UAV Trajectory Design for Emergency and Disaster Management Network

Phetxomphou Zalita – m5282024

Supervisor: Prof. Pham Tuan Anh

Computer Communication Lab., The University of Aizu.

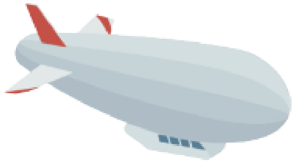
2025.07.10

Outline

1. Research Background
2. Research Proposal
3. Research Progress

UAVs: Transforming Industries, Expanding Human Capabilities

What are UAVs?



Unmanned Aerial Vehicles (UAVs), commonly known as drones, are remotely controlled or autonomous aircraft.

Why UAVs Matter?

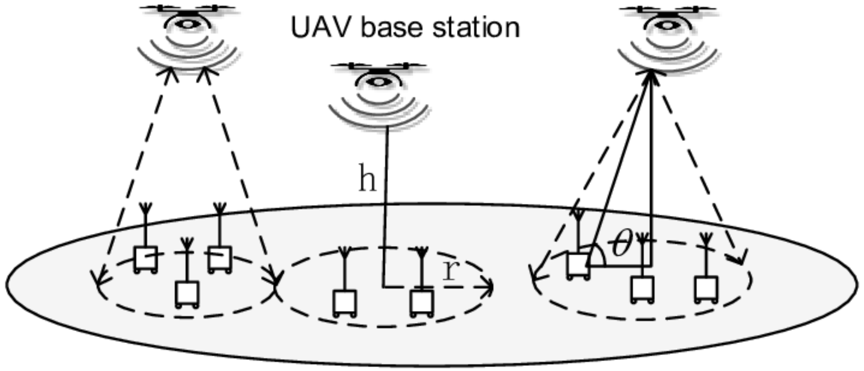
UAVs offer unique capabilities to perform tasks with enhanced speed, safety, and efficiency, especially in hazardous, remote, or data-intensive environments.

Their Diverse Applications:

- Agriculture: Crop monitoring, spraying, and yield assessment
- Logistics: Fast and contactless delivery
- Surveillance: Border security, traffic monitoring, and disaster response
- Filmmaking: Aerial cinematography and photography
- Infrastructure: Inspection of bridges, towers, and power lines
- Environmental Monitoring: Wildlife tracking and air quality assessment

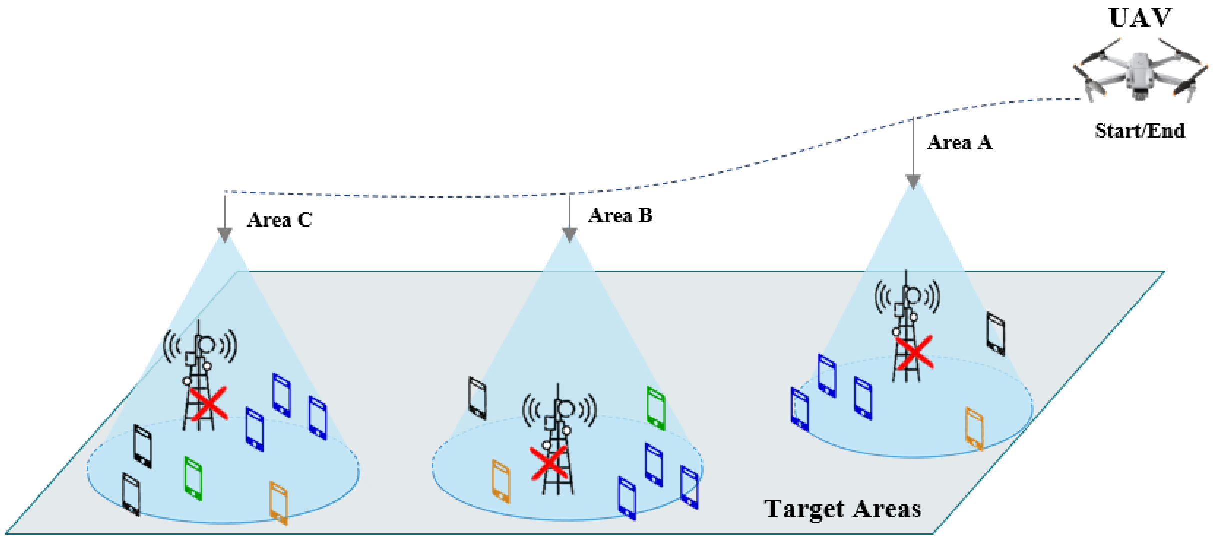
Motivation: Enhancing Emergency Connectivity with UAVs

- Natural disasters critically impact terrestrial communication infrastructure, leaving communities isolated and at risk.
- UAVs serve as rapidly deployable aerial base stations (UAV-BSs), providing temporary, flexible connectivity in compromised or hard-to-reach areas.
- Their mobility and altitude offer a unique advantage for restoring critical communication links, especially where ground infrastructure is damaged or unavailable.



UAV Trajectory Design Impacts Emergency Services

- Some people might just need to send a quick "I'm safe" text, others a crucial video call to emergency services, and some might try to download urgent evacuation maps.
- The UAV has limited coverage, trajectory optimization is crucial ensuring reliable services for all users in each urgency areas.
  - ➢ If the UAV leaves early, it cuts off critical communications for the users it's currently serving
  - ➢ if the UAV stays too long, it wastes valuable time that could be spent serving other critical areas



Users in each region has different amount of data to send.

Outline

1. Research Background
2. Research Proposal
3. Research Progress

Inspired Approaches: Foundational Concepts

Trajectory Planning with Multi-Agent PPO [1]

Inspired from: Hoang et al., IEICE ComEx 2024

Framework:

Centralized Training, Decentralized Execution (CTDE) using PPO agents
UAVs are trained to utilize local observations and a shared reward

Objective: Maximize number of users meeting data rate threshold

User satisfaction is binary.

$$\delta\{r_{n,t} \geq r_{\text{th}}\} \triangleq \begin{cases} 1, & \text{if } r_{n,t} \geq r_{\text{th}} \\ 0, & \text{otherwise.} \end{cases}$$

$$r_{n,t} = \sum_{k=1}^{M} \alpha_{n,k,t} \cdot r_{n,k,t}^{\text{UAV}} + \beta_{n,t} \cdot r_{n,t}^{\text{mBS}}$$

where $r = W \log_2 (1 + \gamma)$

QoE-based Evaluation [2]

Inspired from: Dai et al., Electronics Letters, 2024

Key Contribution:

- Mean Opinion Score (MOS) maps R to user satisfaction levels
- Their goal is to design the best deployment of multiple UAVs
- Game-theoretic strategy is used to optimize UAV positions and power.

Objective:

Maximize the sum-QoE $\sum_{k=1}^{k} MOS(R_k(w_{mk}, p_{mk}, k))$

MOS Formula:

$$MOS(R) = \begin{cases} 5 & a\log_{10}\frac{R}{b} \geq 4.5, \\ 4 & 3.5 \leq a\log_{10}\frac{R}{b} < 4.5, \\ 3 & 2.5 \leq a\log_{10}\frac{R}{b} < 3.5, \\ 2 & 1.5 \leq a\log_{10}\frac{R}{b} < 2.5, \\ 1 & a\log_{10}\frac{R}{b} < 1.5, \end{cases}$$

Where:
R is user rate

$$a = \frac{3.5}{\log_{10}(\theta_2/\theta_1)}, \; b = \theta_1 \left(\frac{\theta_1}{\theta_2}\right)^{\frac{1}{3.5}}$$

θ_1 : minimum acceptable rate
θ_2 : recommended rate

Example Calculation:

$\boldsymbol{\theta_1 = 3\ Mbps\ \&\ \theta_2 = 10\ Mbps}$

→ $a = \frac{3.5}{\log_{10}\left(\frac{10}{3}\right)} = 6.69$

→ $b = 3\left(\frac{3}{10}\right)^{1/3.5} = 2.13$

Assume R = 5Mbps

$x = 6.69 \log_{10} \frac{5}{2.13} \approx 2.47$

Then it falls on MOS = 2
Because $1.5 \leq 2.47 \geq 2.5$

Proposed: QoE-Driven UAV Trajectory Optimization

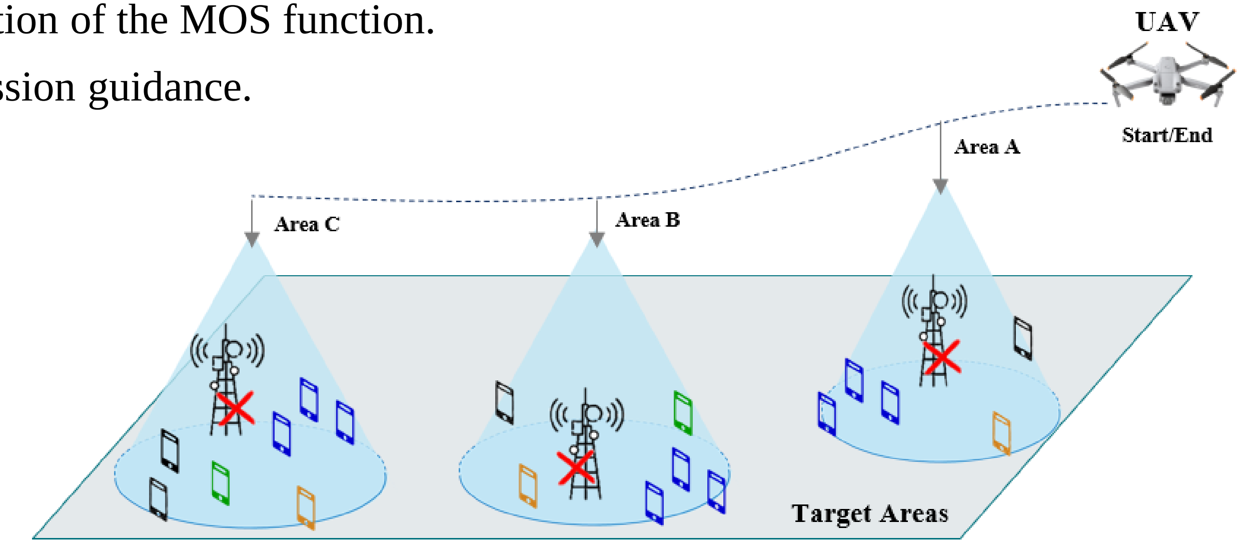
Objective: To learn an optimal trajectory for a single UAV to:

- Maximize user QoE across multiple geographically distinct regions.
- Ensure safe and timely return to home base within operational limits.

Approach: Reinforcement Learning (PPO) for single agent trajectory control.

Key Pillars of Our Solution:

- <u>Customized QoE Model</u>: Utilizing a specific calibration of the MOS function.
- <u>Dynamic Phase-Based Rewards:</u> For multi-stage mission guidance.



Users in each region has different amount of data to send.

Proposed: QoE Model - From Data Rate to MOS

Customized QoE Model: From Data Rate to MOS Maps physical data rate to human perception.

Continuous Score (x):

Original paper's 'a' was 3.5 (4.5 - 1.0 = 3.5). They consider MOS 1 as its lowest baseline for 'minimum acceptable rate'. Their purpose is to maximize the sum of QoE, having MOS running from 1 to 5..

$$x = a \cdot \log_{10}\left(\frac{R}{b}\right) \qquad \text{where} \qquad a = \frac{2.0}{\log_{10}(\theta_2/\theta_1)} \qquad b = \frac{\theta_1}{10^{(2.5/a)}}$$

Our calibration uses 2.0 (4.5 - 2.5 = 2.0) to span from MOS 3 to MOS 5. We consider MOS 3 as our baseline. Our UAV optimizes its trajectory to maximize MOS ≥3 for every serving user.

The calibration constants a and b are derived to align with key service thresholds:

- θ_1: Data rate for "Good" service (e.g., 2 Mbps), mapping to $\mathrm{MOS} = 3$.
- θ_2: Data rate for "Excellent" service (e.g., 5 Mbps), mapping to $\mathrm{MOS} = 5$.

Discrete MOS: Based on thresholds of x:

$$\mathrm{MOS}(R) = \begin{cases} 5 & \text{if } x \geq 4.5 \\ 4 & \text{if } 3.5 \leq x < 4.5 \\ 3 & \text{if } 2.5 \leq x < 3.5 \\ 2 & \text{if } 1.5 \leq x < 2.5 \\ 1 & \text{if } x < 1.5 \text{ and } R > R_{\text{unconnected}} \\ 0 & \text{if } R \leq R_{\text{unconnected}} \end{cases}$$

Introduce MOS 0 for unconnected

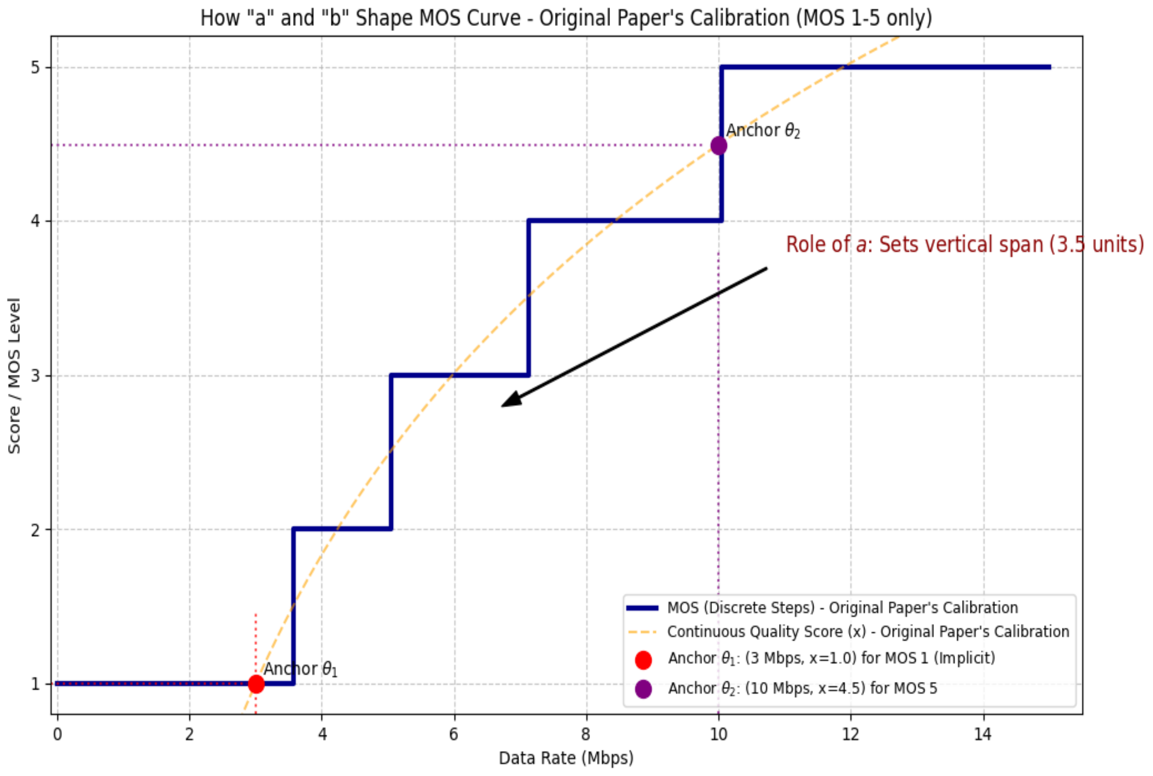
Where $R_{\text{unconnected}}$ is the minimum data rate to be considered connected.

Why MOS Over Direct Data Rate?
This approach prioritizes human experience and system efficiency.

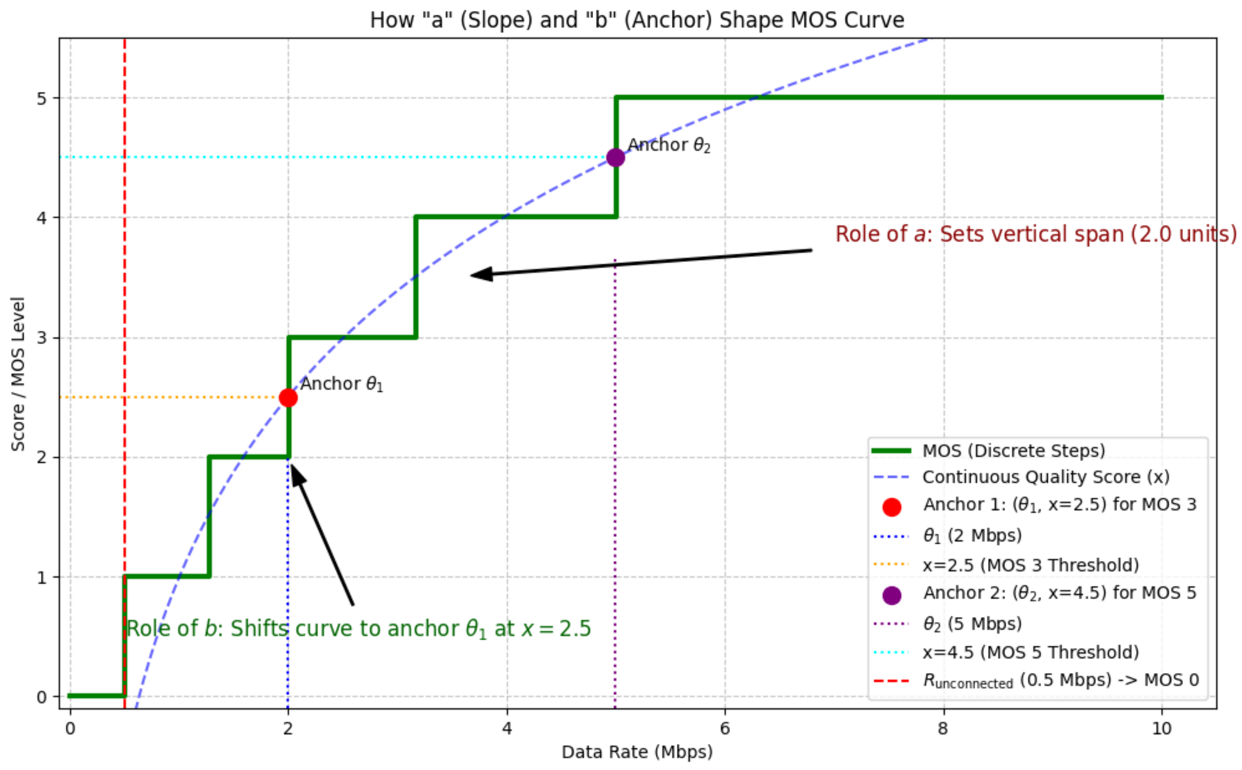
- **You Don't Feel Every Bit:** After a certain point, higher data rates don't improve the perceived quality.
- **Ensuring Usable Service for All:** Focusing solely on total raw data can create "selfish" agents that give extreme speeds to a few, while many others are left with poor or no service.
- **MOS Delivers True Quality:** MOS measures what people actually perceive as good service.

Comparison: MOS Calibration Approaches

Their case: They consider **MOS 1** as Theta_1
For their minimum acceptable rate



Our case: We consider **MOS 3** as Theta_1
for our minimum acceptable rate
(align with our target for “Good Service”)



Proposed: Phase-Based Reward Function

Phase-Based Reward Function (Rt) Dynamically adapts to mission objectives, integrating MOS.

a "Multi-Region Return-to-Home" mission.

Phase 1: Serving Users (while regions unserved)

$$R_t = \underbrace{w_{serve} \cdot N_t^{inc}}_{\text{Satisfied users with target MOS}} + \underbrace{w_{move} \cdot (d_{t-1}^{CoM} - d_t^{CoM})}_{\text{Effective location}} + \underbrace{B_{region} \cdot I_t^{region}}_{\text{Regional completion}} - P_{total}$$

Phase 2: Return to Base (after all users served)

$$R_t = w_{move} \cdot (d_{t-1}^{home} - d_t^{home}) - P_{distance} - p_{base}$$

Terminal Rewards:

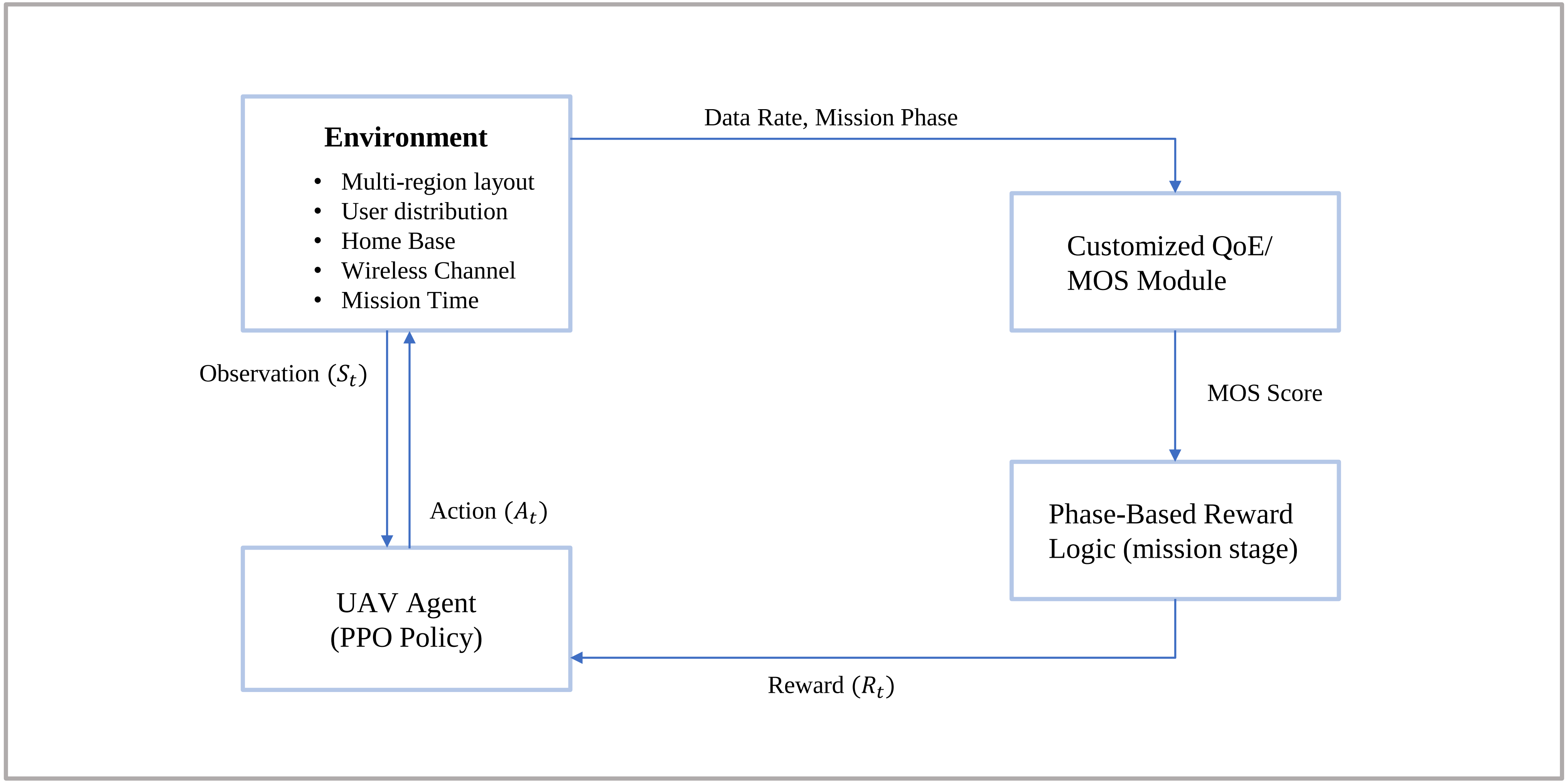
- Success: Large bonus if UAV returns home within time limit.
- Failure: Catastrophic penalty if time limit reached before mission complete.

Table: Parameters of the Phase-Based Reward Function

Parameter	Explanation
R_t	Total reward at timestep t.
w_{serve}	Weight for serving reward.
N_t^{inc}	Number of users with improving MOS (≥ 3).
w_{move}	Weight for guidance reward.
d_{t-1}^{CoM}	Distance to Center of Mass of unserved users at $t-1$.
d_t^{CoM}	Distance to Center of Mass of unserved users at t.
B_{region}	Bonus for completing a region (5x for final region).
I_t^{region}	Indicator (1 if region completed, 0 otherwise).
P_{total}	Combined step and urgency penalties in Phase 1.
d_{t-1}^{home}	Distance to home base at $t-1$.
d_t^{home}	Distance to home base at t.
$P_{distance}$	Penalty for distance from home.
p_{base}	Step penalty in Phase 2.
$B_{terminal}$	Large bonus for mission success (terminal reward).

Overall System Design

Reinforcement Learning Framework



Expected Contribution & Implementation Plan

Expected Contribution:

- Development of a QoE-driven UAV trajectory optimization framework tailored for complex multi-region, return-to-home missions.
- Design and implementation of a Customized QoE/MOS model with a specific calibration for user satisfaction evaluation.
- Proposal and implementation of a Phase-Based Reward Function to effectively guide multi-stage UAV behaviors.
- Demonstration of the framework's effectiveness in maximizing user satisfaction and mission completion in simulated emergency scenarios.

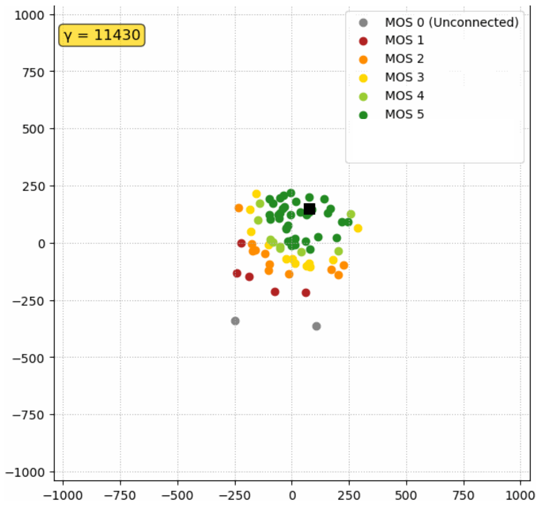
Implementation Plan

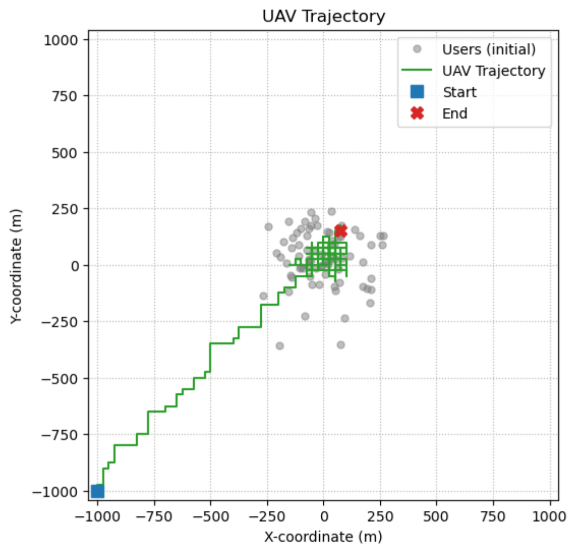
Phase	Objective	RL Strategy	Implementation Plan
Phase I	UAV Trajectory Optimization based on QoE Evaluation (Current Work)	PPO	- Maximizing MOS in a single region. - Extending to multi-region scenarios with operational constraints.
Phase II	Advanced Emergency Zone Handling & Obstacle Integration (Future Work)	MAPPO	- Expand to Multi-Agent PPO (MAPPO) for fleet coordination. - Apply QoE-based reward using MOS for multiple urgency regions - Extend to support multi-level priority in QoE. - Integrate link-impacting obstacles into the environment.

Outline

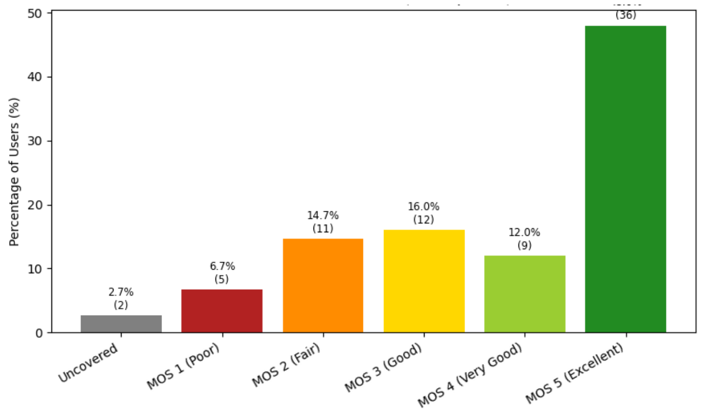
1. Research Background
2. Research Proposal
3. Research Progress

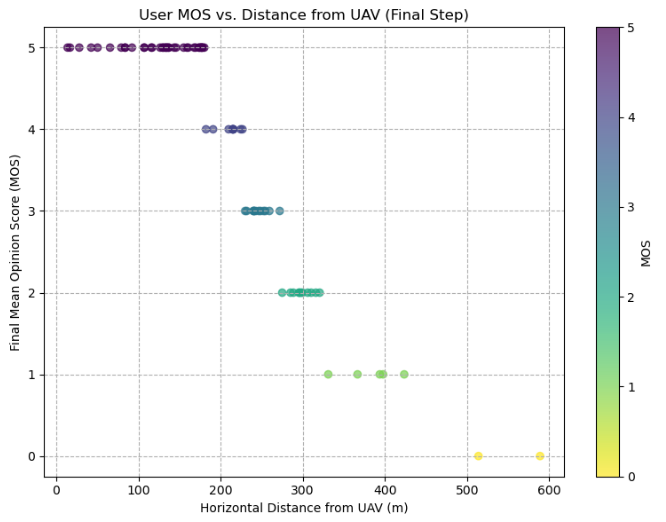
Research Progress: Phase I - Single Region MOS Maximization

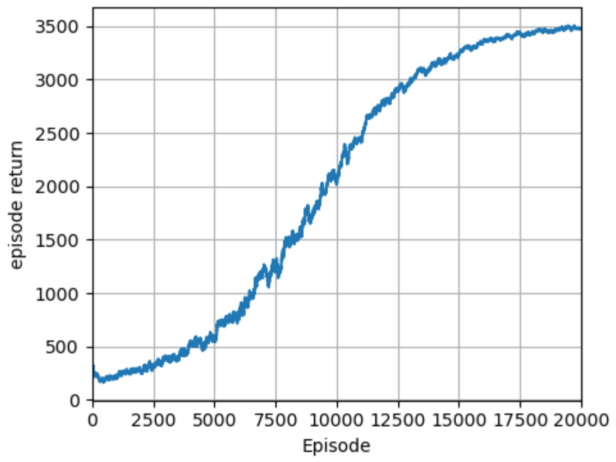


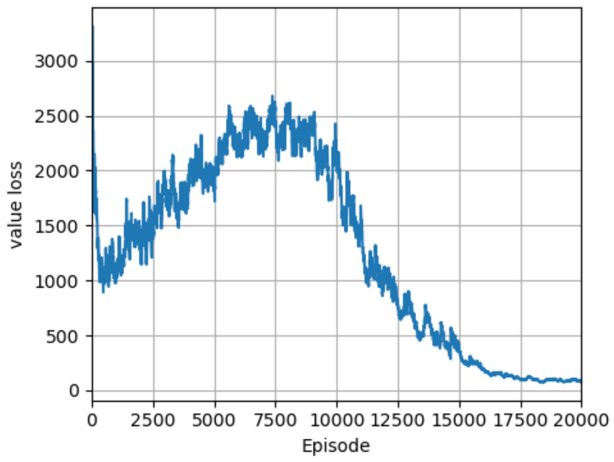


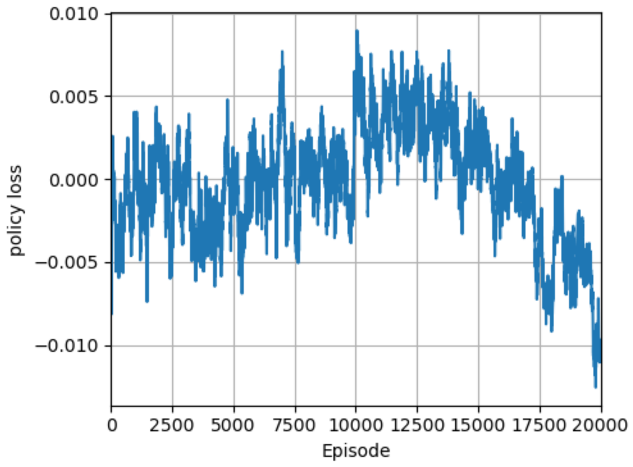
Note: User distribution is wider than UAV's converge range, to test UAV can find the best placement and maximize MOS.

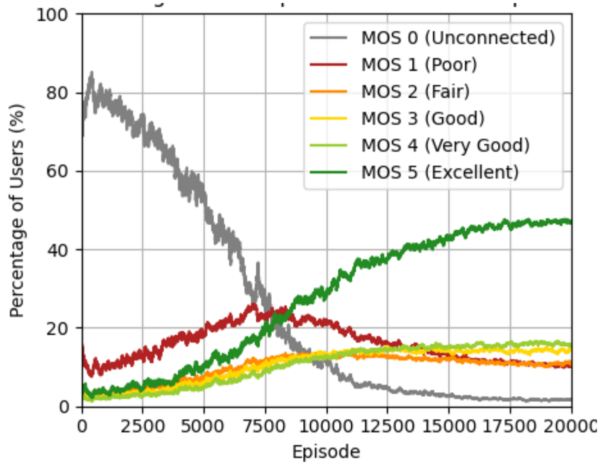






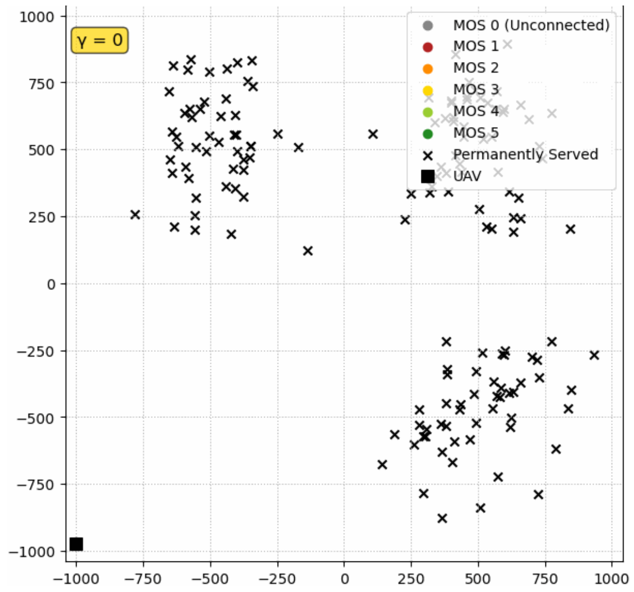


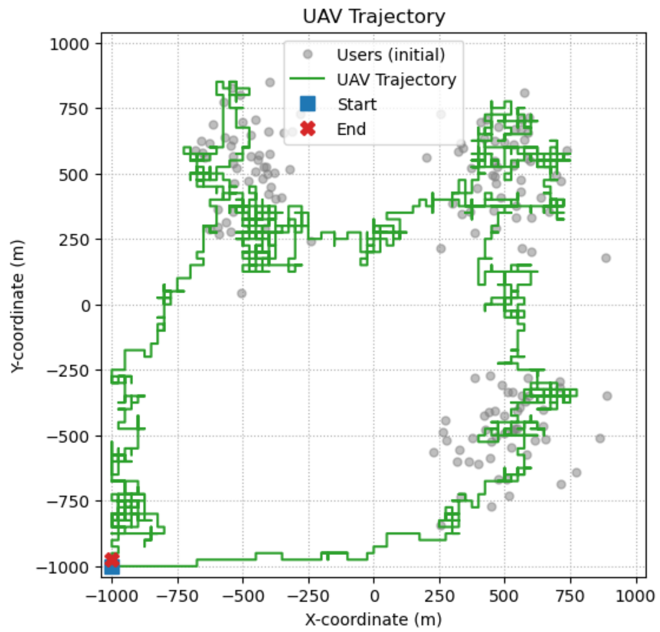


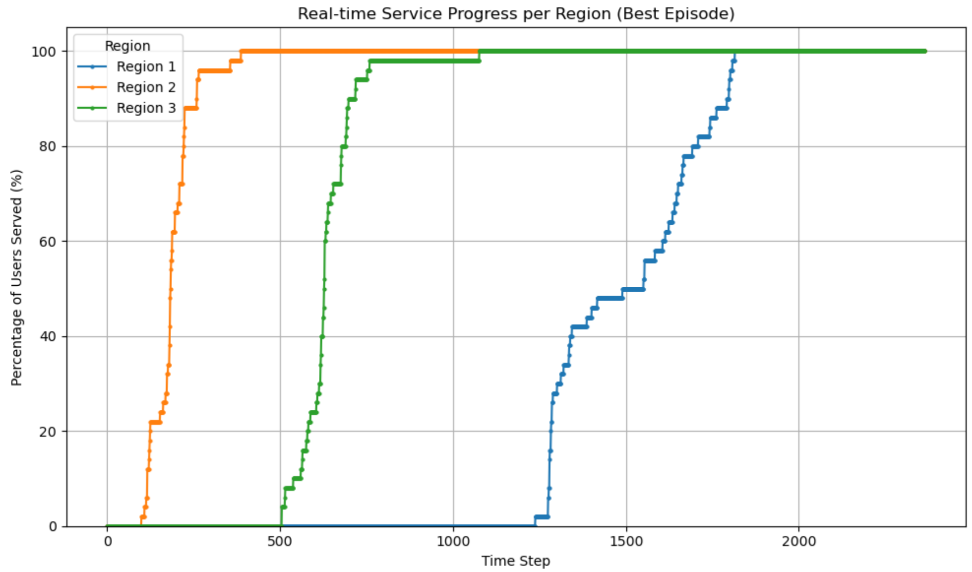


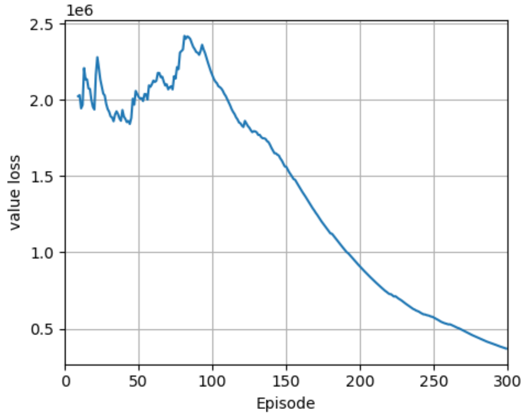
@20k run: UAV provide stable service in a region, maximizing MOS for 75 users.

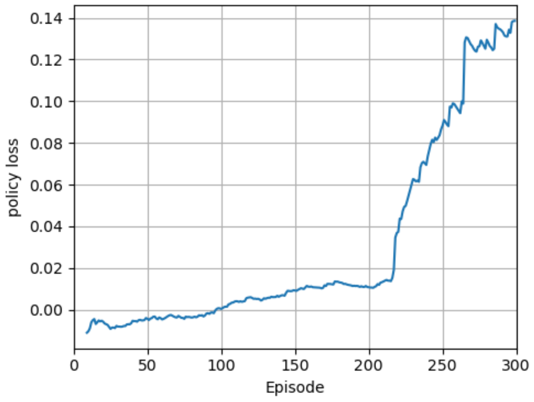
Research Progress: Phase I - Multi-Region Return-to-Home: Initial Progress & Challenges

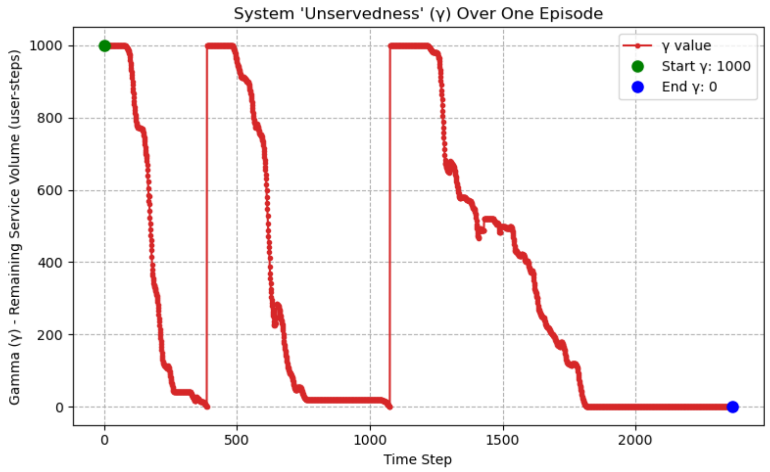












Acknowledge Challenges: Inconsistency (episode return) and instability (policy loss spike)

Future Works

1. Continue to stabilize the multi-region return-to-home mission, focusing on consistent learning and robust policy convergence (e.g., resolving policy loss instability).
2. Extend the operational constraint from fixed time steps to dynamic data amount/energy budget, reflecting real-world UAV limitations.
3. Conduct a comprehensive performance comparison of our Machine Learning (RL) approach against relevant baseline methods (e.g., Game Theory) for MOS maximization.
4. Further develop Phase II research, including multi-level emergency zone handling, priority levels, and obstacle integration, as outlined in the implementation plan.

References

[1] L. T. Hoang, C. T. Nguyen, H. D. Le and A. T. Pham, "Multi-Agent Reinforcement Learning for Cooperative Trajectory Design of UAV-BS Fleets in Terrestrial/Non-Terrestrial Integrated Networks," in IEICE Communications Express, vol. 13, no. 8, pp. 327-330, August 2024, doi: 10.23919/comex.2024XBL0084.

[2] H. Dai, Y. Ju, Y. Liang, Z. Zhang, H. Xu, and B. Wang, "QoE-driven multi-UAV deployment scheme for emergency communication networks," Electronics Letters, vol. 60, no. 1, pp. 1–10, Jan. 2024, doi: 10.1049/ell2.13188.

THANK YOU!