Search & Rescue Swarm Drone
Perception Architecture
(Synchronized Camera + LiDAR Training Strategy)

This document defines the perception architecture for the 150 g class Search & Rescue swarm drone system. It is aligned with:

- The Electronics Stack (STM32H7 / STM32N6 / STM32WL architecture)
- The Mesh Communications & Mission Concept specification

This document is structured to be clear to both human engineers and large language models.

1. Architectural Alignment with Electronics Stack

Processor Roles

STM32H7:
- Flight control
- Stabilization and motor control
- Navigation and failsafe logic
- Interface to 900 MHz control radio

STM32N6:
- Camera processing
- 54×42 ToF LiDAR processing
- AI inference (Neural-ART NPU)
- Object detection
- Semantic mapping generation

STM32WL:
- 900 MHz sub-GHz mesh communications
- Store-and-forward relay behavior

Data Flow

OV2640 Camera + 54×42 ST ToF → STM32N6 → detections / map deltas → STM32H7 → STM32WL mesh

STM32H7 ↔ 900 MHz long-range radio (pilot control + failsafe)

2. Real-Time RGB Perception (Primary Mode)

Sensor: OV2640 (~1 MP)

Model:
- YOLOv8n
- INT8 quantized
- TFLite format
- Input: 256×256 RGB

Pipeline on STM32N6:
1. Capture 1 MP frame
2. Downscale or crop to 256×256
3. Quantize to INT8
4. NPU inference
5. Confidence threshold ≥ 0.5
6. Select highest-confidence detection
7. Output: label confidence%

Target inference: 20–30 FPS.
Single top detection for bandwidth efficiency.

3. Low-Light Fallback: LiDAR-Only Perception

Sensor: ST 3D ToF 54×42 depth grid (~2,268 pixels)
Depth[y][x] → distance (mm)

If RGB confidence falls or lighting insufficient, switch to LiDAR mode.

Depth Model:
- Tiny CNN
- INT8 TFLite
- Input: 54×42×1

Initial classes:
- person_like_blob
- doorway_opening
- wall_plane
- obstacle_blob
- empty_open

4. Teacher–Student Training Strategy

No pretrained models exist for 54×42 LiDAR grids.

Training approach:

1. Collect synchronized RGB + LiDAR frames
2. Run YOLOv8n on RGB
3. Filter labels (confidence ≥ 0.8)
4. Align with LiDAR grid
5. Train tiny CNN on depth only

Recommended dataset: ~10,000 synchronized frames.

5. Mesh Integration

Perception outputs are structured and compact.

Detection packet includes:
• Object type
• Confidence
• Position estimate

Data sizes align with mesh constraints:
• Object detection: ~100–500 bytes
• Map delta: 100–800 bytes
• LiDAR keyframe (compressed): 1–5 KB

No video streaming over 900 MHz mesh.

6. Operational Impact

Enables:
• Indoor mapping
• Human detection
• Corridor recognition
• Hazard detection
• Operation in darkness or smoke

Supports distributed store-and-forward mesh behavior.