ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

Anomaly Detection, Localization and Classification for Railway Inspection

Riccardo Gasparini, Stefano Pini, Guido Borghi, Giuseppe Scaglione,
Simone Calderara, Eugenio Fedeli, Rita Cucchiara

{name.surname}@unimore.it {g.scaglione, e.fedeli}@rfi.it

University of Modena and Reggio Emilia, Italy

Rete Ferroviaria Italiana, Gruppo Ferrovie dello Stato

UNIMORE UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

AImage Lab

RFI RETE FERROVIARIA ITALIANA GRUPPO FERROVIE DELLO STATO ITALIANE

Visual Inspection of Railways: Why?

- The **inspection of railways**, *i.e.* the activity to check the **absence of obstacles** placed on the railroad that could **damage or derail trains**, is a key element to **guarantee the safety of transports**
- Inspection activities are usually conducted **during the night**, when the train circulation is suspended

1

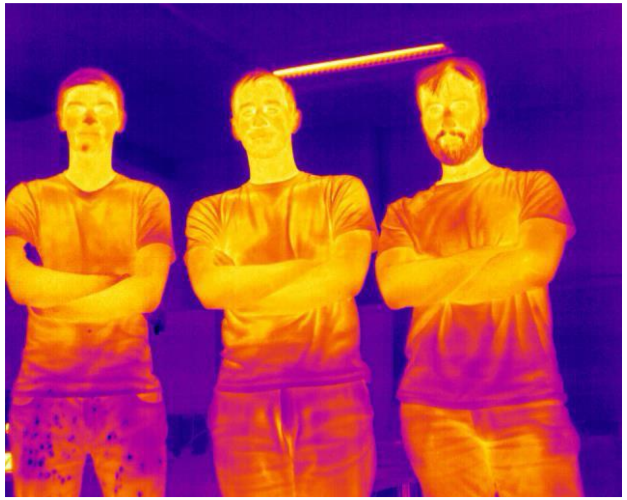
- Due to the vastness of railways, an automatic inspection is strongly demanded
- Therefore, we propose a **vision-based framework to detect obstacles** in videos acquired from a **rail drone**, *i.e.* a self-powered vehicle moving along railways and operated by remote control.

1. https://bari.repubblica.it/cronaca/2018/04/24/news/treni_bari-matera_frigo_sui_binari-194731081 *(news in italian)*

Requirements

- We identify 3 main **requirements**:
  - **Fast Acquisition**: the **frame rate** and the **shutter speed** of the acquisition devices must be sufficiently high **to avoid motion blur** caused by the high speed of the drone (up to 100 km/h)
  - **Night Vision**: the adoption of external **light sources** and the use of **thermal** cameras is required. Due to energy consumption, we focus our attention on thermal cameras
  - **High Resolution**: to detect even small-sized anomalies at long distances, cameras must have a high spatial resolution.

Motion blur



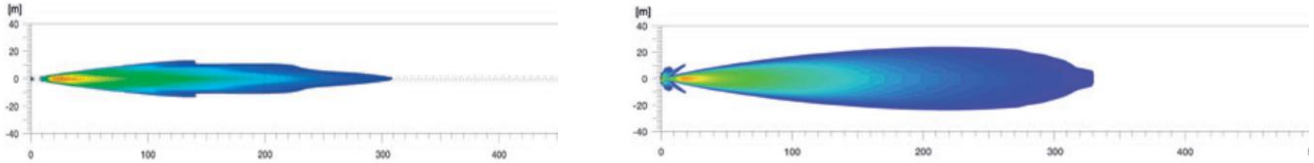
Thermal frame

Hardware Solution

The following **cameras** and **light sources** are employed:

- **Basler ac800-510ac**[1]
  - RGB camera
  - Up to **500 fps**
  - Spatial resolution: 800x500 pixels
  - Equipped with a **12.5 – 75mm zoom lens**

- **2 Light sources**
  - **2 x LED Light Bar 470**[2]
    - Wide areas close to the drone
  - **2 x Comet 200 LED**[3]
    - Areas far from the drone



1. https://www.baslerweb.com/en/products/cameras/area-scan-cameras/ace/aca800-510uc
2. https://www.hella.com/truck/it/LED-LIGHT-BAR-470-Single-Twin-3950.html
3. https://www.hella.com/offroad/it/Comet-200-LED-1626.html

Hardware Solution

- **Flir Boson 640**[1]
  - Thermal camera
  - 640x480 pixels
  - Up to 60 fps
  - Equipped with a 14mm lens

In this work, we focus on this camera (thermal data)

- **Zed Stereo camera**[2]
  - RGB stereo camera
  - 4416x1242
  - Up to 20 meters
  - From 15 to 100 fps

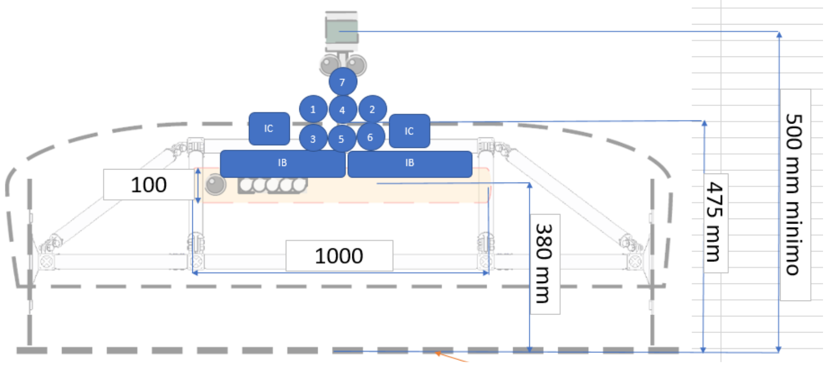
RGB Disparity Map

1. https://www.flir.it/products/boson
2. https://www.stereolabs.com/zed

Hardware solution analysis

- Analysis of **weights** and **power** consumptions:

Item	Weight	Power
Basler ac800-510ac	200g (body) + 483g (lens)	3,2W
2 x LED Light Bar 470	2 x 900g	2 x 35W
2 x Comet 200 LED	2 x 495g	2 x 13W
Flir Boson 640	7,5g	500 mW
Zed Stereo camera	159g	1,5W



Spatial locations of cameras and external light sources

- **Total**: **3,6 kg** and **100 W**, suitable for a self-powered drone

The *Vesuvio* Dataset

- We collect a **new dataset** using a variety of cameras and light sources
- We placed them on a **real train**
- We select and employ the following objects, which are the **common tools** used in the construction sites along the railways:
  - Electrical Insulator
  - Fuel Tank
  - Rail Signal
  - Pickaxe
  - Locking turnout
  - Track lifting jack
  - Traffic light
  - Insulating stick
  - LPG tank
  - Balise
  - Oiler

Overview of obstacles in **thermal domain.**

(a) Electrical Insulator

(b) Fuel tank

(c) Rail signal

(d) Pickaxe

(e) Locking turnout

(f) Track lifting jack

(g) Traffic light

(h) Insulating stick

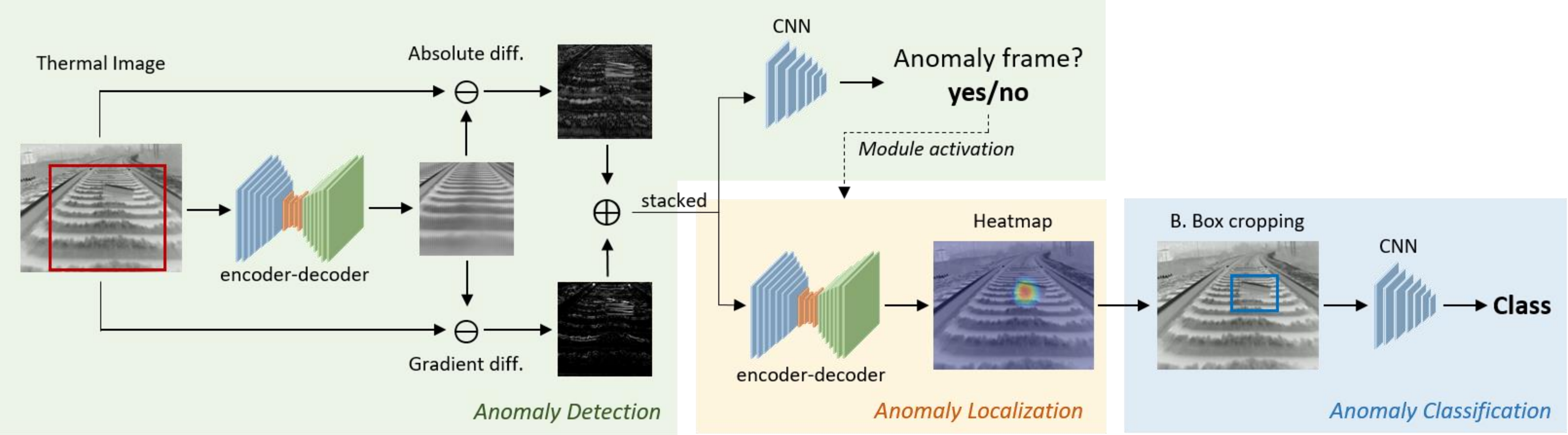
(i) LPG tank

(j) Balise

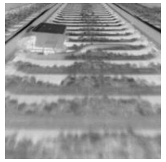
(k) Oiler

(l) No anomaly

- The proposed framework is a combination of **3 sequential blocks**:

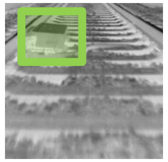


Anomaly Detection



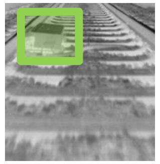
Frame contains an anomaly

Anomaly Localization



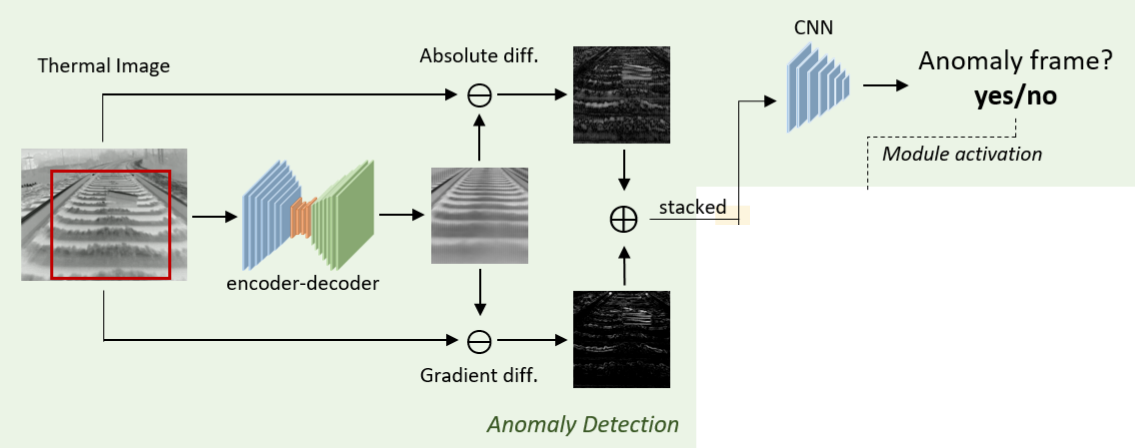
Anomaly is inside the green bounding box

Anomaly Classification



Anomaly inside the green bounding box is an oiler

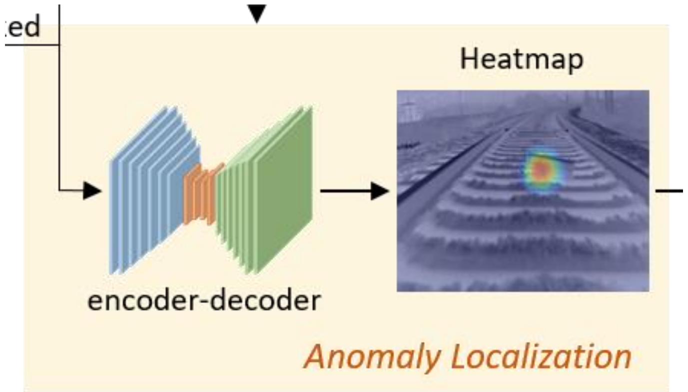
Anomaly Detection

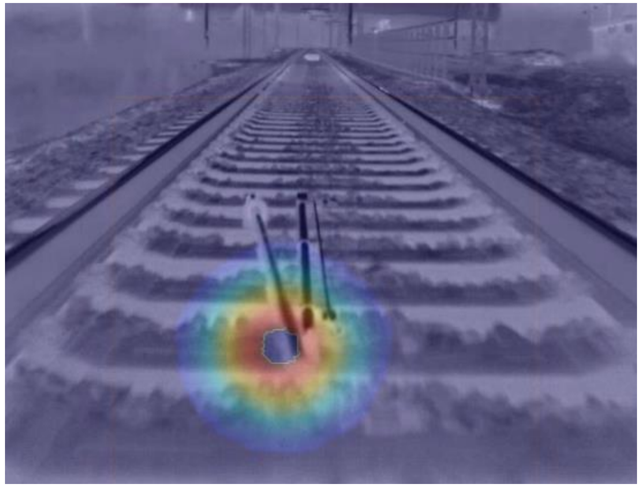


- **Input**: a single thermal frame (192x192)
- **Output**: the reconstructed frame
- **Training**:
  - The encoder-decoder receives as input **only regular frames**, *i.e.* frames without any anomaly
  - The reconstructed frame is compared with the original input frame, through:
    - An **absolute** difference: $L_{MSE} = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} \| I_I(m,n) - I_R(m,n) \|_2^2$
    - A **gradient** difference, *i.e.* a difference computed on the gradients of the two images: $L_G = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} \| G_{I_I}(m,n) - G_{I_R}(m,n) \|_2^2$
  - The resulting two difference images are **stacked** and classified as anomaly/not anomaly by the second architecture, the classifier.

Anomaly Localization

- **This module is activated only when the first one classifies the frame as anomaly**
- **Input**: stacked difference images (2x192x192)
- **Output**: heatmap (the location of the anomaly is expressed with a bi-variate Gaussian function)
  - Heatmap is then used to **extract a crop** of the detected anomaly which will be classified by the last module of the framework
  - The autoencoder is lighter, in terms of number of parameters, than the one used in the first module.
- **Training**:
  - **Supervised training**: anomalies which have been manually annotated with bounding boxes
  - *Adam* optimizer and the *Mean Squared Error* function are used

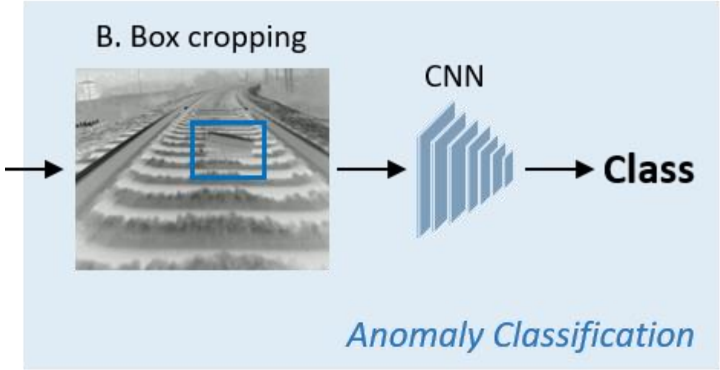


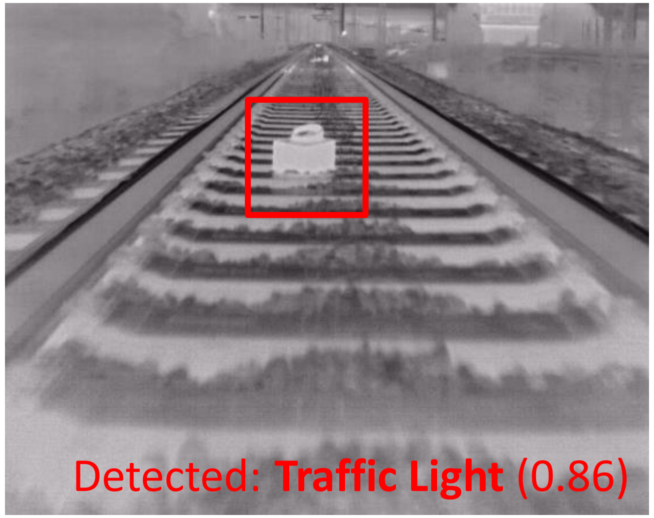


Heatmap of the anomaly

Anomaly Classification

- **Input**: the anomaly is cropped (64x64)
  - smaller images are zero-padded
  - bigger images are appropriately resized
- **Output**: anomaly class
- **Training**:
  - Supervised approach (classes are manually labelled)
  - The *Categorical Cross Entropy* loss is employed as objective function
  - The network outputs a probability score on the list of available classes





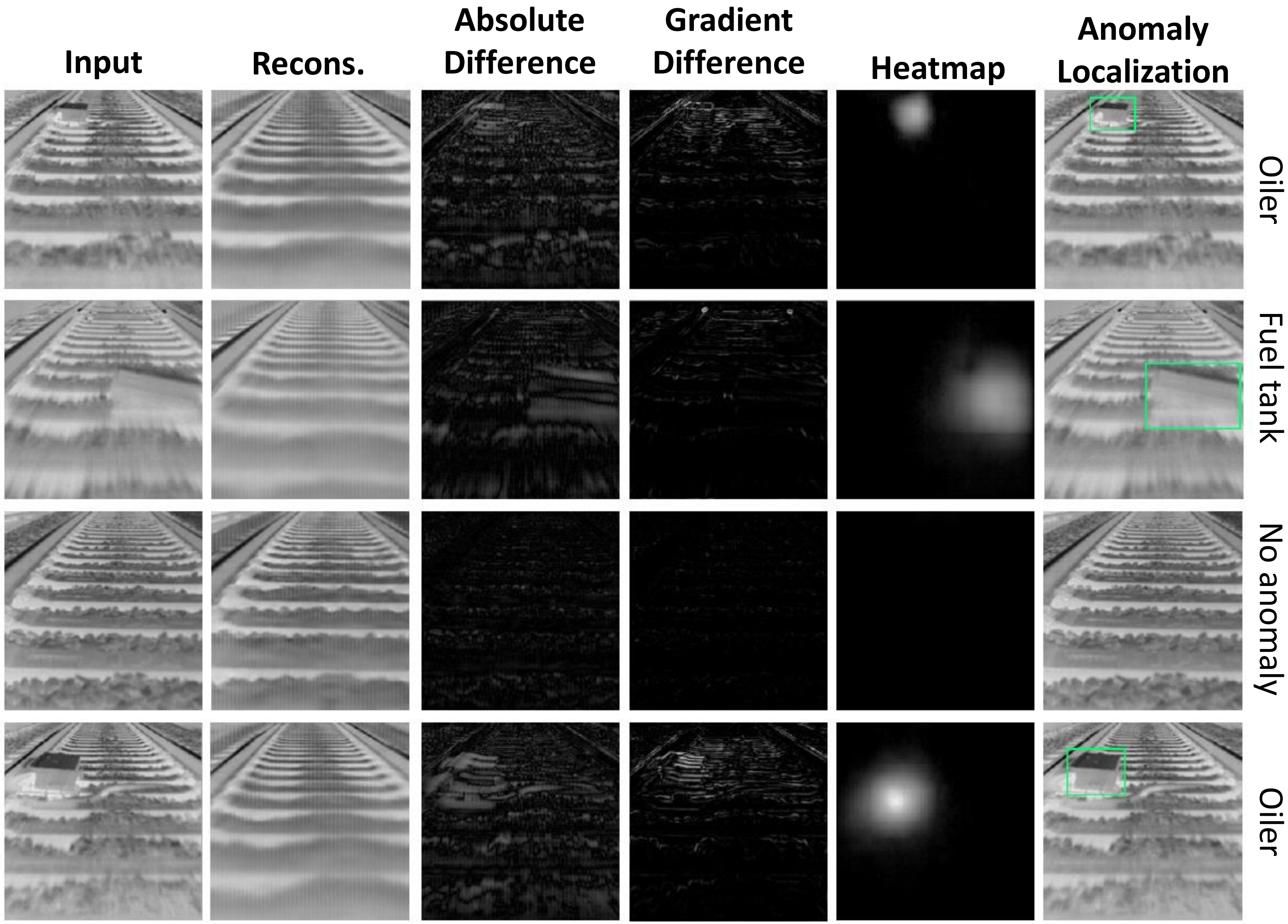
Results

- Results obtained on 2 different split configuration:
  - **80-20**: we put 80% of data in train and 20% in test
  - **Cross-class**: 3 classes are in the test set, the remaining in the training one
- For the anomaly localization task, we consider an object as correctly located only if IoU>0.3

	80-20 split				Cross-class split			
Modules	**Accuracy**	**Precision**	**Recall**	**F1-score**	**Accuracy**	**Precision**	**Recall**	**F1-score**
Anomaly Detection	0.966	0.989	0.957	0.973	0.848	0.903	0.776	0.835
Anomaly Localization	0.903	0.741	0.989	0.847	0.785	0.521	0.997	0.684
Anomaly Classification	0.970	0.795	0.794	0.785	-	-	-	-

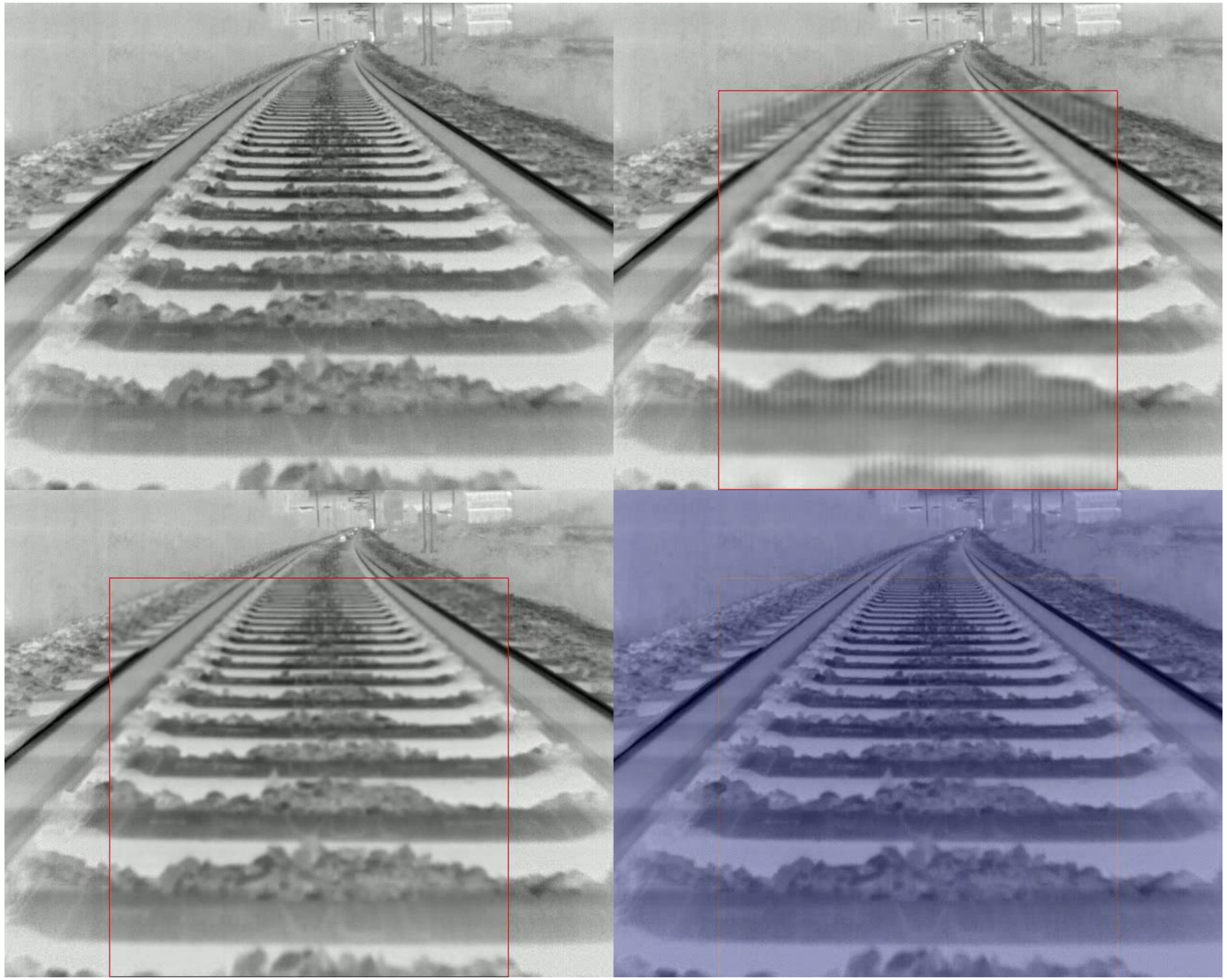
Class	Accuracy	Precision	Recall	F1-score
Electrical insulator	0.94	0.54	0.44	0.48
Fuel tank	0.96	0.65	0.76	0.70
Rail signal	0.99	1.0	0.77	0.87
Pickaxe	0.97	0.58	0.92	0.71
Locking turnout	0.99	0.88	0.78	0.82
Track lifting jack	0.98	0.81	0.94	0.87
Traffic light	0.99	0.88	0.95	0.91
Insulating stick	0.98	0.97	0.90	0.94
LPG tank	0.97	0.92	0.79	0.85
Balise	0.93	0.73	0.53	0.62
Olier	0.97	0.79	0.96	0.87

Visual Results



Demo Video

Input (thermal) **Reconstructed**



Classification **Heatmap**