**Paper Information**

**Authors**: Ryuto Tsuruta, Kazuyuki Morioka

**Title**: Autonomous Navigation of a Mobile Robot with a Monocular Camera using Deep Reinforcement Learning and Semantic Image Segmentation

**Type (journal/conference)**: Conference

**Publication Date**: 2024

**Number of Citations**: 2

**Contributors**

**Primary Reader**: Xinyang

**Proofreader**: Ben

**Research Focus**

Unlike the traditional navigation based on pre-mapped environments and lidar sensors, the paper research autonomous navigation for mobile robots using a monocular camera as the primary sensor input. With the help of reinforcement learning models(RL) and semantic segmentation technology, this paper aims to improve the performance of robots navigation in dynamic environments, reducing the gap between real-world application and simulation.

**Technical Details**

**Models**: The research employs Proximal Policy Optimization (PPO) for deep reinforcement learning, integrating a semantic segmentation model (U-Net with ResNet50 as the encoder) to process monocular camera images. With this method the robot can understand its surroundings by classifying floor regions and obstacles.

**Datasets**: The experiments used a monocular camera, but they used a ZED2 camera to acquire the dataset for automatic floor segmentation with masks. Totally 191 image pairs were collected（150 used for training, 16 for validation, and 25 for testing.）

**Training**: The Unity ML-Agents toolkit was used to train the DRL model in a simulated indoor corridor environment, where the robot learned to navigate 5 to 8 meters without colliding with obstacles.

**Key Evaluation Metrics**:

* Navigation Success Rate by calculating the Dice loss
* IoU (Intersection over Union) for segmentation accuracy
* Smoothness of motion (linear/angular velocity stability)
* Robustness when test varying camera heights

**Outcomes**

**Model Performance**: The trained PPO model successfully completed **994/1000 episodes**, proving high reliability. The ROS-based system was tested on a real robot, confirming good generalization.

**Semantic Segmentation Accuracy**: The segmentation model achieved a 0.9147 IoU score, indicating high accuracy in distinguishing floor regions from obstacles.

**Impact of Camera Height**: The study found that the camera height will affect navigation performance. Optimal performance was achieved at 33cm, while heights of 44cm and 71cm will lead to unstable velocity outputs and failed turns.

**Obstacle Avoidance Strategy**: When an obstacle is detected, the robot avoids collision by reducing its speed. The model also demonstrates the ability to make smooth turns using continuous angular velocity adjustments.