The Challenge of Machine Perception

For an autonomous vehicle to navigate safely, it must do more than just ‘see’ the world; it must understand it at a granular level. Every single pixel in its field of view contains information critical for decision-making. This requires a level of environmental comprehension that traditional computer vision struggles to achieve amidst real-world complexity like occlusions, shadows, and unpredictable weather.

Translating Pixels into Meaning with Semantic Segmentation

RAW IMAGE

SEGMENTED MAP

Semantic Segmentation is the task of classifying each pixel in an image into a predefined category. Instead of identifying an object with a bounding box, we create a detailed, pixel-perfect map of the environment. This process transforms a raw image into a structured, machine-readable format, distinguishing road from sidewalk, vehicle from pedestrian, and building from sky.

The Perfect Sandbox: Training Perception in a Simulated World

To build a robust model, we begin in a controlled environment. The CARLA Simulator provides a dataset of 5,000 synthetic driving scenes with perfectly labeled ground truth masks. This offers two critical advantages:

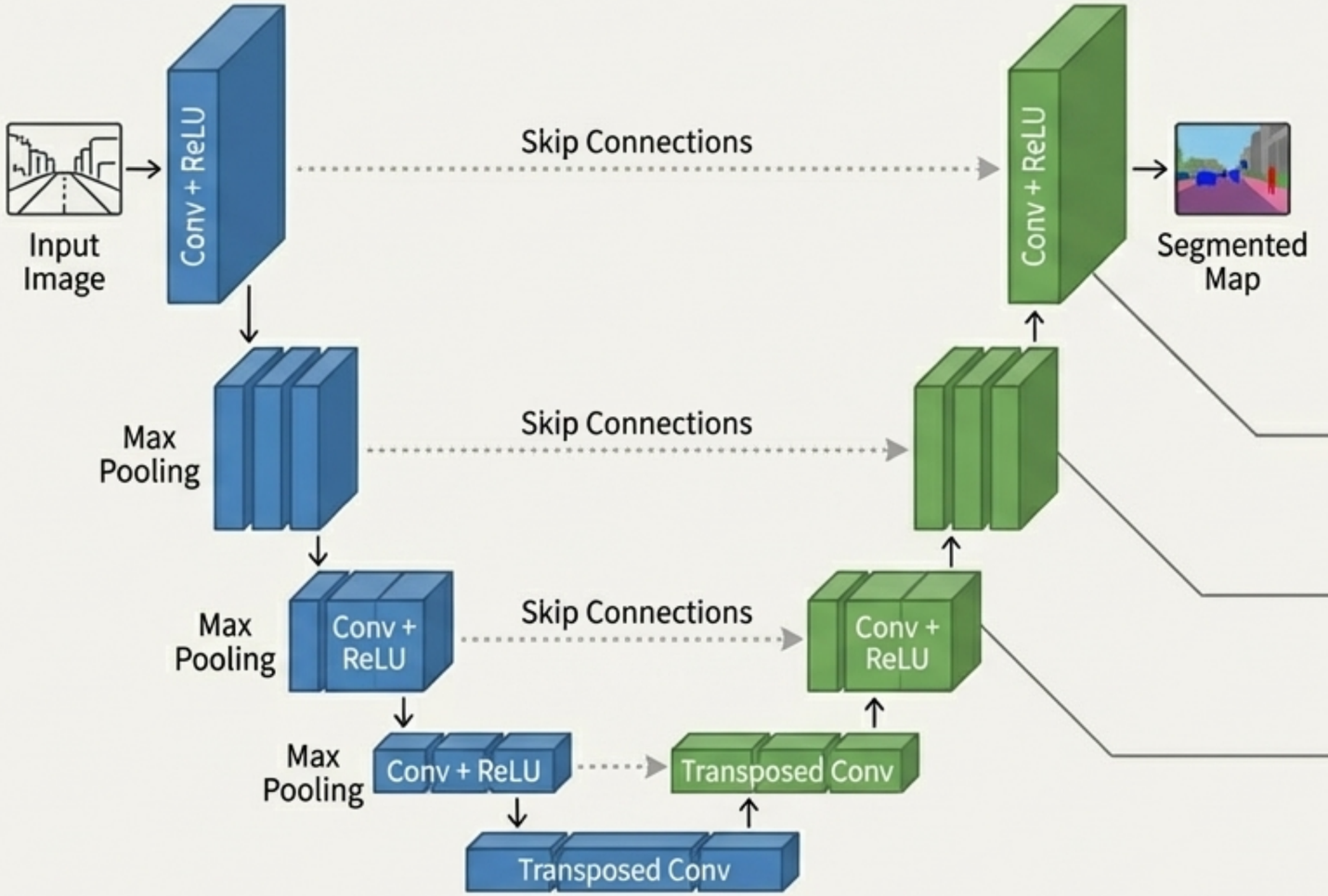
Pixel-Perfect Labels

Eliminates the noise and annotation errors common in real-world datasets.

Baseline Foundation

Allows us to build and validate a strong baseline model before tackling the complexities of real-world data.

The U-Net Architecture: Designed for Precise Segmentation



We selected the U-Net model, an architecture renowned for its effectiveness in biomedical image segmentation and perfectly suited for this task. Its key strength lies in an elegant design that captures context and enables precise localization.

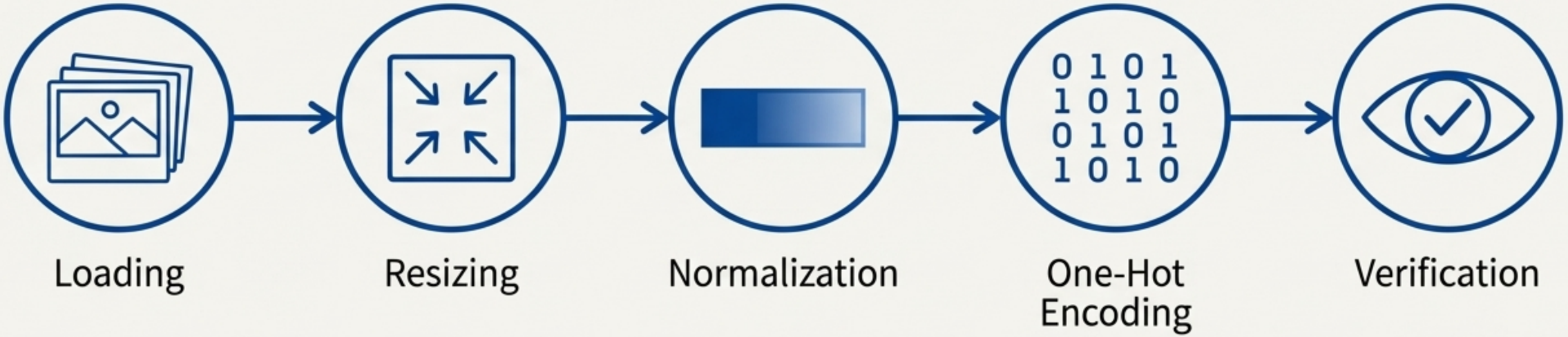
Encoder (Contracting Path): Extracts feature representations from the image.

Decoder (Expanding Path): Reconstructs a full-resolution segmentation map.

Skip Connections: The critical innovation that preserves fine spatial details often lost during downsampling.

Step 1: Preparing the Data for Training

Meticulous data preprocessing is crucial for model performance and convergence. Our pipeline involved several key steps:



- **Loading**: Importing the CARLA RGB images and their corresponding labeled masks.
- **Resizing**: Standardizing all images and masks to a uniform shape of 256×256 pixels.
- **Normalization**: Scaling pixel values to a [0, 1] range to accelerate training.
- **One-Hot Encoding**: Converting ground-truth masks into the required format for multi-class segmentation.
- **Verification**: Overlaying masks on images to visually confirm correct alignment and labeling before training.

Step 2: Forging the Model with a Defined Training Regimen

The model was trained using a standard yet effective configuration designed to optimize learning and prevent overfitting.

Loss Function

Categorical Cross-Entropy to measure the pixel-wise difference between predicted and true masks.

Optimizer

Adam Optimizer for its adaptive learning rate and fast convergence.

Training Configuration

Parameter	Value
Batch Size	`32`
Epochs	`20`
Learning Rate	`0.001`
Validation Split	`20%`

Intelligent Training Callbacks

- **ModelCheckpoint**: To save the best-performing model during training.
- **EarlyStopping**: To halt training when validation performance ceased to improve, preventing overfitting.

The Verdict: High Accuracy Across All Datasets

98.77%
Training Accuracy

98.67%
Validation Accuracy

98.58%
Test Accuracy

Consistent Performance Demonstrates Strong Generalization

The model achieved exceptionally high accuracy scores not only on the data it was trained on, but also on the unseen validation and test sets. This consistency demonstrates that the model successfully learned meaningful **scene features** rather than just memorizing the training data, and it can generalize well to new images.

A Comprehensive Evaluation Beyond Accuracy

To ensure a robust evaluation of segmentation quality, we employed a suite of standard industry metrics:

Mean IoU (Intersection over Union)

Measures the overlap between predicted & ground truth masks.

F1 Score / Dice Coefficient

The harmonic balance between precision and recall.

Precision

The ratio of correctly predicted positive pixels.

Recall (Sensitivity)

The model's ability to find all true positive pixels.

Specificity

The ability to correctly identify negative pixels (background).

True Detection Rate (TDR)

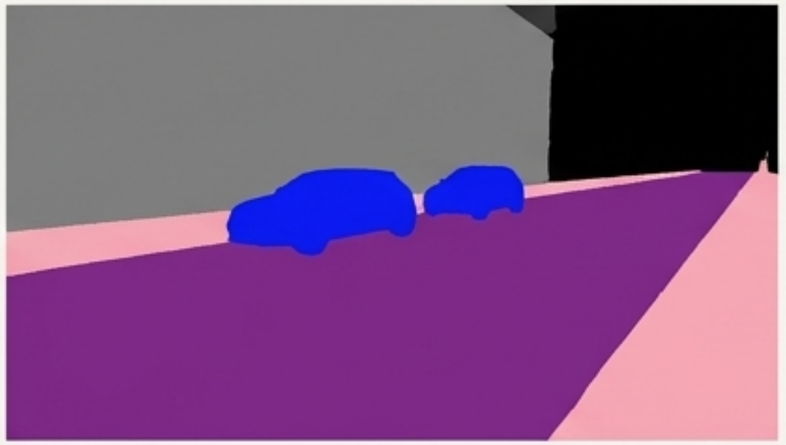
Measures the correct segmentation rate for target objects.

These metrics provide deep insights into both pixel-level performance and class-wise segmentation quality.

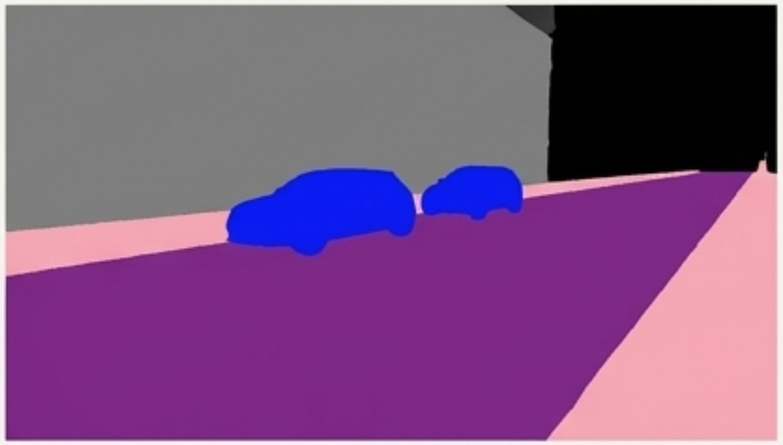
Visualizing the Results: From Raw Image to Segmented Scene

Original RGB Image

Ground Truth Mask



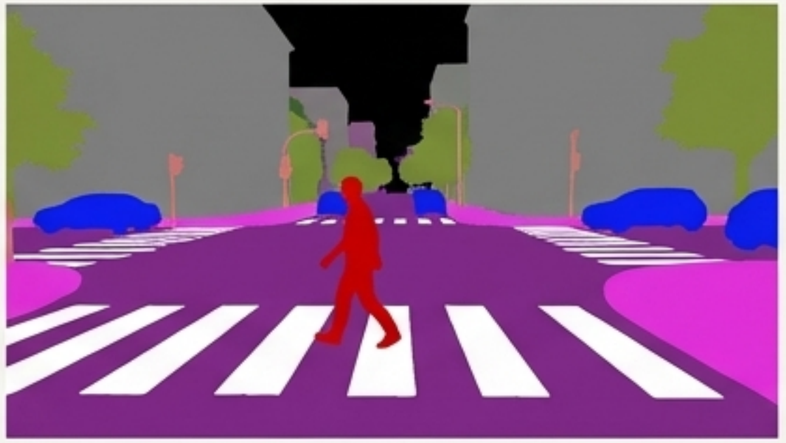
Model's Prediction



Clear separation of vehicles and road.

Original RGB Image

Ground Truth Mask



Model’s Prediction

Accurate detection of pedestrian and sidewalk.

Key Observations on Model Performance

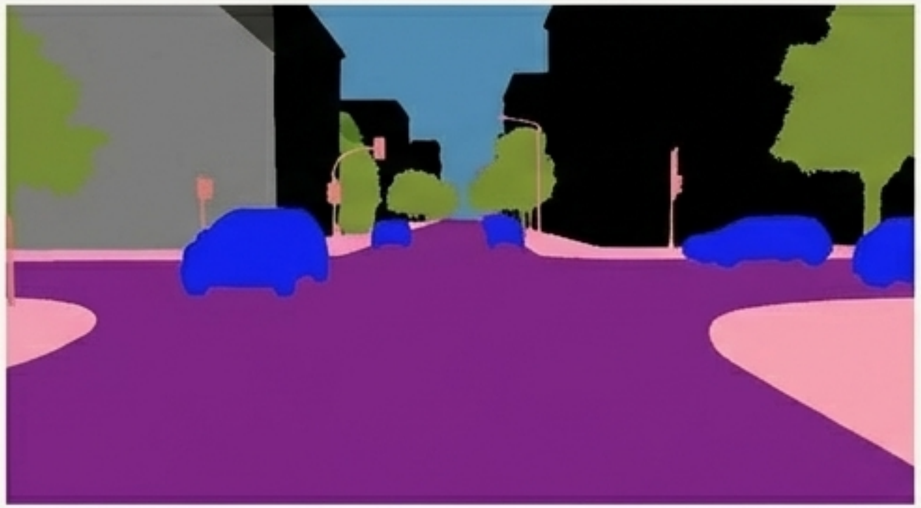
Strengths

- **Structured Objects:** Achieved strong, clean segmentation on large, well-defined objects like roads, buildings, and sky.
- **Vehicle & Pedestrian Detection:** Reliably identified and delineated key dynamic objects critical for safe navigation.
- **Training Stability**: The high quality of the CARLA dataset contributed to a stable and effective training process.

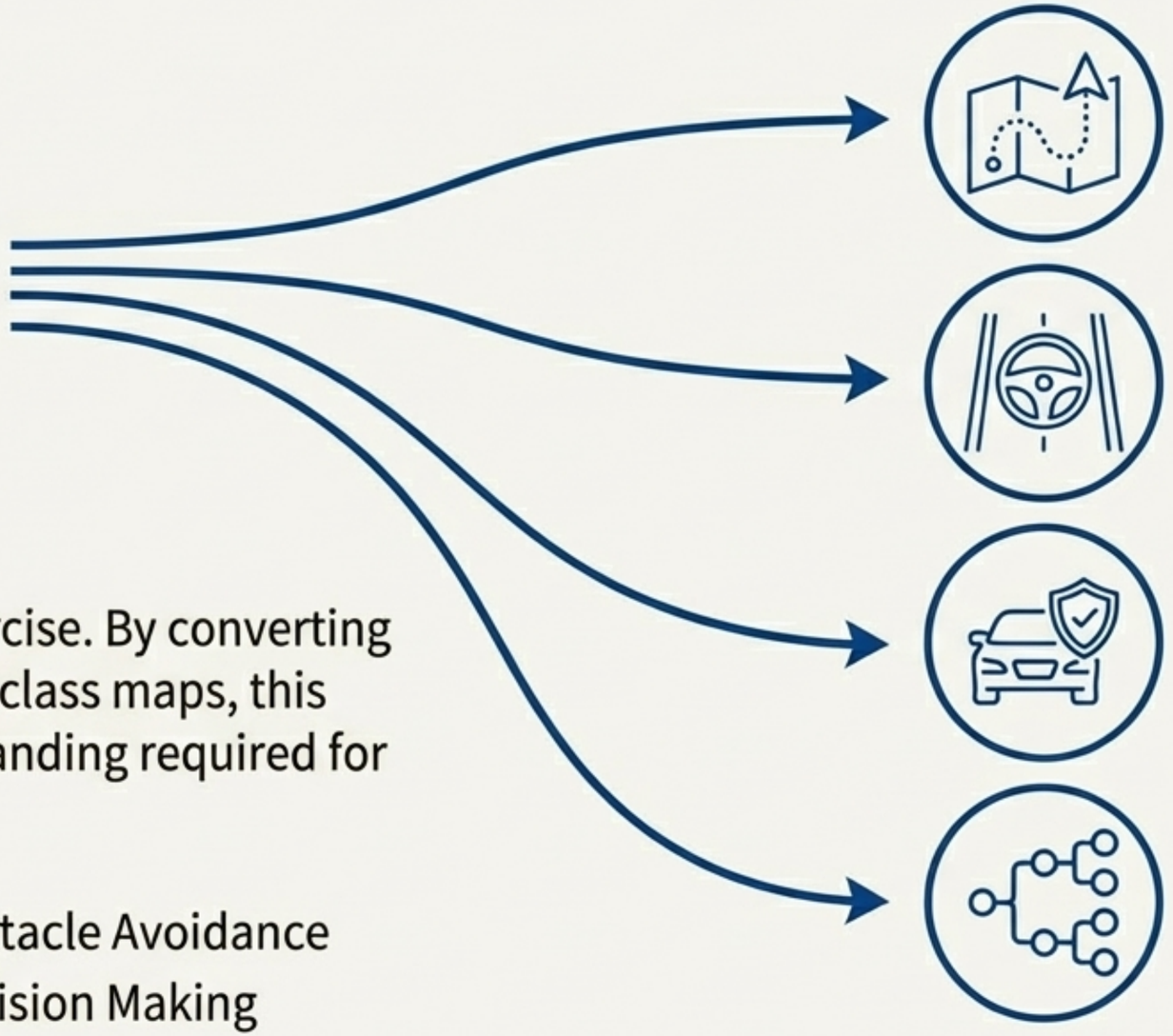
Areas of Nuance

- **Fine Structures:** Showed minor confusion in segmenting very thin or small structures, such as traffic poles and wires, a common challenge in segmentation tasks.

From Pixels to Pathways: The Practical Impact



Segmented Scene Data



This project is more than an academic exercise. By converting raw visual data into structured, pixel-level class maps, this system provides the foundational understanding required for critical autonomous driving tasks:

- Path Planning & Navigation
- Lane Detection and Keeping
- Obstacle Avoidance
- Decision Making

The Roadmap to Real-World Robustness

While the model is highly effective on simulated data, the path to real-world deployment involves several key enhancements:

1 Augment the Data

Introduce artificial variations (brightness, rotation, weather effects) to improve resilience.

2 Refine the Training

Experiment with advanced loss functions (e.g., Dice loss) better suited for imbalanced classes.

3 Evolve the Architecture

Incorporate state-of-the-art models like Attention U-Net or DeepLabV3+ for potentially higher accuracy.

4 Bridge to Reality

Train on challenging real-world datasets and optimize the model for real-time inference on edge devices.

A Successful Implementation of High-Fidelity Scene Segmentation

This project successfully implemented the U-Net architecture for multi-class semantic segmentation on CARLA driving scenes. The model achieved high accuracy and strong segmentation quality, demonstrating that this approach is a powerful tool for enhancing perception in autonomous systems.

By converting raw images into pixel-level class maps, the system provides the critical scene understanding necessary for safer, more intelligent navigation.

Core Project Toolkit

Language & Libraries

Dataset & Environment