# *A Contrario* Paradigm for YOLO-based Infrared Small Target Detection


Alina CIOCARLAN[1,2], Sylvie LE HEGARAT-MASCLE[2], Sidonie LEFEBVRE[1], Arnaud WOISELLE[3], Clara BARBANSON[3]

[1]DOTA ,ONERA, Université Paris-Saclay, F-91123 Palaiseau, France [2]SATIE, Université Paris-Saclay, 91405 Orsay, France [3]Safran Electronics & Defense, F-91344 Massy, France

*alina.ciocarlan@onera.fr*


## Motivations

### *Context*

- **Surveillance:** detection of aircrafts on infrared (IR) images
- **Applications:** military intelligence, anti-collision systems, decision-making tools

### *Application-related locks*

- **Small** and **rare** targets with **low contrast** compared with the background
- **Complex** and **textured** backgrounds (clouds, vegetation, buildings…)
- **Very few annotated data** (thermal-IR, multispectral), **high** annotation **costs**

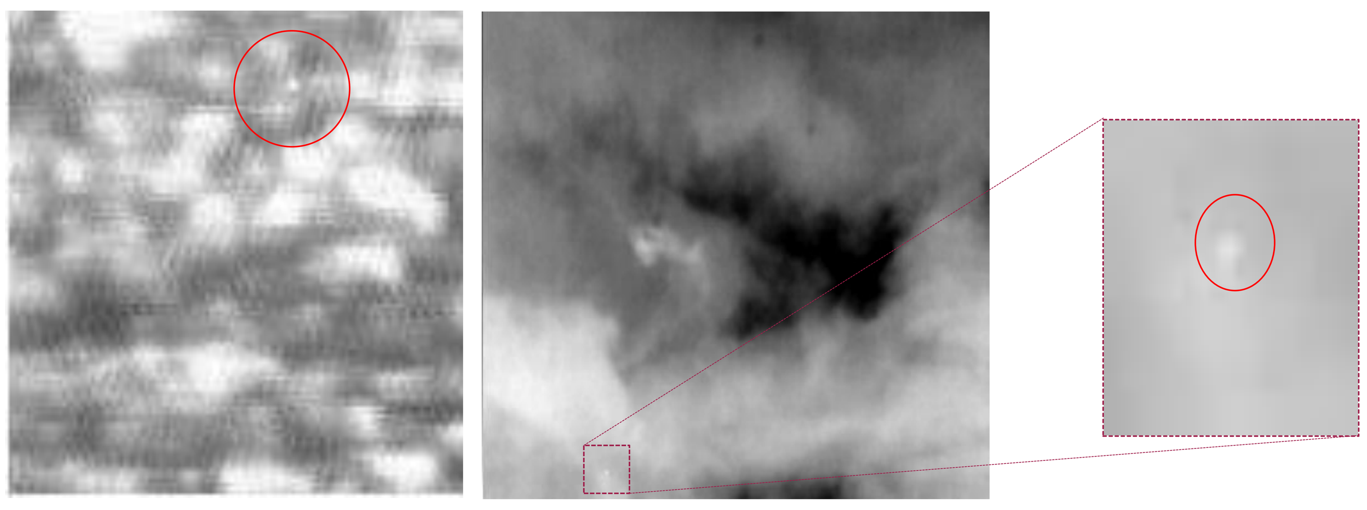


Fig. 1: Images taken from the publicly available SIRST dataset [1]. The target is drowned in cluttered and challenging backgrounds.

## Proposed method

### *A contrario paradigm*

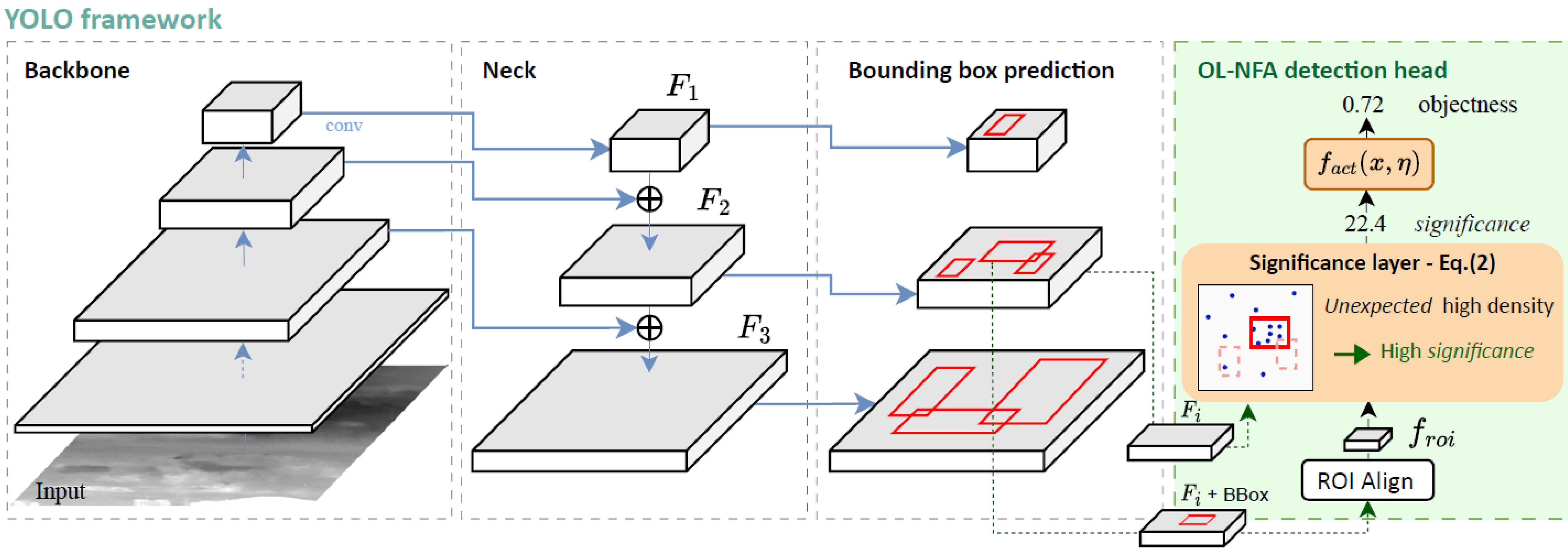
- Detecting what differs from the naive modelling of the background [2]
- Usual case - binary image:
  - Naive model = Bernoulli distribution
  - Detection of unexpectedly dense structures in terms of pixels with value 1

$$NFA(\kappa,\nu,p) = N_{test}\times P(|X_i| \geq \kappa) = N_{test}\sum_{i=\kappa}^{\nu}\binom{\nu}{i}p^i(1-p)^{\nu-i}$$

- $\kappa$ Nb. of pixels with value 1 in the tested area
- $\nu$ Area of the tested rectangle
- $p$ Density of points in the image
- $N_{test}$ Nb. of tested rectangles within an image

Fig. 2: Illustration of the Gestalt laws. An abnormal grouping of points allows us to perceive objects in a scene. [3]

### *Differentiable integration of the a contrario criterion into YOLO [4] framework*




#### Object-level NFA head (OL-NFA)

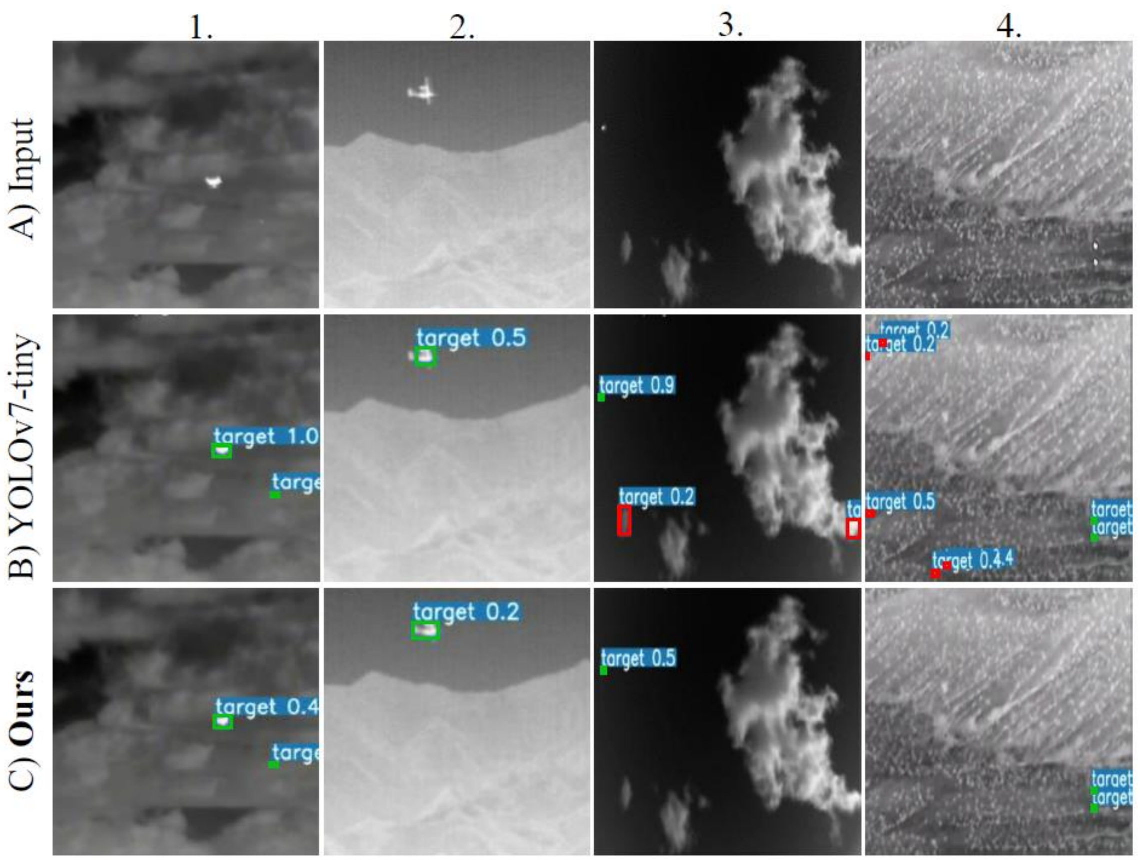
- Fuzzy belonging coefficients
- *Significance* rather than NFA

$$S(\kappa,\nu,p) = -\ln(\mathrm{NFA}(\kappa,\nu,p))$$

- Specific activation function

$$f_{act}(x,\eta) = 2\sigma(x+\ln\eta)-1$$

- **State-of-the-art results on SIRST**




| Method | F1 | AP | Prec. | Rec. | FPS |
|---|---|---|---|---|---|
| Segmentation networks for IRSTD | | | | | |
| ACM | $95.4^{\pm1.7}$ | $95.2^{\pm3.8}$ | 95.1 | 95.8 | 251 |
| LSPM | $85.0^{\pm2.9}$ | $90.2^{\pm0.8}$ | 86.6 | 83.5 | 125 |
| DNANet | $\underline{96.9}^{\pm0.5}$ | $\underline{98.1}^{\pm1.2}$ | 96.6 | **97.2** | 33 |
| Object detection methods | | | | | |
| YOLOv3 | $96.1^{\pm0.3}$ | $97.5^{\pm0.1}$ | 96.9 | 95.4 | 144 |
| YOLOR | $95.7^{\pm2.2}$ | $96.7^{\pm1.1}$ | 96.5 | 94.9 | 136 |
| YOLOv7 | $96.5^{\pm1.2}$ | $97.6^{\pm0.7}$ | 97.2 | 95.9 | 147 |
| YOLOv7-tiny | $96.5^{\pm0.6}$ | $97.8^{\pm0.4}$ | 96.9 | $\underline{96.2}$ | 256 |
| **Ours** | $\mathbf{97.2}^{\pm0.6}$ | $\mathbf{98.2}^{\pm0.2}$ | **98.6** | 95.9 | 208 |

#### *Results*

- ➢ Our OL-NFA head **bridges the performance gap** between SOTA segmentation and detection methods
- ➢ It controls the **number of false alarms** while mainting a high detection rate

- **Robustness towards few-shot training**

| Method | 15-shots | | 25-shots | |
|---|---|---|---|---|
| | F1 | AP | F1 | AP |
| YOLOv7-tiny | $50.7^{\pm7.0}$ | $51.3^{\pm7.0}$ | $68.0^{\pm6.6}$ | $69.6^{\pm8.4}$ |
| **Ours** | $\mathbf{85.0}^{\pm5.0}$ | $\mathbf{90.5}^{\pm5.2}$ | $\mathbf{89.7}^{\pm4.2}$ | $\mathbf{93.4}^{\pm2.0}$ |

#### *Results for few-shot setting*

- ➢ Our OL-NFA detection head significantly improves the performance in a few-shot training
- ➢ This is because our network learns a representation of **background** elements rather than the targets themselves

## Perspectives

- Adapt the *a contrario* block for other detection backbones [5], test other naïve models
- Limit the drop in performance observed for large object detection

## References


[1] Yimian Dai, et al. « Asymmetric contextual modulation for infrared small target detection ». In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 950–959, 2021.

[2] A. Desolneux, et al. “From gestalt theory to image analysis : a probabilistic approach”, vol. 34. *Springer Science & Business Media*, 2007.

[3] https://www.usabilis.com/definition-theorie de gestalt/

[4] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi, “You only look once: Unified, real-time object detection,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[5] Ciocarlan, Alina, et al. "Deep-NFA: A deep A Contrario framework for tiny object detection." Pattern Recognition (2024): 110312.


ONERA THE FRENCH AEROSPACE LAB SAFRAN université PARIS-SACLAY SATIE