# UMM: A Unified Multi-Modal Model for Low-Level Vision Tasks with Dual-Driven Prompting

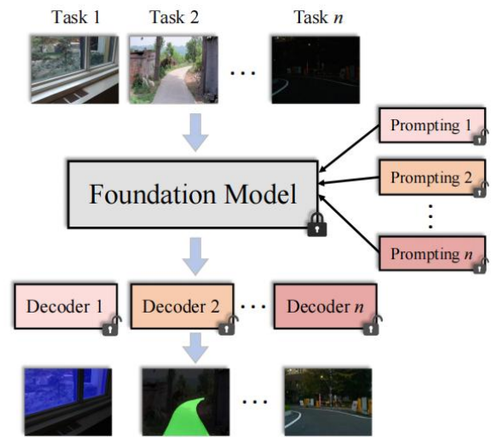
BMVC 2025

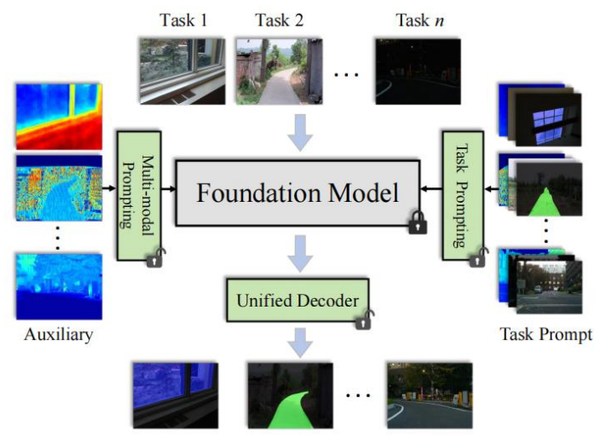
## Motivation

There is a lack of exploration on unified multi-task models for low-level tasks in the visual multi-modal field

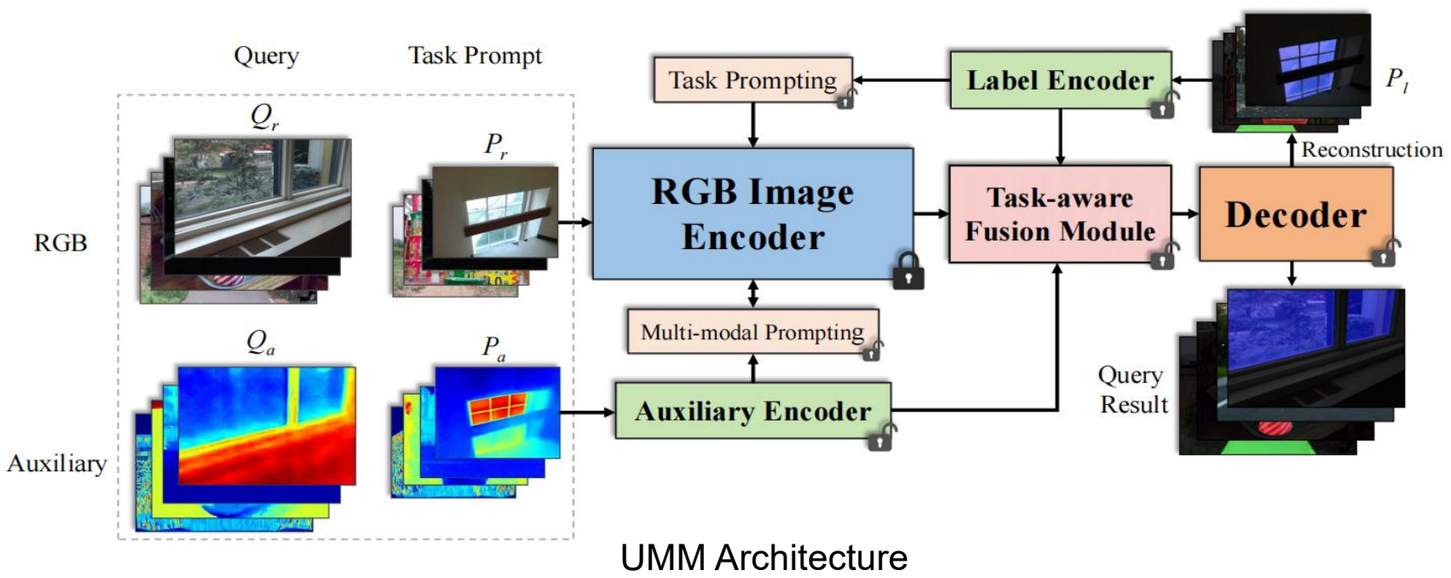
- Existing unified models rely on multiple decoders or task-specific parameters and fail to exploit the shared knowledge among tasks
  → the **dual-driven prompting learning paradigm** is proposed
- Traditional multi-modal fusion methods struggle to adapt to multi-task scenarios and tend to cause modality interference
  → the **Task-Aware Fusion Module** is proposed

UMM exhibits excellent performance in various visual multi-modal low-level tasks

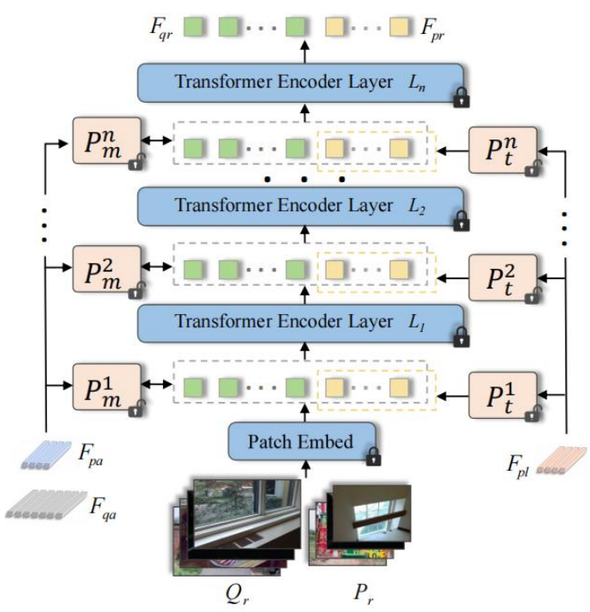


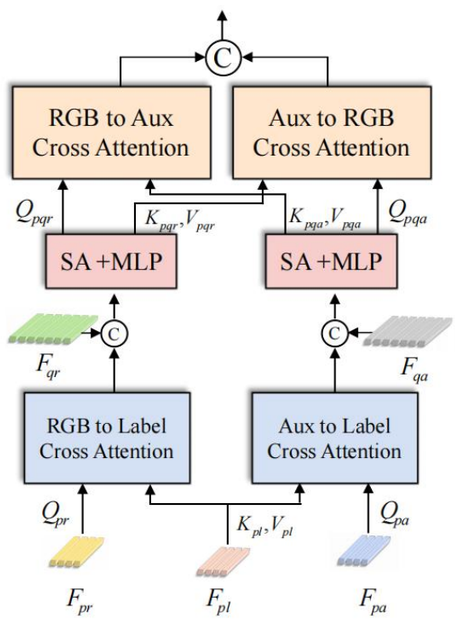

## Approach




UMM Architecture

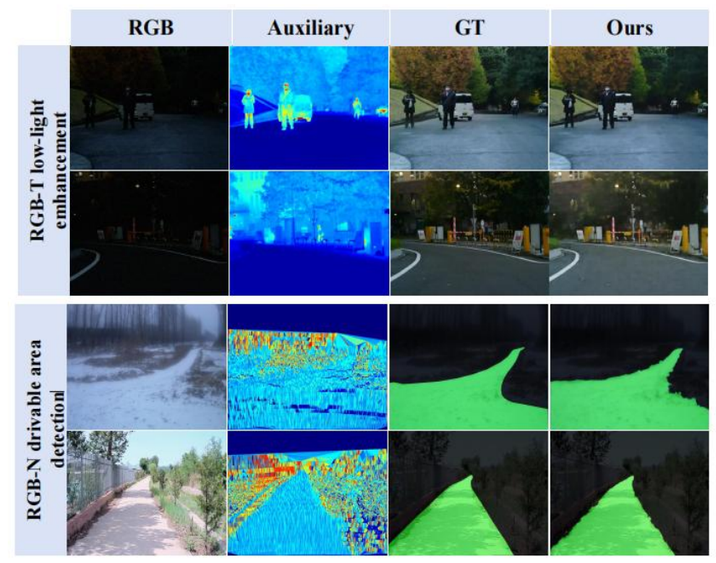


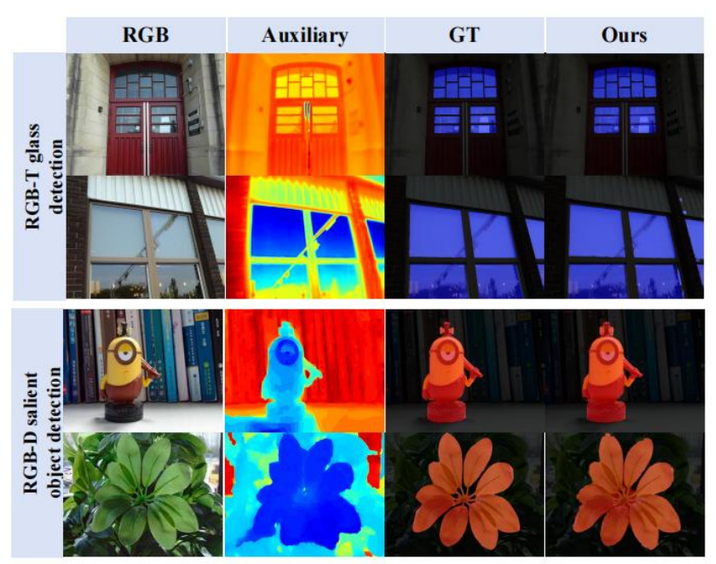

Dual-driven Prompting




Task-aware Fusion Module

## Visualization




## Quantitative Evaluation

| Task | | RGB-T glass detection | RGB-T low-light enhancement | | RGB-D salient object detection | | | | RGB-N drivable area detection |
|---|---|---|---|---|---|---|---|---|---|
| Metrics | | MAE↓ | PSNR↑ | SSIM↑ | $E_\xi$ ↑ | $F_\beta^\omega$ ↑ | $F_\beta$ ↑ | MAE↓ | mIoU↑ |
| RGB-only | VPT [22] | 0.066 | 27.01 | 0.8439 | 0.947 | 0.912 | 0.924 | 0.030 | 49.3 |
| | Adaptformer [4] | 0.060 | 27.32 | 0.8526 | 0.955 | 0.926 | 0.937 | 0.024 | 47.1 |
| Multi-modal | ProTrack [44] | 0.075 | 25.13 | 0.8233 | 0.941 | 0.890 | 0.909 | 0.037 | 79.2 |
| | ViPT [55] | 0.053 | **27.85** | **0.8537** | 0.954 | 0.927 | 0.938 | 0.023 | 67.0 |
| | Ours(ST) | 0.053 | 27.26 | 0.8451 | 0.959 | 0.930 | 0.938 | 0.021 | 69.2 |
| | Ours(JT) | **0.035** | 27.45 | 0.8468 | **0.967** | **0.948** | **0.956** | **0.019** | **93.5** |

Comparison of UMM with other methods

| Task | RGB-T glass detection | RGB-D salient object detection | | | | RGB-N drivable area detection |
|---|---|---|---|---|---|---|
| Metrics | MAE↓ | $E_\xi$ ↑ | $F_\beta^\omega$ ↑ | $F_\beta$ ↑ | MAE↓ | mIoU↑ |
| DAD | - | - | - | - | - | 69.2 |
| GD | 0.053 | - | - | - | - | - |
| DAD+SOD | - | 0.964 | 0.942 | 0.951 | 0.020 | 88.5 |
| GD+SOD | 0.042 | **0.970** | 0.946 | 0.954 | **0.019** | - |
| DAD+GD | **0.041** | - | - | - | - | 93.1 |
| DAD+GD+SOD | **0.041** | 0.966 | **0.947** | **0.956** | **0.019** | **93.5** |

Ablation study of joint training