) for virtual scene generation from point clouds, specifically adapted to the synthetic_rooms_dataset. The primary objective is minimizing inference time while preserving reconstruction quality, enabling deployment on embedded devices.

State of the Art: Occupancy Networks

Instead of explicit meshes, Occupancy Networks learn a continuous function that maps any 3D point to its occupancy probability. This allows for extracting meshes at any desired resolution.

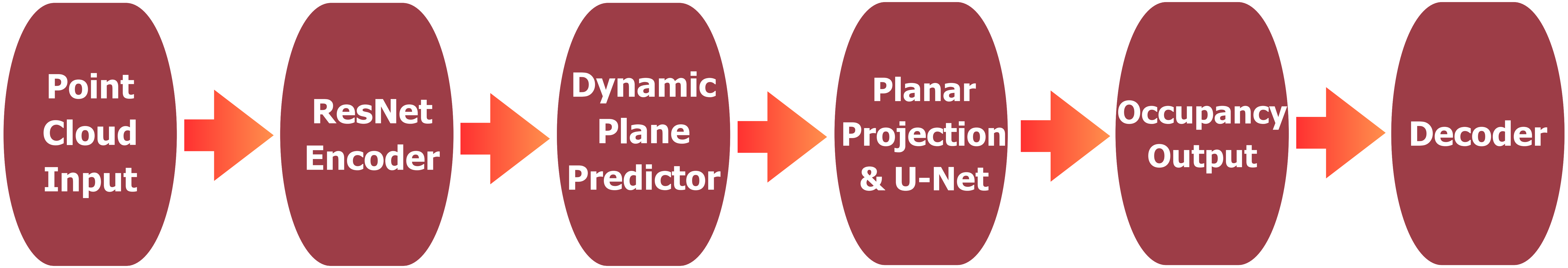
Key Architectures

- **Convolutional Occupancy Networks** (Peng et al.): 3D convolutional features, powerful but computationally intensive.
- **Dynamic Plane Convolutional Occupancy Networks** (Lionar et al.): Proposes projecting 3D features onto dynamic 2D planes, significantly improving efficiency.

Proposed Method: An implementation and methodical experimentation and analysis based on the ALCOR Lab TransONet architecture, based on the Dynamic Plane CONet concept.

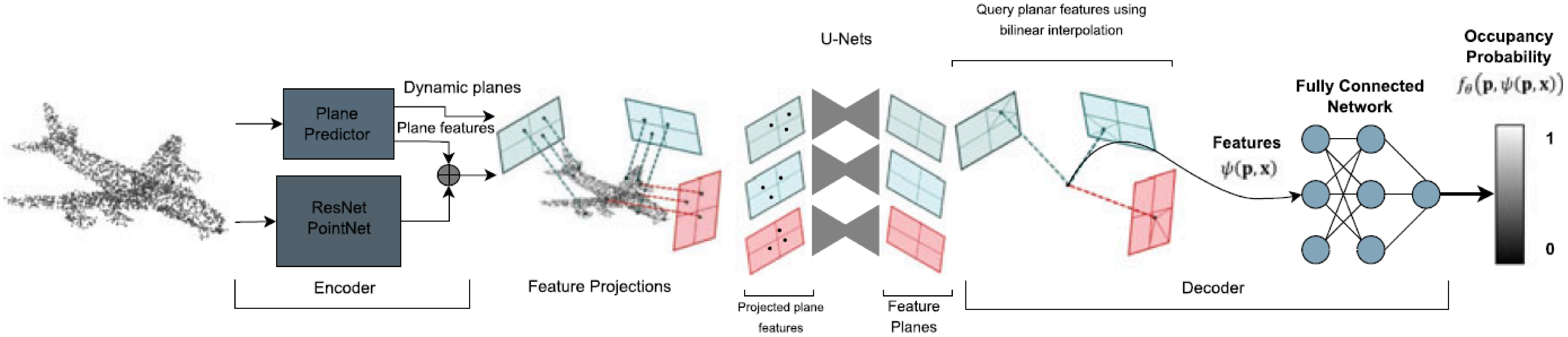
Approach: The TransONet Model

High-Level Diagram:



Challenge:
The original ALCOR Lab repository was Docker-based, which is incompatible with cloud platforms like Google Colab/Kaggle.
Solution:
Re-engineered the entire code structure to be cloud-native, enabling GPU-accelerated training and evaluation for this project.

Approach: The TransONet Model



(Figure extracted from reference [2])

Key Components:

- **Network**: A ResNet-based PointNet encoder extracts point-wise features.
- **Dynamic Planes**: A predictor learns 3 orthogonal planes to project features onto.
- **U-Net**: Processes these 2D feature planes to capture hierarchical spatial information.
- **Decoder**: A final MLP predicts the occupancy value for any given 3D query point

Dataset and Experimental Configuration

Dataset:

- **Dataset:** Synthetic-Rooms subset
- Provided dataset composed of structured scenes from ShapeNet.
- **Training Set:** 1600 samples
- **Validation Set:** 400 samples

Best parameters:

- **Batch Size**: 64
- **Learning Rate**: 5e-4 with Adam Optimizer and **ReduceLROnPlateau**
- **Epochs**: 100 (with Early Stopping)
- **Sampling**: WeightedRandomSampler with a 16x multiplier for positive samples

Model Evaluation: An Iterative Journey

Methodology: A series of experiments was conducted, methodically addressing performance bottlenecks.

Key Iterations & Insights:

- **(Baseline):** Low Precision (6.83%) and Slow Inference (2.31s).
- **(Speed):** Resolved Inference Bottleneck. Vectorized the PlanarProjection module, achieving a 23x speedup to 0.098s.
- **(Stability):** Introduced Early Stopping & LR Scheduler. Improve in precision with longer trains.
- **(Regression):** Learned that overly aggressive sampling (25x) and dropout (0.2) harmed performance.
- **(Breakthrough):** Recalibrated hyperparameters (Batch Size 64, LR 5e-4, Sampling 16x).

Model Evaluation: An Iterative Journey

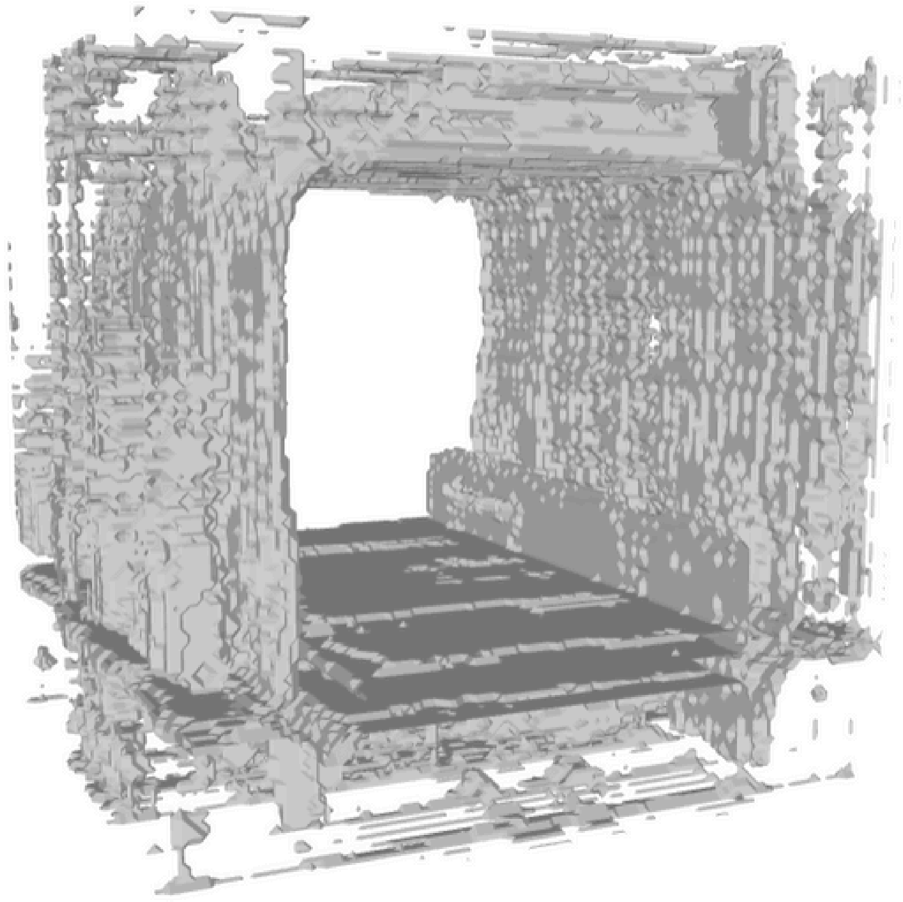
From Bottlenecks to Breakthroughs:

Challenge:	Solution	Impact:
1. Prediction Collapse	Strategic Bias Initialization & weighted sampling	Enabled Meaningful Learning
2. Severe Over-Prediction and long training times	Systematic Weighted Sampling	Precision Improvement
3. Crippling Inference Speed	Vectorizing the PlanarProjection	Reduced inference time to ~0.1s, Speedup

Final Results & Objectives Achieved

Final Results & Objectives Achieved:

Metric	Baseline	Final	Improvement
F1-Score	12.58%	19.87%	+58%
Precision	6.83%	11.49%	+68%
Volumetric IoU	7.11%	11.03%	+55%
Inference Time	2.31s	0.09 - 0.232s	-90%



Conclusions & Future Work

Key Learnings:

- **Systematic Tuning is Crucial:** Success was driven by a balanced, data-driven approach.
- **The 3-Layer U-Net is Essential**: A shallower, 2-layer U-Net, was tested hypothesizing it could be more efficient. This led to a significant performance regression.
- **The Architecture Resisted Simple Pruning**: Experiments with L1 pruning showed that the model's performance is intrinsically tied to its dense 1.77M parameters. This presents a compelling challenge for future work in model compression.

Future Work:

- **Refine Sampling**: Explore the optimal sampling multiplier range with more granularity.
- **Advanced Loss Functions**: Experiment with Focal Loss to better handle the class imbalance.
- **Model Compression**: Investigate knowledge distillation as a more effective compression technique than pruning for this architecture

References

1. Lionar, S., Emtsev, D., Svilarkovic, D., and Peng, S. (2020). Dynamic Plane Convolutional Occupancy Networks. arXiv.
2. Tonti, C. M., Papa L. and Amerini, I., "Lightweight 3-D Convolutional Occupancy Networks for Virtual Object Reconstruction," in IEEE Computer Graphics and Applications, 2024.
3. Peng, S., Niemeyer, M., Mescheder, L., Pollefeys, M., and Geiger, A. (2020). Convolutional Occupancy Networks. arXiv.