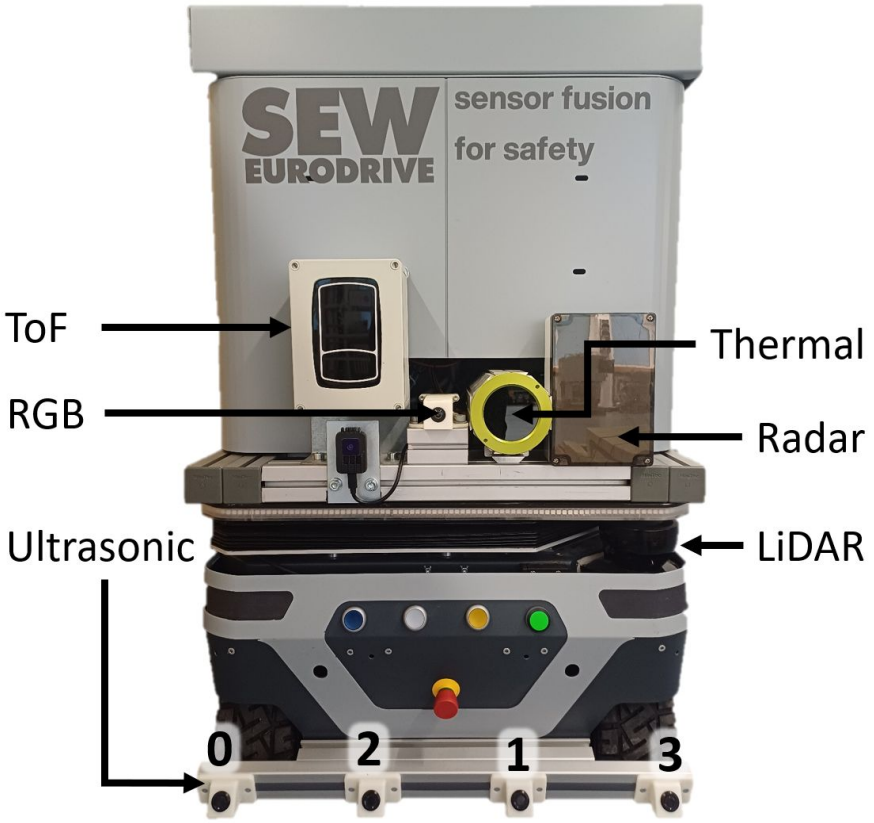
TEAM 4 / PRELIMINARY RESEARCH PROPOSAL

Sensor Interplay, Degradation and Fallback

Thermal, ToF depth and laser
for person detection on an industrial mobile robot

Supervisor: Myroslav Mishchuk
Team:
Oluwasesan Cornelius Akinbode (SUT)
Andrii Baran (LPNU)
Daniel Dąbrowski (SUT)



SEW-EURODRIVE dataset documentation
Image: CC BY 4.0

The research problem

Which sensor data helps us detect the person?

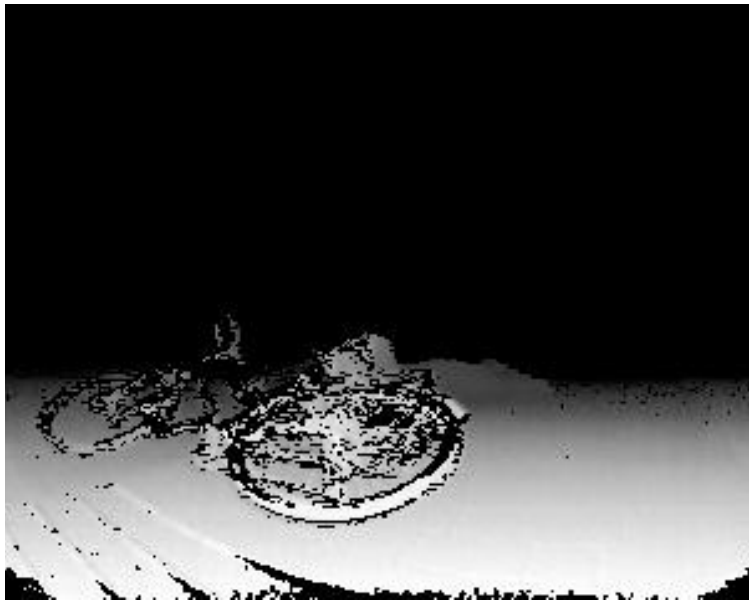
RGB (context)



Thermal



ToF depth



When a stream degrades, does it still earn its processing cost?

We will compare sensor utility, XAI interactions and a confidence-based fallback.

Official SEW nighttime example. RGB for context only. Images: CC BY 4.0.

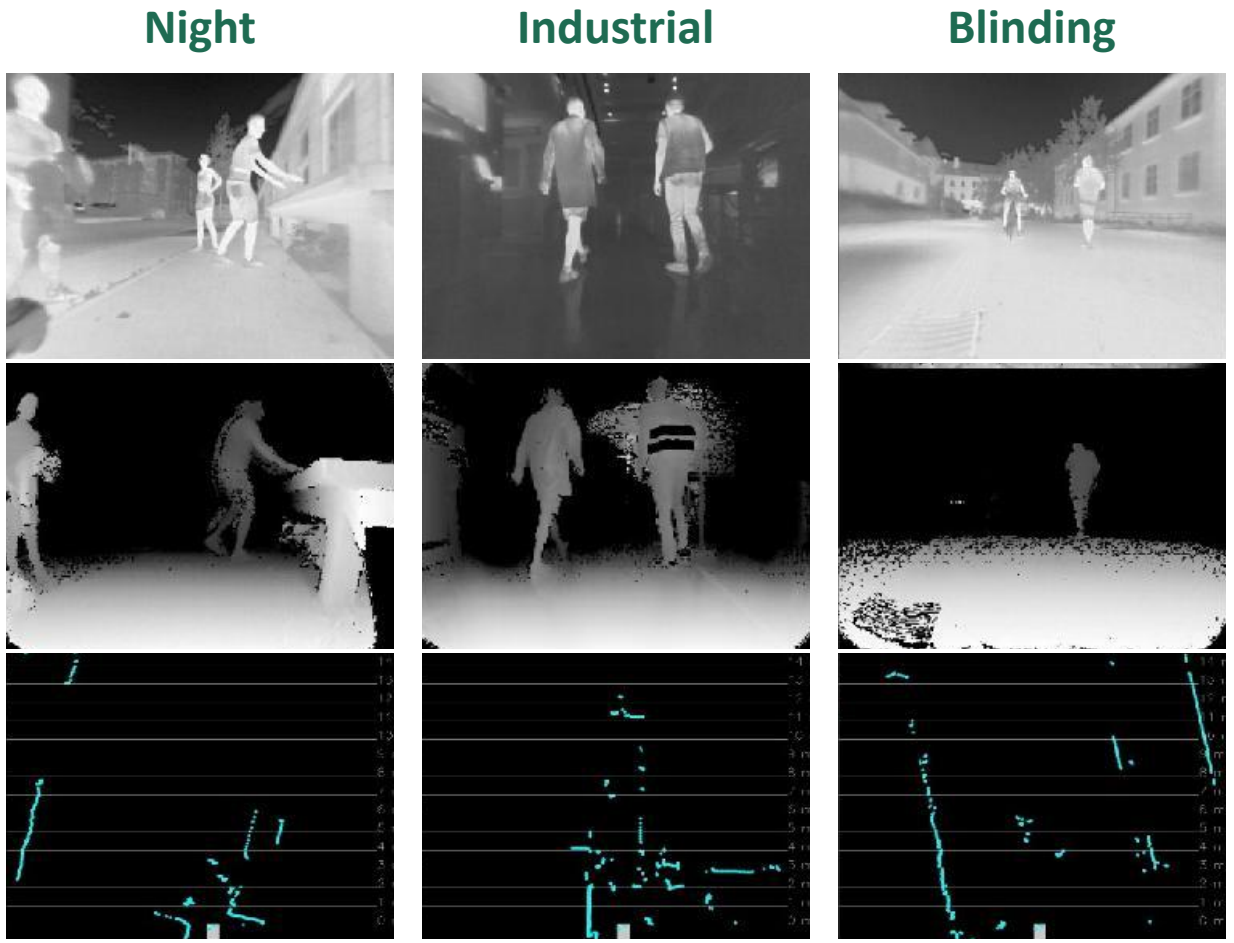
Main hypothesis and research questions

XAI interaction scores reveal whether sensor streams are complementary, redundant or conflicting under clean and degraded conditions. This knowledge can make a lightweight multimodal model more efficient or more robust without a significant loss in prediction quality.

RQ	Research question
RQ1	Do interaction scores agree with ablation, the ground truth for which streams are needed?
RQ2	As a stream degrades, do its interaction scores still reflect its benefit?
RQ3	Can a cheap run-time signal tell when to stop relying on a degraded stream?

RQ1 and RQ2 test the interaction claim. RQ3 tests the practical fallback.

SEW dataset and person-presence task



15,921
synchronized key frames

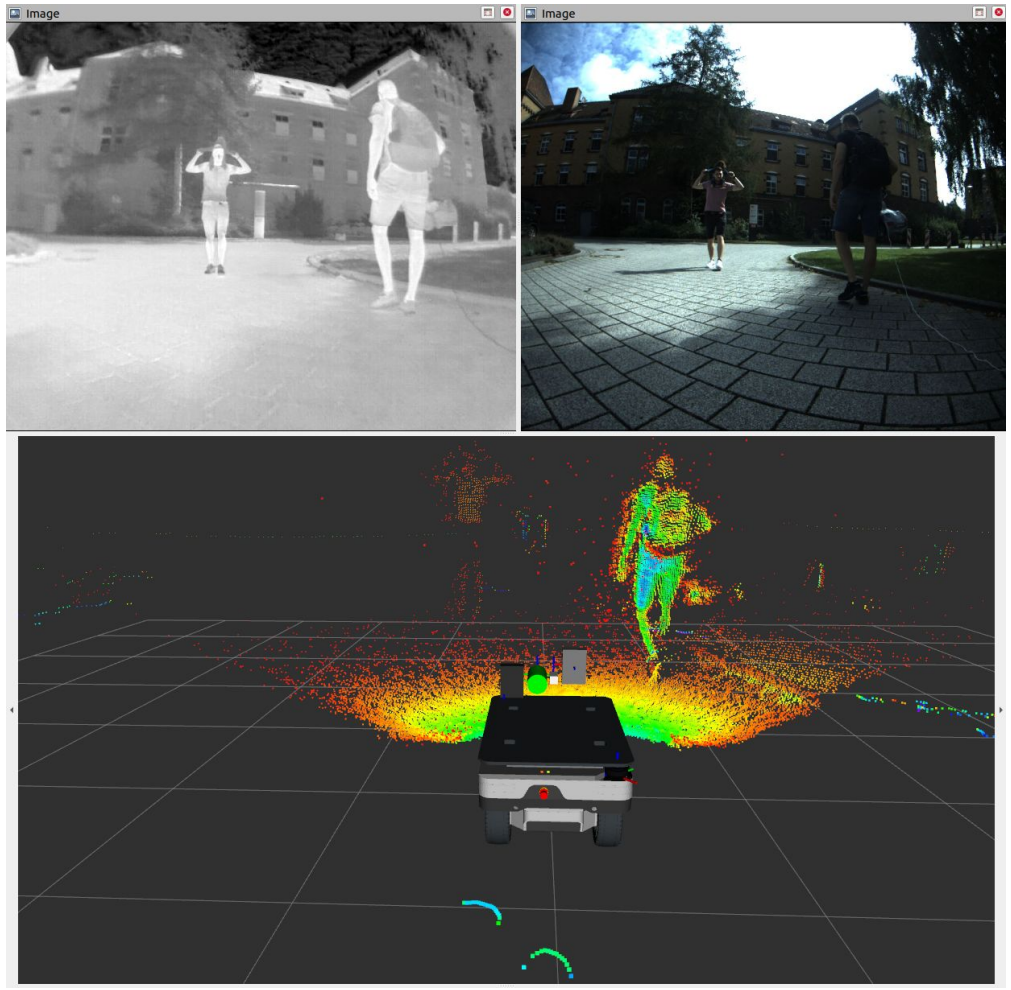
Person present in a fixed frontal region?

Industrial indoor and urban outdoor recordings

3D and 2D object boxes
Season, weather and daytime

Official SEW examples (cropped). RGB checks labels only. Data: CC BY-SA 4.0. Image: CC BY 4.0.

General approach



Training and split

Train on clean data.
Split by recording.

Sensor comparisons

Lightweight branches, all sensor subsets.
Intermediate and late fusion.

XAI and evaluation

Stream interactions against ablation.
Quality, fusion gain, XAI fidelity and stability.
Latency and memory.

SEW documentation example. RGB supports label checking only. Image: CC BY 4.0.

Degradation options

X = degraded stream R = remaining streams

Option	What we change
A. Measurement dropout	Remove valid measurements of X, such as depth pixels.
B. Feature-level noise	Add Gaussian noise to extracted features of X.
C. Signal-level noise	Add Gaussian noise to the raw signal of X, before feature extraction.

Train on clean data. Apply degradation at evaluation only.

Choose one degradation after a pilot. The three options test different failures.

RQ1

Do interaction scores agree with ablation?

H1. Stronger positive interactions go with larger non-additive gains in ablation. Weak or negative interactions go with redundancy or conflict.

Step	Comparison on clean data
1	Ablate every stream subset to establish the empirical reference.
2	Compute interactions on the fused model and compare them with ablation.
3	Check explanation stability and compare interaction methods.

If supported

Screen streams to keep or prune before costly ablation.

If rejected

Ablation remains necessary to judge sensor utility.

RQ2

Does a stream still pay off as it degrades?

H2. Changes in the degraded stream's interaction scores track the direction of changes in its contribution to prediction quality.

Worked case: depth added to laser–thermal fusion

Step	Comparison across severity levels
1	Measure by ablation how much X adds to R at each level.
2	Compare the trend with X-related interactions. Check replacement sensitivity.
3	Find changes to redundant or conflicting behaviour and test a simpler branch.

Consistently diverging trends reject H2. Negligible changes are inconclusive.

Agreement could guide simplification. Otherwise, interaction strength alone is insufficient.

RQ3

Can the model fall back to the remaining streams?

H3. A gate driven by calibrated predictive entropy loses less prediction quality than fixed-weight fusion as degradation grows.

Use R alone

When X's prediction is clearly less certain than R's.

Use fixed-weight fusion

Otherwise.

Calibrate and tune on validation only. The gate never sees the true severity.

Compare single streams, R alone, fixed fusion and the gate across severities.
Aim for a benefit under severe degradation with little cost on clean data.

Where R matches fusion, remove X's branch and measure latency and memory.

How the research questions connect

Stage	What it passes to the next stage
Shared setup	One dataset and split, stream-subset comparisons, fusion designs and one degradation.
RQ1. Interplay	Clean-data interaction regimes and a shortlist for pruning.
RQ2. Degradation	Severity ranges where the degraded stream stops paying off.
RQ3. Fallback	Settings where the remaining streams suffice.
Simplify and measure	Shrink or remove the branch, then recheck quality, latency and memory.

RQ1’s pruning shortlist also feeds the final simplification step.

Expected impact

Industry decision	RQ	Contribution
Sensor configuration	RQ1	Which streams add unique information, duplicate it or conflict.
On-board compute	RQ2	Where a degraded stream stops paying off, and a simplified branch.
Robustness at run time	RQ3	A tested fallback rule for an unreliable stream.

Possible actions

Keep or prune a stream. Skip, downsample or shrink its processing.
Down-weight an unreliable input at run time.

Risks and the first step

Risk	Mitigation
A degraded branch stays confidently wrong.	Check confidence against severity before relying on the gate.
Interaction scores depend on missing-stream replacement.	Compare at least two replacement strategies.
Degradation options test different failures.	Commit to one option after a pilot.

First step: a small SEW pilot

Check samples and labels, then agree on models and one degradation.

A preliminary four-stage plan

Stage	Main steps
1. Pilot	Check SEW samples, alignment and labels. Define the frontal region. Select models and one degradation.
2. Clean baselines	Use recording-based splits. Compare stream subsets and at least two fusion designs.
3. Degradation and fallback	Compare XAI with ablation across severities. Check stability and replacements. Tune the entropy gate on validation.
4. Simplification	Test a smaller or removed branch where justified. Recheck quality, latency and memory.

Pilot decisions

Ablation definition, XAI comparison rule and acceptable loss in prediction quality

Questions for discussion

Problem

Does the problem exist, and is it sufficiently challenging?

Hypotheses

Are the hypotheses clear and directly related to the problem?

Scientific insight

Would confirming or rejecting them provide a meaningful insight?

Research plan

Can the preliminary plan reasonably test these hypotheses?

Open choices: models, degradation, ablation definition and quality-loss tolerance

References

1. SEW-EURODRIVE

SEW Multimodal AMR Dataset 2026 documentation

2. Wenderoth et al. (2025)

Measuring Cross-Modal Interactions in Multimodal Models. AAAI.

3. Muschalik et al. (2024)

shapiq: Shapley Interactions for Machine Learning. NeurIPS.

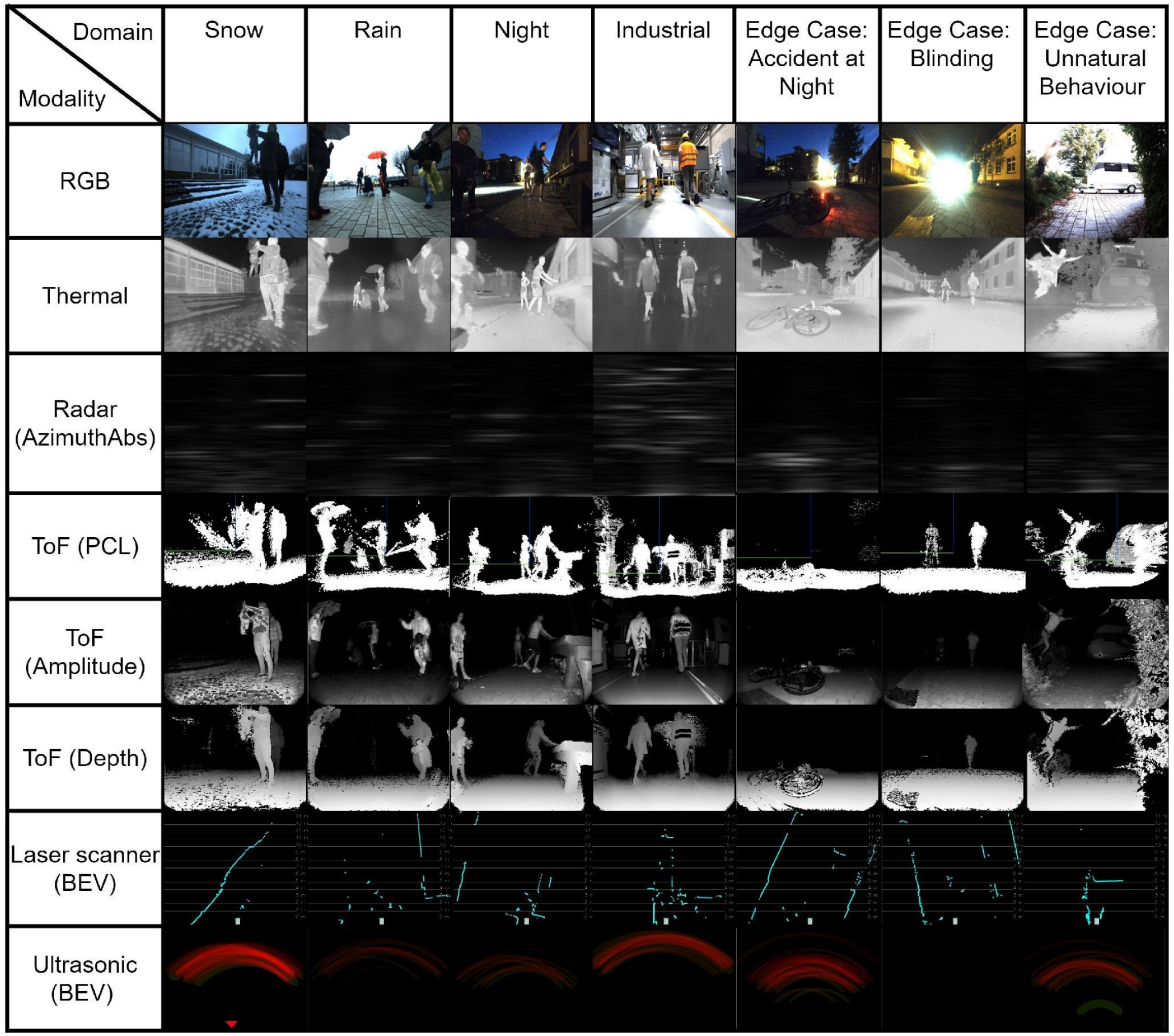
4. Zhang et al. (2023)

Provable Dynamic Fusion for Low-Quality Multimodal Data. ICML.

5. Bijelic et al. (2020)

Seeing Through Fog Without Seeing Fog. CVPR.

SEW modalities and recording conditions



Our inputs

Thermal image
Independent ToF depth
2D laser scans

RGB supports label checking.

Other modalities in this figure provide dataset context.

Source: SEW documentation, CC BY 4.0