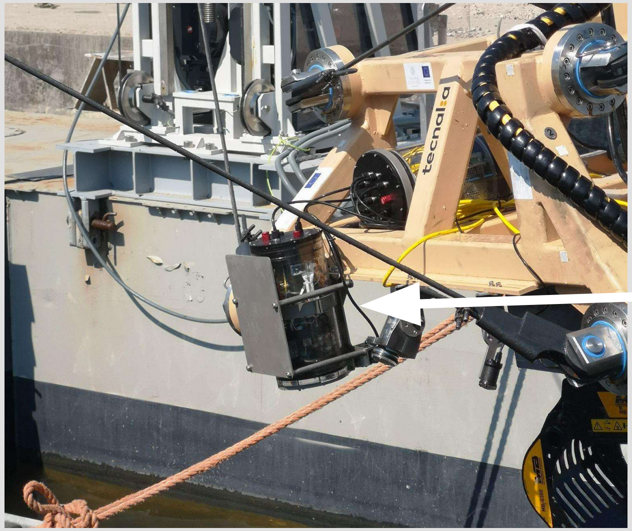
From TrashCan to UNO: Deriving an Underwater Image Dataset To Get a More Consistent and Balanced Version

Cyril Barrelet[1], Marc Chaumont[1, 3], Gérard Subsol[1], Vincent Creuze[1], Marc Gouttefarde[2]

[1] Research team ICAR, LIRMM, Univ Montpellier, CNRS, Montpellier, France
[2] Research team DEXTER, LIRMM, Univ Montpellier, CNRS, Montpellier, France
[3] Univ Nîmes, France

Context

- Remove macro-litter from the seabed



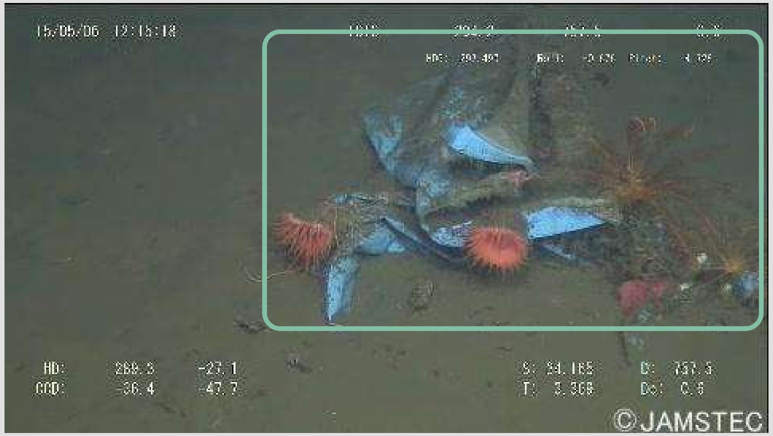
Available underwater macro-litter databases

DeepSeaWaste dataset

Aluminium can

- 544 images + labels
- 76 classes

TrashCan dataset



- 7,212 images + 8634 labels
- 16-22 classes (8 litter categories)

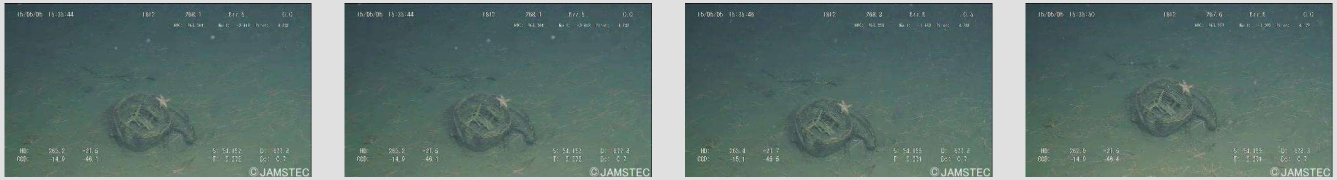
Deep learning issues

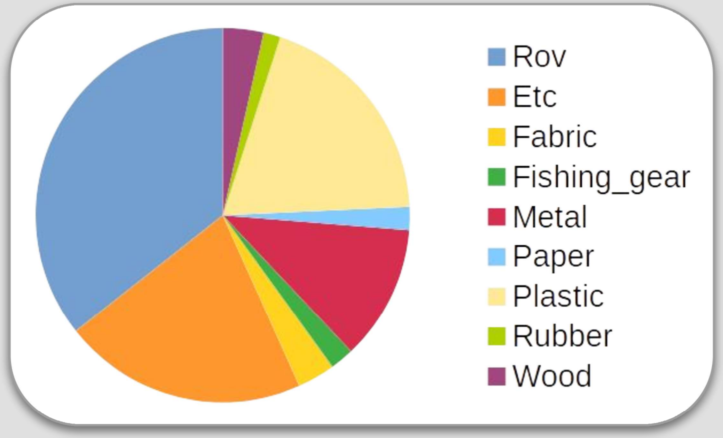
- TrashCan construction bias
  - 7,212 frames extracted from 312 sequences

Deep learning issues

- TrashCan construction bias
  - 7,212 frames extracted from 312 sequences

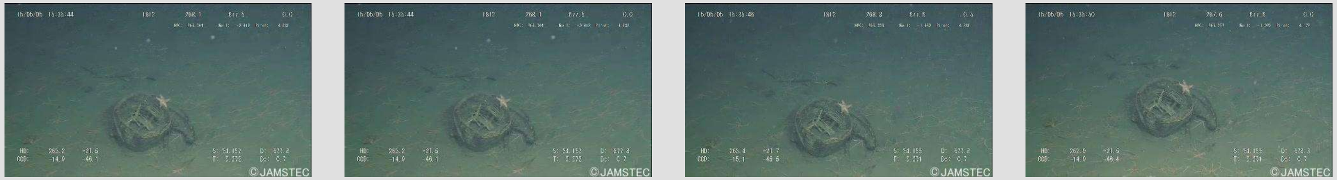
- Class unbalance

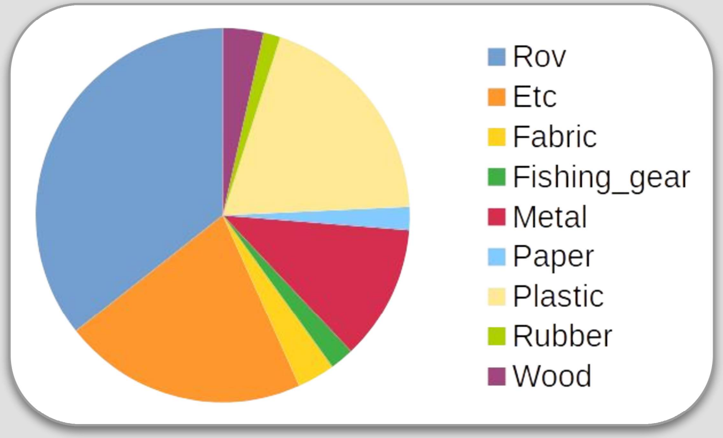




Deep learning issues

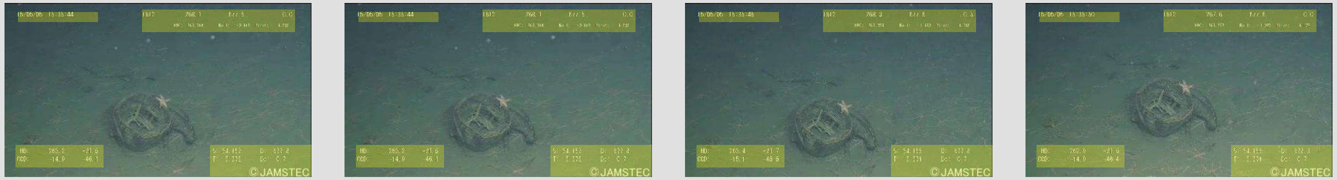
- TrashCan construction bias
  - 7,212 frames extracted from 312 sequences
- Class unbalance
- Annotations quality
  - Incorrect annotations
  - Missing annotations
  - Poor localization

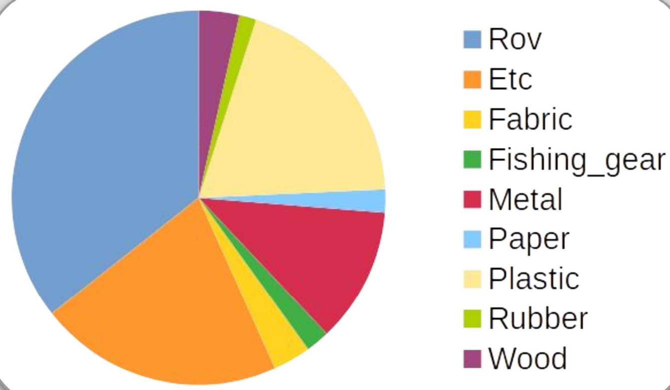


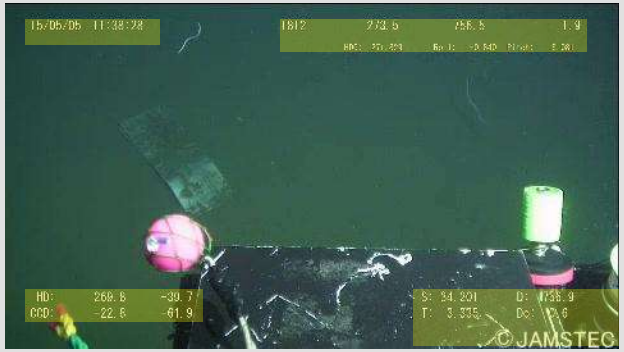


Deep learning issues

- TrashCan construction bias
  - 7,212 frames extracted from 312 sequences
- Class unbalance
- Annotations quality
  - Incorrect annotations
  - Missing annotations
  - Poor localization
- Metadata overlay







Contributions

- **New** Underwater Non-natural Object dataset: UNO
- Methodology to compare networks using a **well-balanced** k-fold
- **Comparison** of TrashCan and UNO using YOLOv5
- **Covariate shift** test using underwater images from AQUALOC

UNO construction

- Label redefinition
  - Non-natural objects (one class)

Original TrashCan

7,212 images

8,634 labels

Derived UNO

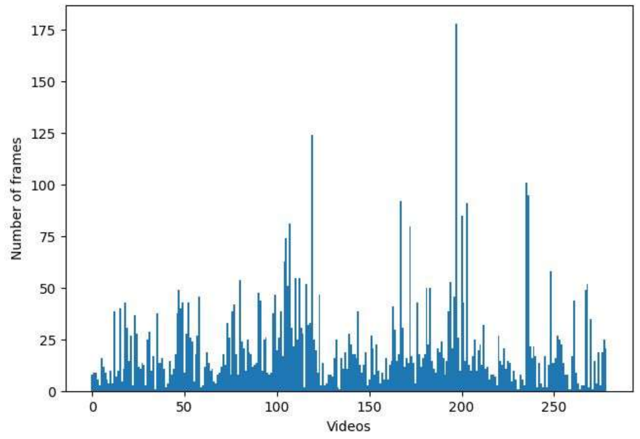
5,902 images

10,773 labels

A methodology to obtain a **well-balanced** k-fold

$$f^* = \underset{f \in \{1..5\}^{279}}{arg\ min} (\sigma_F + \sigma_{BB})$$

Bin packing problem



Fold	Videos	Frames	BBs
1	63	1180	2159
2	64	1182	2137
3	49	1185	2152
4	44	1179	2163
5	57	1176	2162
Mean	55.4	1189.2	2154.6
Std	7.81	3.00	9.60

Experiments and results

Model and hyperparameters

- YOLOv5m pre-trained on ImageNet
- SGD optimizer
- OneCycle scheduler
- Initial learning rate: 0.0032
- Final learning rate: 0.000384
- Warmup: 20%
- Batch size: 28
- 5 trainings of 300 epochs each

Augmentations

- Color transformation
- Rotations
- Translations
- Scaling
- Shearing
- Flip-UP and Flip-LR
- Mosaic
- Mixup

Training set	Evaluation set	Split	F1-score	mAP@.50
TrashCan	TrashCan	Random	79.7	80.8

Experiments and results

Model and hyperparameters

- YOLOv5m pre-trained on ImageNet
- SGD optimizer
- OneCycle scheduler
- Initial learning rate: 0.0032
- Final learning rate: 0.000384
- Warmup: 20%
- Batch size: 28
- 5 trainings of 300 epochs each

Augmentations

- Color transformation
- Rotations
- Translations
- Scaling
- Shearing
- Flip-UP and Flip-LR
- Mosaic
- Mixup

Training set	Evaluation set	Split	F1-score	mAP@.50
TrashCan	TrashCan	Random	79.7	80.8

Experiments and results

Model and hyperparameters

- YOLOv5m pre-trained on ImageNet
- SGD optimizer
- OneCycle scheduler
- Initial learning rate: 0.0032
- Final learning rate: 0.000384
- Warmup: 20%
- Batch size: 28
- 5 trainings of 300 epochs each

Augmentations

- Color transformation
- Rotations
- Translations
- Scaling
- Shearing
- Flip-UP and Flip-LR
- Mosaic
- Mixup

Training set	Evaluation set	Split	F1-score	mAP@.50
TrashCan	TrashCan	Random	79.7	80.8
TrashCan	TrashCan	K-folded	58.4 ± 4.2	56.6 ± 6.3

Experiments and results

Model and hyperparameters

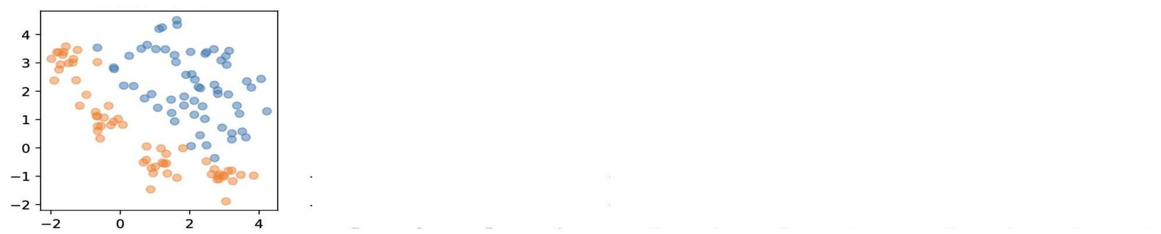
- YOLOv5m pre-trained on ImageNet
- SGD optimizer
- OneCycle scheduler
- Initial learning rate: 0.0032
- Final learning rate: 0.000384
- Warmup: 20%
- Batch size: 28
- 5 trainings of 300 epochs each

Augmentations

- Color transformation
- Rotations
- Translations
- Scaling
- Shearing
- Flip-UP and Flip-LR
- Mosaic
- Mixup

Training set	Evaluation set	Split	F1-score	mAP@.50
TrashCan	TrashCan	Random	79.7	80.8
TrashCan	TrashCan	K-folded	58.4 ± 4.2	56.6 ± 6.3
UNO	UNO	K-folded	67.3 ± 1.5	68.8 ± 1.2

Domain shift evaluation



Domain shift:

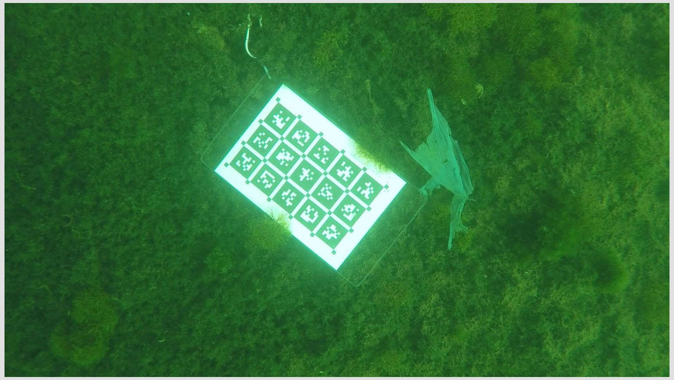
Change in the data distribution between an algorithm's training dataset, and a dataset it encounters when deployed.

Images color

Objects shape

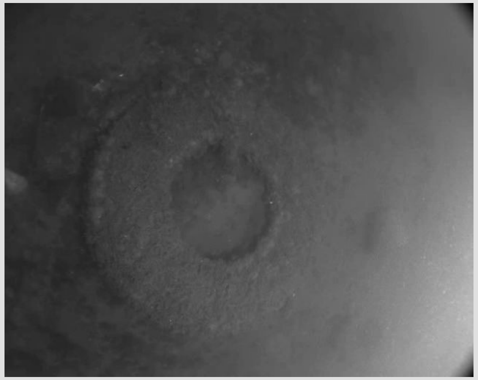
Fouling

AQUALOC (Target)



AQUALOC (Target)

AQUALOC (Target)



TrashCan (Source)

Domain shift evaluation

Training set	Evaluation set	Split	F1-score	mAP@.50
TrashCan	TrashCan	Random	79.7	80.8
TrashCan	TrashCan	K-folded	58.4 ± 4.2	56.6 ± 6.3
UNO	UNO	K-folded	67.3 ± 1.5	68.8 ± 1.2

- Evaluation set: 150 annotated images from AQUALOC dataset

Training set	Evaluation set	Split	F1-score	mAP@.50
TrashCan	AQUALOC	K-folded	55.7 ± 1.6	52.5 ± 1.9
UNO	AQUALOC	K-folded	55.6 ± 4.5	55.2 ± 4.7

Perspectives and conclusion

- Extend the methodology to multi-class
- Work on different adaptation domain scenarios

AQUALOC video

AQUALOC video