

Common Practice on Visual Understanding Models for Road Scenes

Task-specific Models

From instance-level perception to global-level understanding

- Semantic logic
- Graph structure
- Contextual understanding

From closed-set to open-set condition

- Anomaly detection
- Open-world understanding

From single modality to multiple modality

- Vision-centric data
- Vision and language data
- Other modalities.

Unified Multi-task Models

Task-specific outputs

- Solving negative transfer
- Efficacy of unified framework
- Including auxiliary tasks

Unified language outputs

- Language-based models
- Language-based benchmarks

Unified Multi-modal Models

LLM functions as sequence modeling

- LLM serves as brain
- Closed-loop system

Cross-modal interaction in VLM

- Attention mechanism
- Q-Former module

Prompting Foundation Models

Textual prompt

Visual prompt

Multi-step prompt

Task-specific prompt

Geometric prompt

Prompt pool