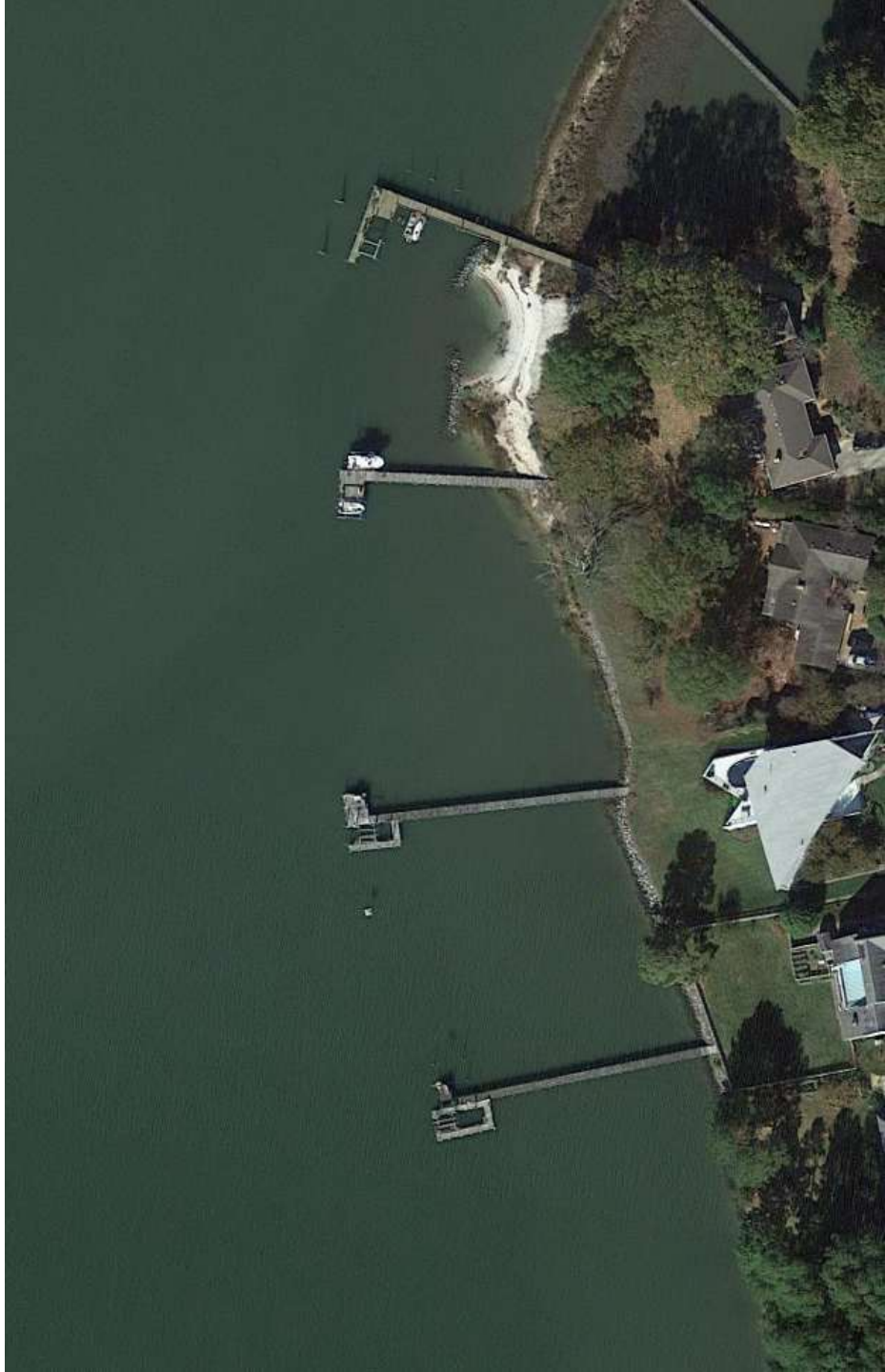
MARS: A Multi-models Agent for Remote Sensing Images

Author: Weijie Cui

Supervisor: Dr. Muhammad Shahzad

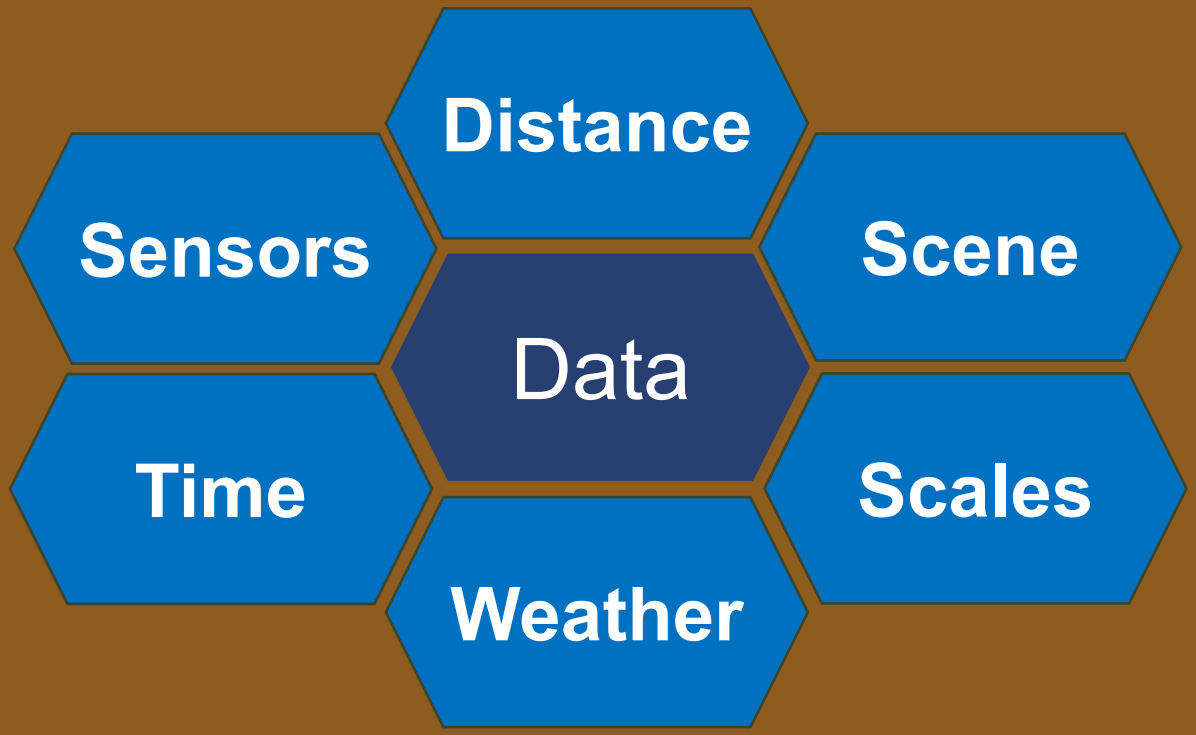


Agenda

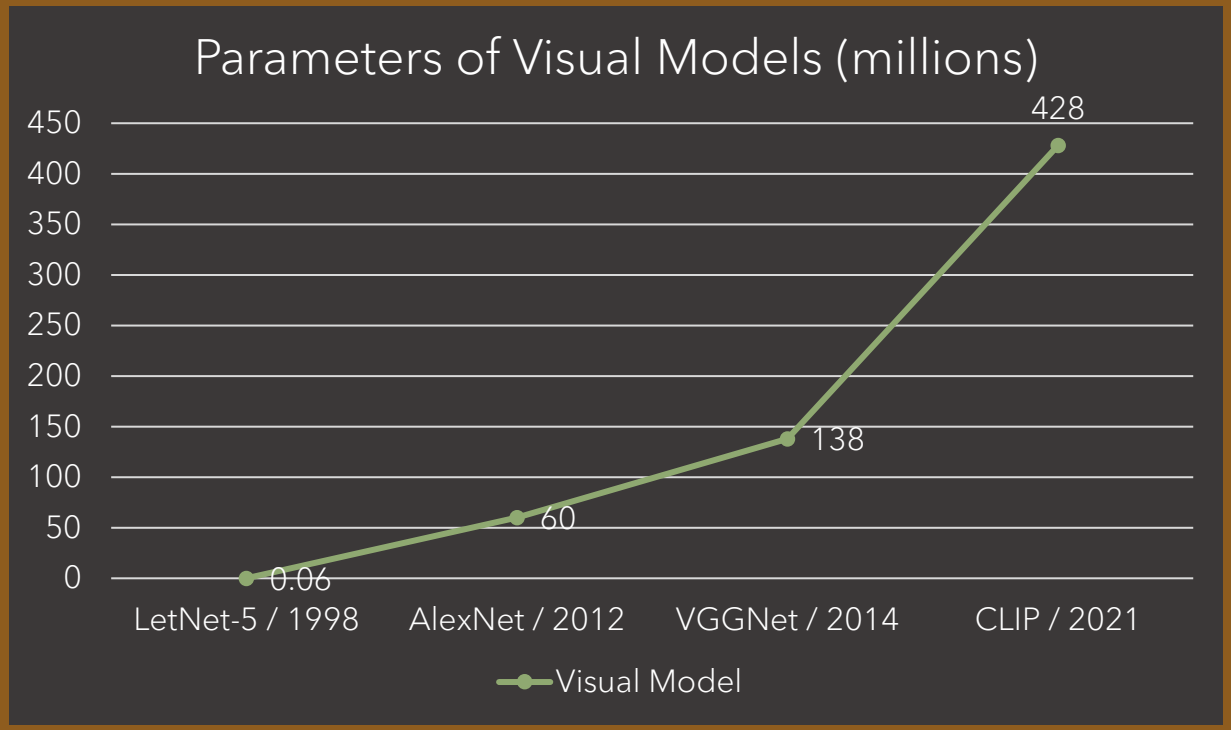
- Background
- Problem Definition
- Literature Review
- Project Aim and Objectives
- Project Methodology
- Results
- Discussions
- Conclusions and Future Work
- Self-Reflection
- Citation of Resources

Background

Remote Sensing (RS) images are vital for environmental monitoring and disaster rescue. Though visual models has already made good progress, they are still struggling with the RS images.



Challenges



Complexity

Problem Definition

- Could Agent improve the accuracy of remote sensing image analysis?
- Can we enhance visual capabilities without increasing computing resource usage?

Literature Review

Visual Models
(Detection, Segmentation)
- YOLO [1] (small & efficient)
- SAMRS [2] (big & complex)

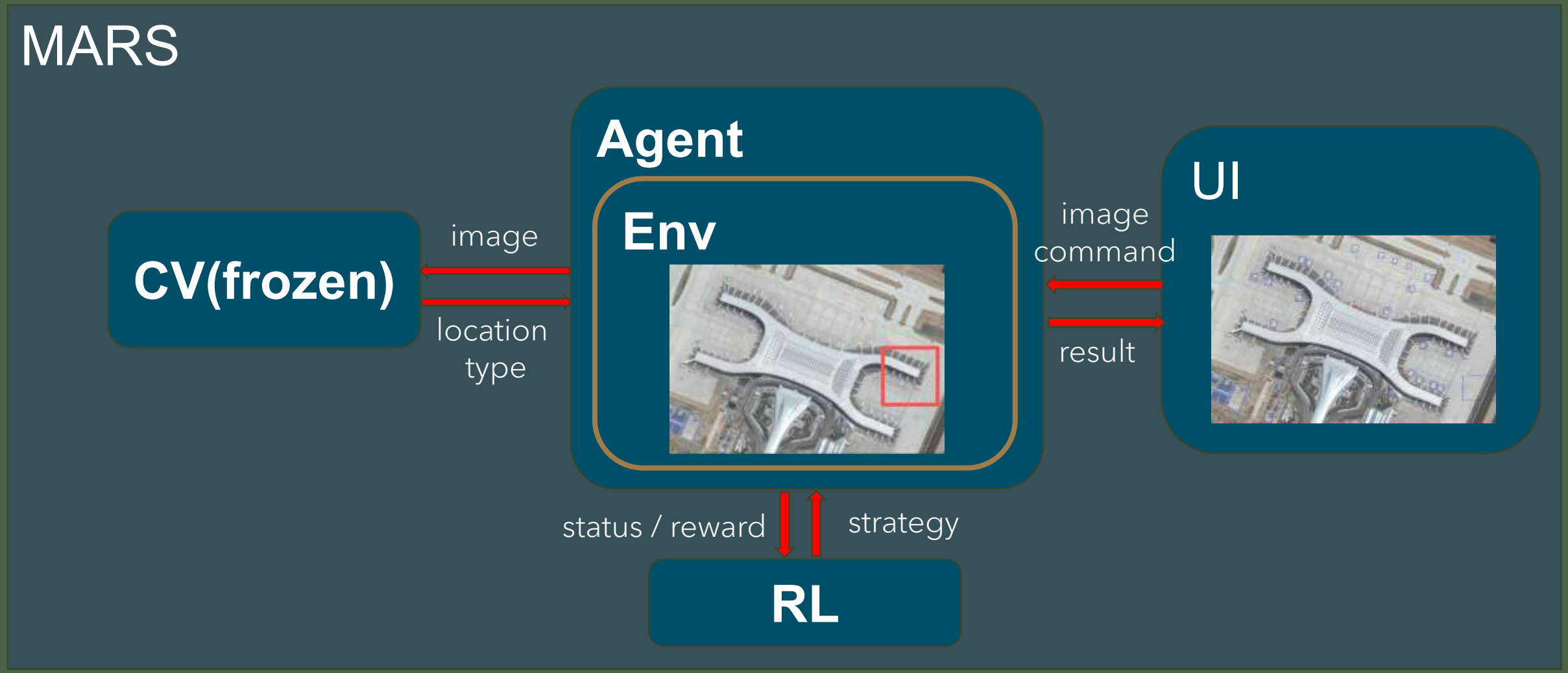
Reinforcement Learning
(Decision-making)
- Scale-Aware classification Network (SAN) [3]

Research Gap
- Poor scalability for detecting objects of various sizes.
- Requires a large amount of computing resources.
- Not suitable for applications of IoT in Remote Sensing.

Project Aim and Objectives

MARS	Reinforcement Learning + Environment + Vision Model
Evaluation	Intersection of Union, Accuracy
Potentiality	Small or partial obscured target

Methodology - Software Structure

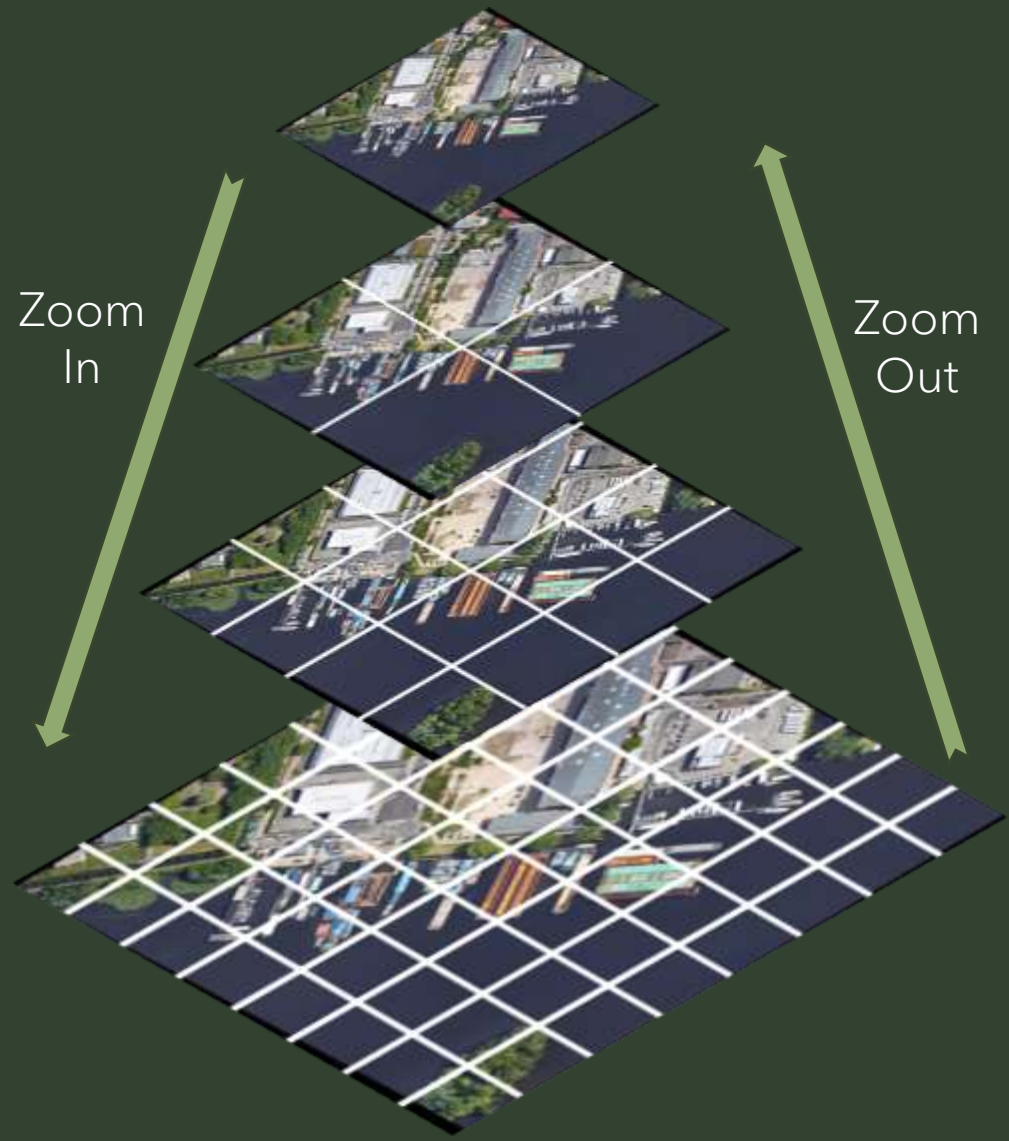


Methodology – Environment

Responsibilities: a source image, current window, clipped images and action spaces.

Reward pseudocode:

if found objects: confidence + 1
else if zoom in: +0.1
else: -0.2



Methodology – Reinforcement Learning

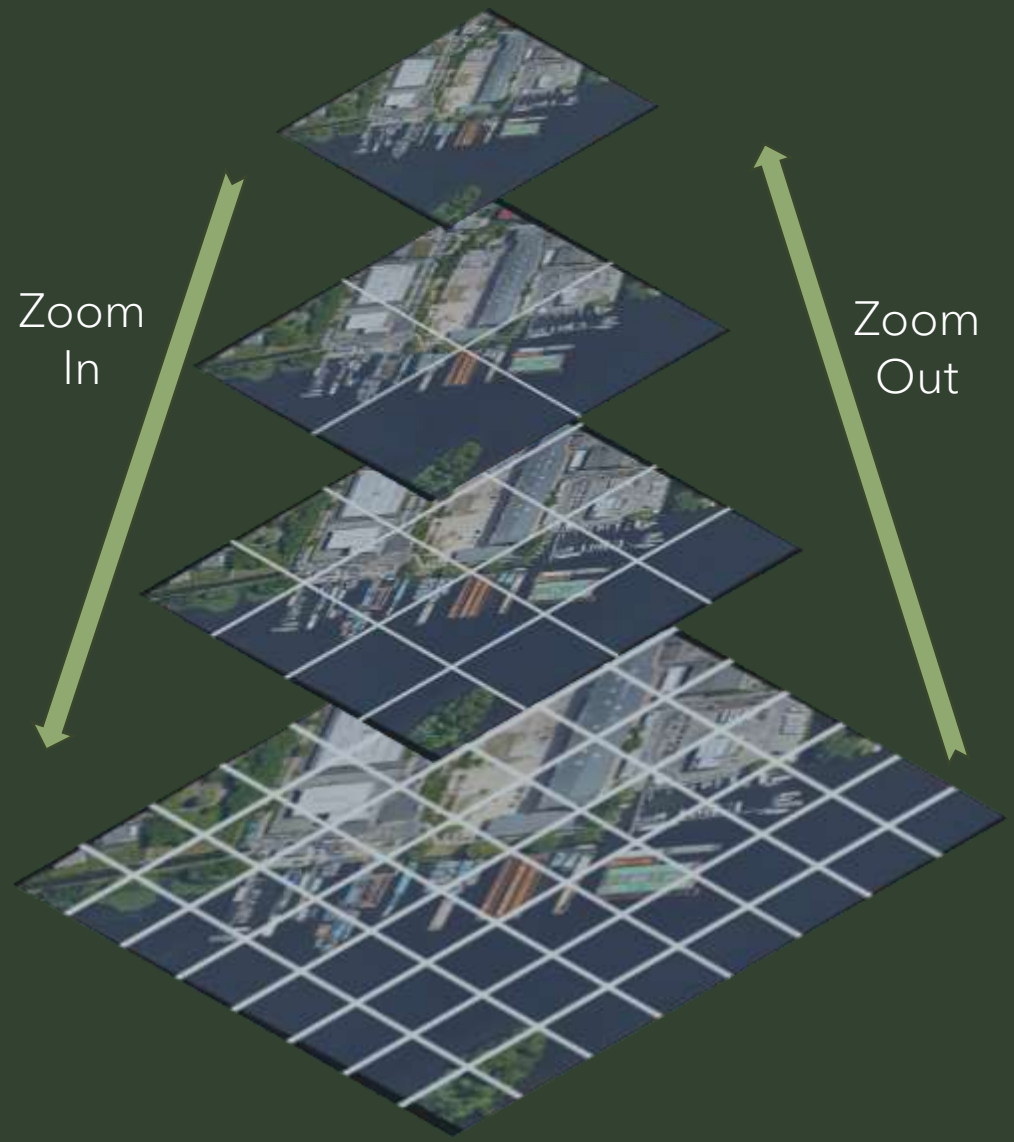
Model: Q-table

Action Space:

- Left, Up, Right, Down
- Zoom In, Zoom Out

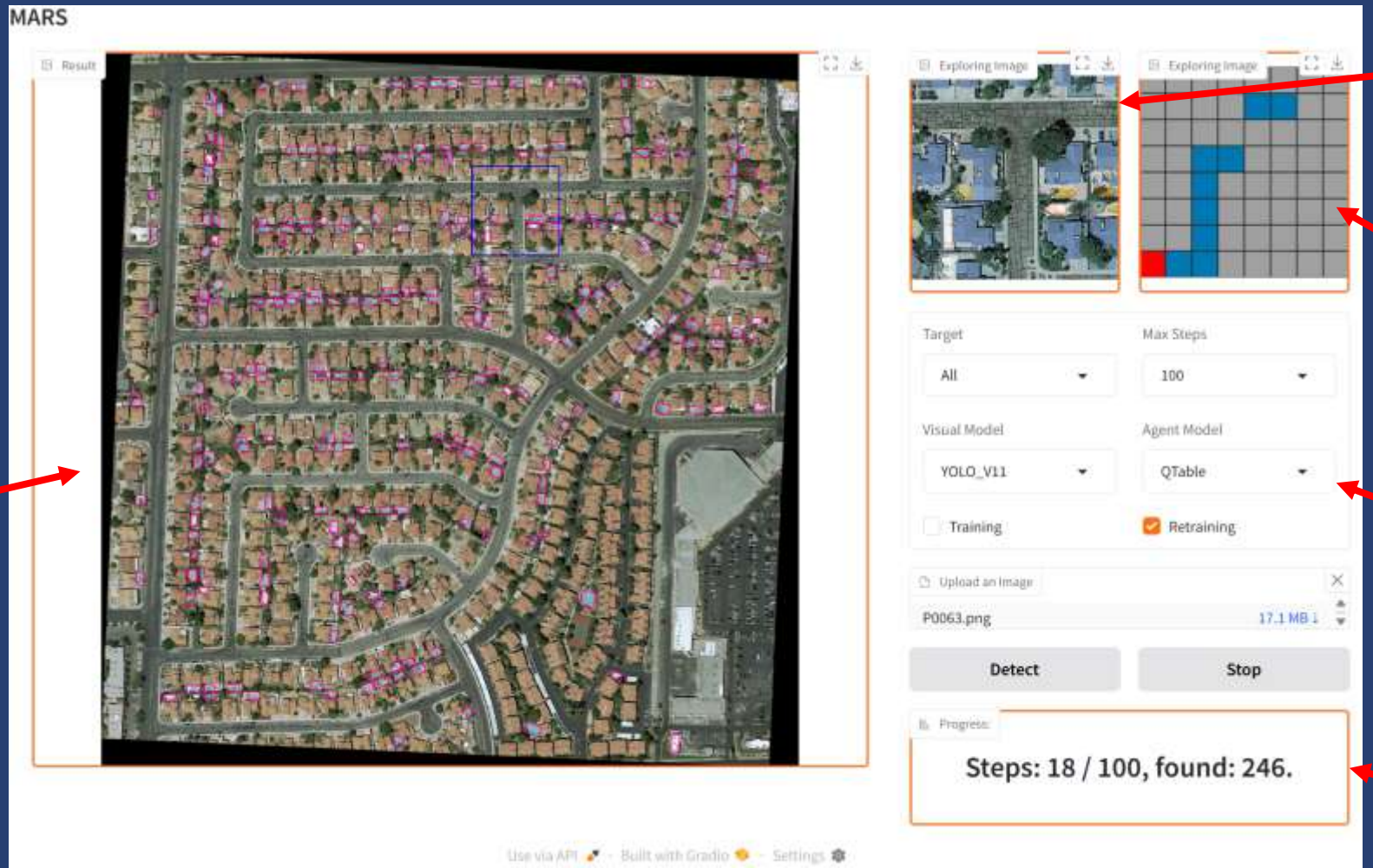
Q-value Equation:

*Q = Q + alpha * (Reward + gamma * Max-Next – Q)*
eps=0.15, alpha=0.5, gamma=0.9.



Methodology – UI

Application framework: Gradio.



Window

Map

Results

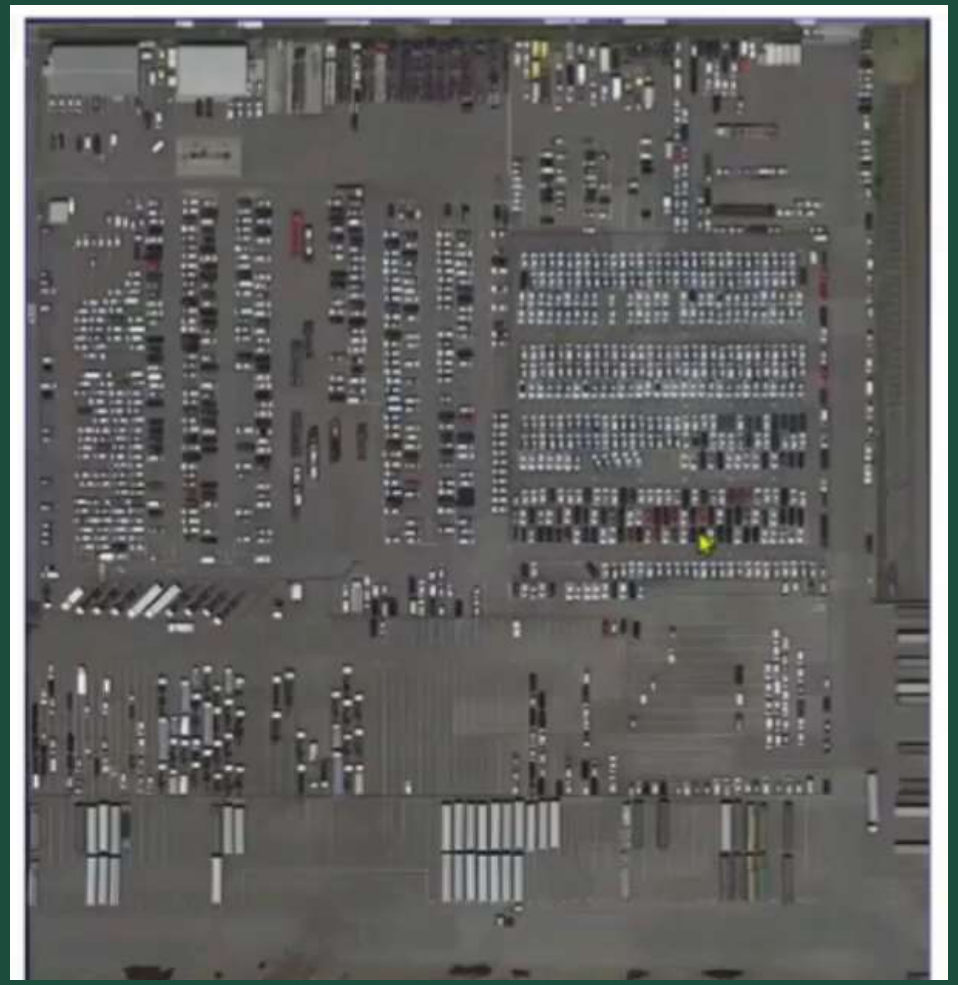
Control Panel

Messages

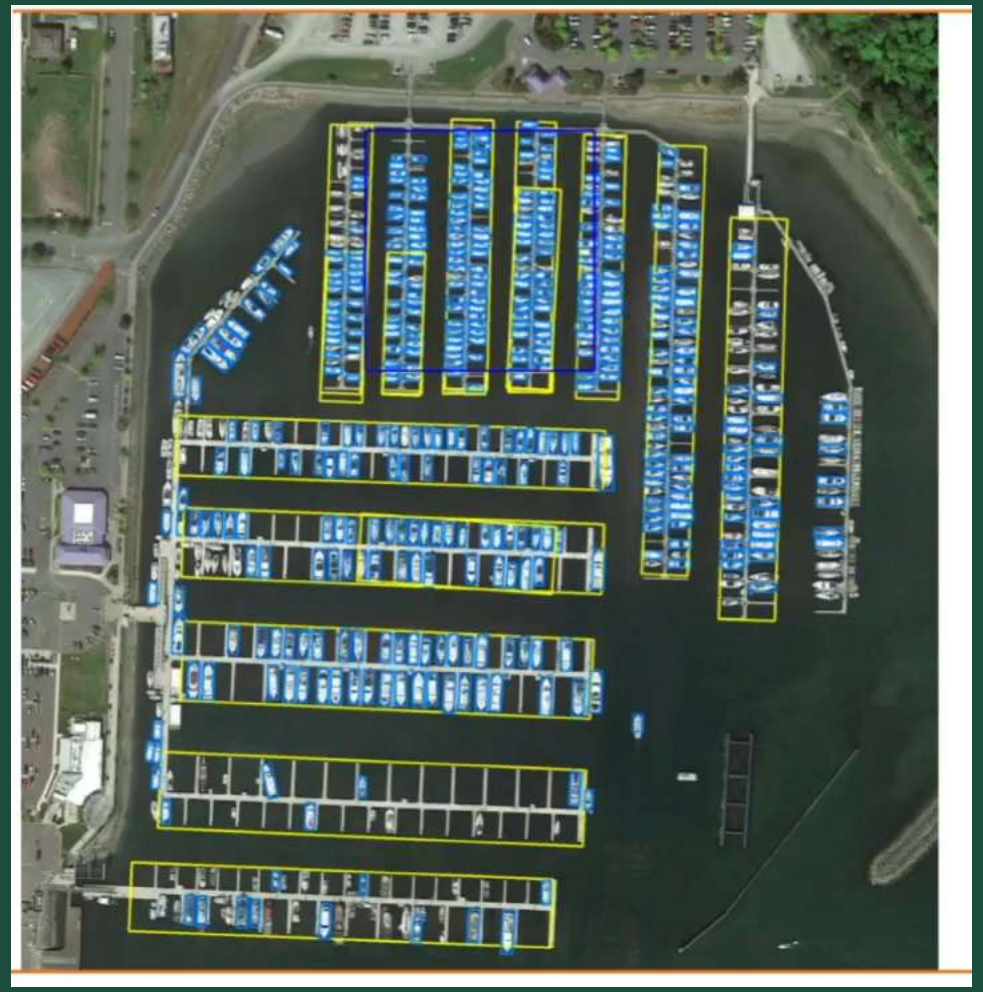
Results

- Size insensitive: magnification levels: x1, x2, x4, and x8.
- Partially obscured √
- Small targets √
- Shadowed targets √

Results – More Samples



A carpark with large and small vehicles



A lake with docks and ships

Discussions

- **Contribution**: MARS built a visual search environment and verified the feasibility of integrating agents and visual models.
- **Limitations**: Currently, only YOLO model is integrated and focuses on classification and positioning tasks.

Conclusions and Future Work

- MARS improved the visual model's recognition accuracy, achieving greater detail without upgrading GPU resource.

Next:

RL + CV — P1

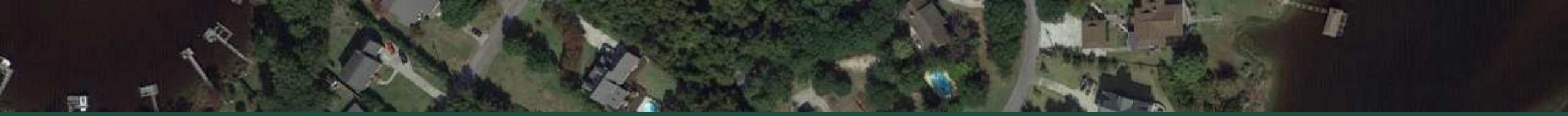
P2 — Visual models for disaster rescue

IoT — P3

Self-reflection

- MARS Framework
  - Setting up Simulation Environment
  - Training Reinforcement Learning Model
  - Integrating and Valuating MARS

- Unfinished Work
  - Quantitative evaluation of performance



Citation of Resources

[1] Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024). YOLOv10: Real-Time End-to-End Object Detection. https://arxiv.org/pdf/2405.14458 .
[2] Wang, D., Zhang, J., Du, B., Xu, M., Liu, L., Tao, D., & Zhang, L. (2023). SAMRS: Scaling-up Remote Sensing Segmentation Dataset with Segment Anything Model. ArXiv.org. https://arxiv.org/abs/2305.02034 .
[3] Liu, Y., Zhong, Y., Shi, S. and Zhang, L. (2024). Scale-aware deep reinforcement learning for high resolution remote sensing imagery classification. ISPRS Journal of Photogrammetry and Remote Sensing, [online] 209, pp.296–311. doi:https://doi.org/10.1016/j.isprsjprs.2024.01.013.

Thank you!

- Author: Weijie Cui
- Supervisor: Dr. Muhammad Shahzad
- Email: fk824514@student.reading.ac.uk
- Git: https://github.com/WeijieCui/MARS