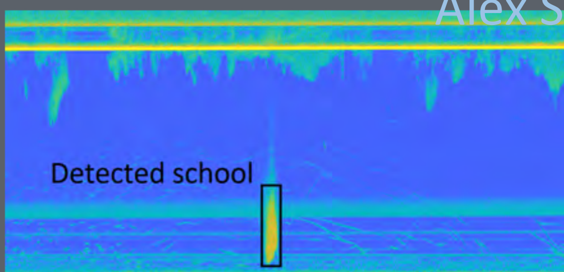
Computer Vision-Based Detection of Schools of Herring from Acoustic Backscatter Time Series

Alireza Rezvanifar[1], Tunai Porto Marques[1], Melissa Cote[1], **Alexandra Branzan Albu**[1],
Alex Slonimer[2], Thomas Tolhurst[2], Kaan Ersahin[2], Todd Mudge[2], and Stéphane Gauthier[3]

[1]University of Victoria, Victoria, BC, Canada. E-mail: aalbu@uvic.ca
[2]ASL Environmental Sciences Inc., Victoria, BC, Canada
[3]Fisheries and Oceans Canada - Institute of Ocean Sciences, Sidney, BC, Canada



PICES 2019 Annual Meeting

W15: Application of machine learning to ecosystem change issues in the North Pacific

Victoria, BC, Canada

October 17-18, 2019

University of Victoria

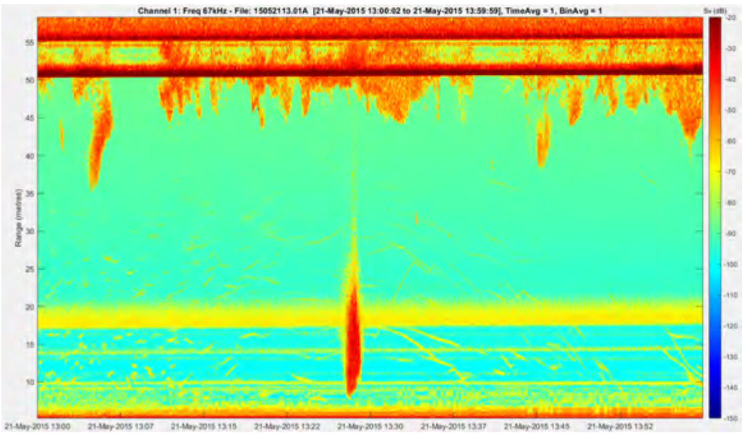
ASL Environmental Sciences

Fisheries and Oceans Canada

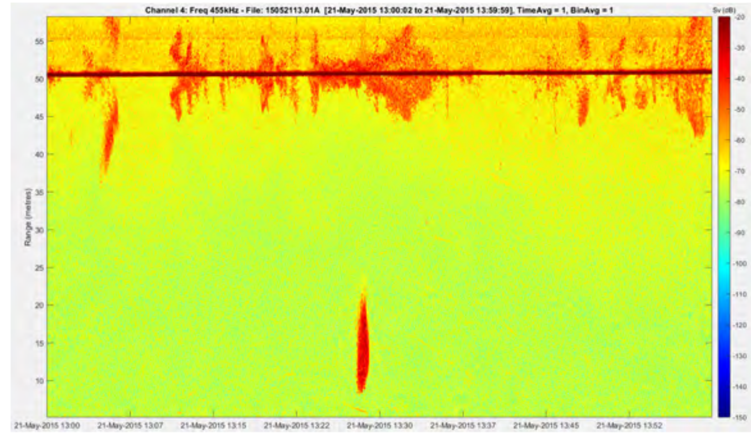
NSERC CRSNG

Context

- Study of acoustic backscatter:
    - Thorough, non-invasive approach
    - Allows to monitor underwater sites for ecosystem changes
- Data:
    - Acquired via multifrequency echosounders (e.g. AZFPs)
    - Visualized as 2D images (echograms)



Sample echogram (67 kHz)



Sample echogram (455 kHz)

Context

- Challenges:
  - Echograms typically analyzed via manual or semi-automatic methods:
    - Time consuming (tons of data to analyze)
    - Prone to errors and inconsistencies
    - Expensive third-party software (e.g. EchoView)

- Solution:
  - Machine learning can improve data processing and interpretation!

Context

- Collaborative project:

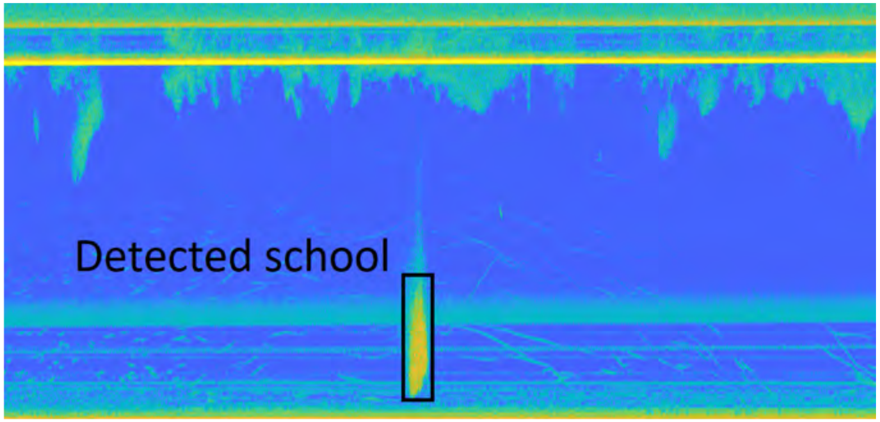
Fisheries and Oceans Canada
(Institute of Ocean Sciences)
Biologists
⇒ Acquire data

University of Victoria
(Computer vision research lab)
Computer engineers
⇒ Develop ML tools

ASL Environmental Sciences
Remote sensing specialists
and acousticians
⇒ Develop echosounders (AZFPs)

Context

- Goal:
  - Explore novel ways to detect visual patterns from echosounder data using computer vision and machine learning techniques
- Case study:
  - Automatic detection of schools of herring from AZFP measurements



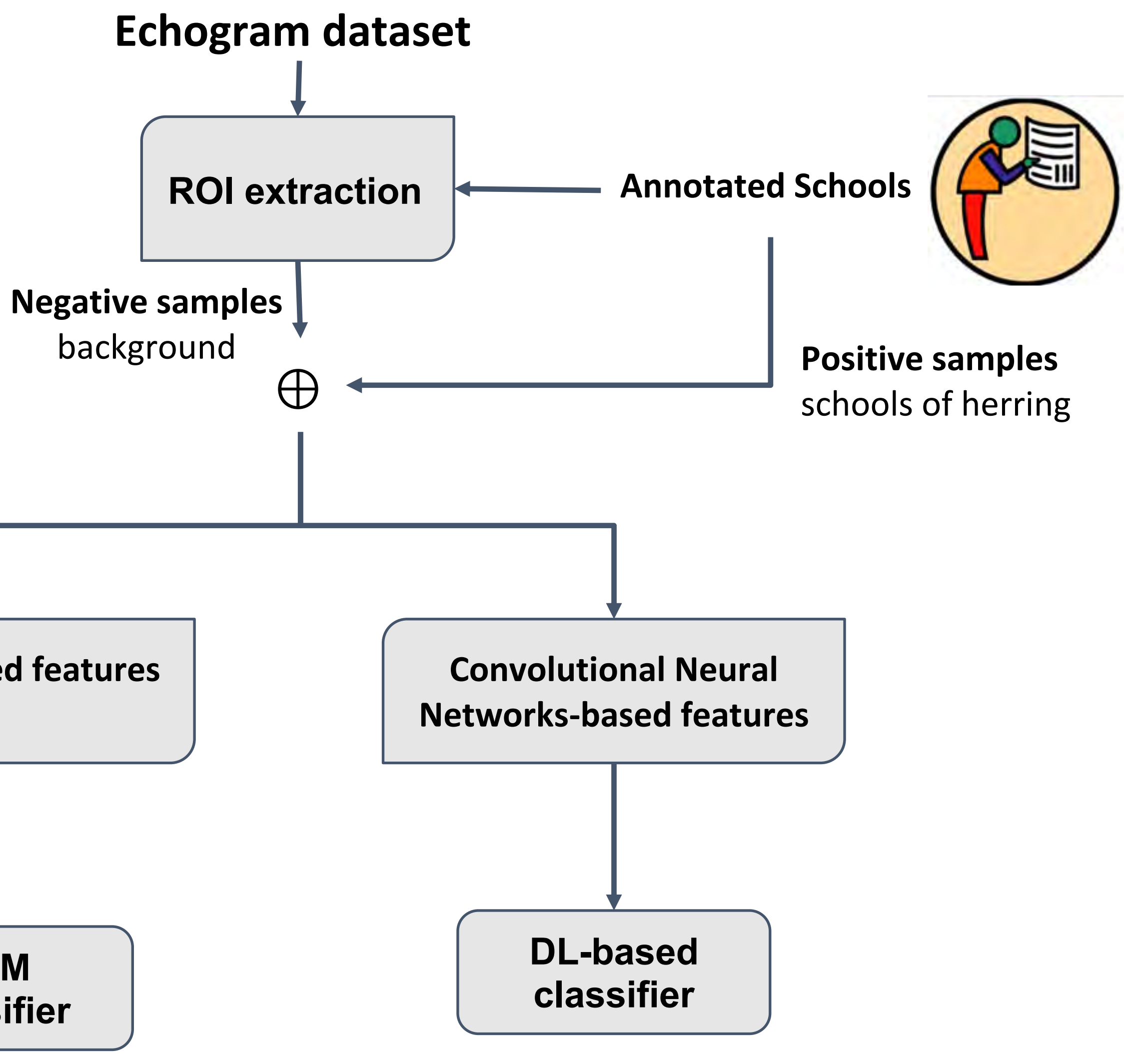
Contributions

1. We propose a dual paradigm approach for fish detection from echograms
   - Classical machine learning paradigm
   - Deep learning paradigm: novel application that goes beyond the few existing works

2. Our framework automates acoustic survey analyses
   - Will reduce processing times, required man-power, and inconsistencies in the results
   - Potential to be scaled to handle additional underwater species (e.g. salmon, zooplankton, etc.)

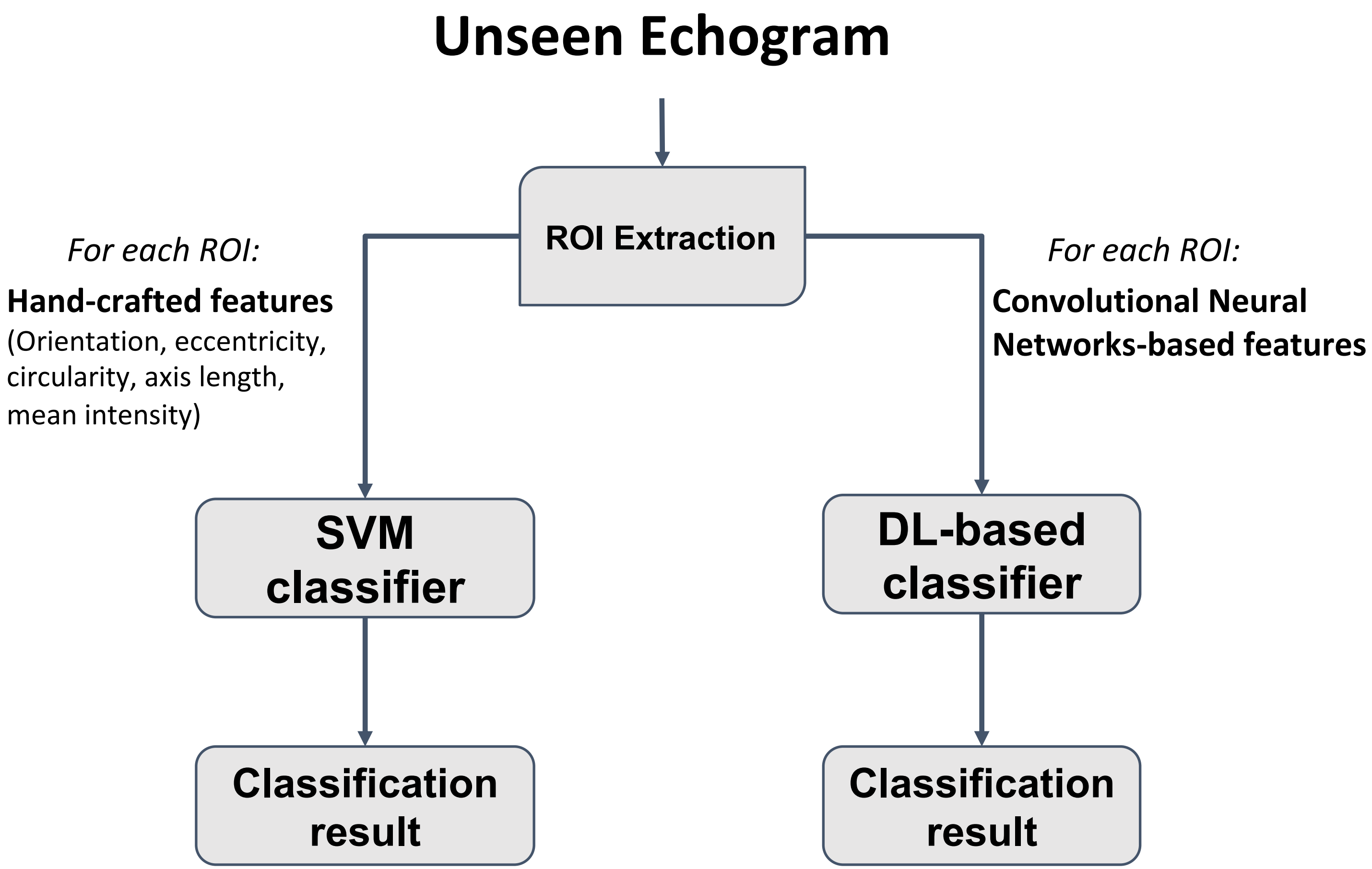
Proposed Method: Overview

***Training* Phase**

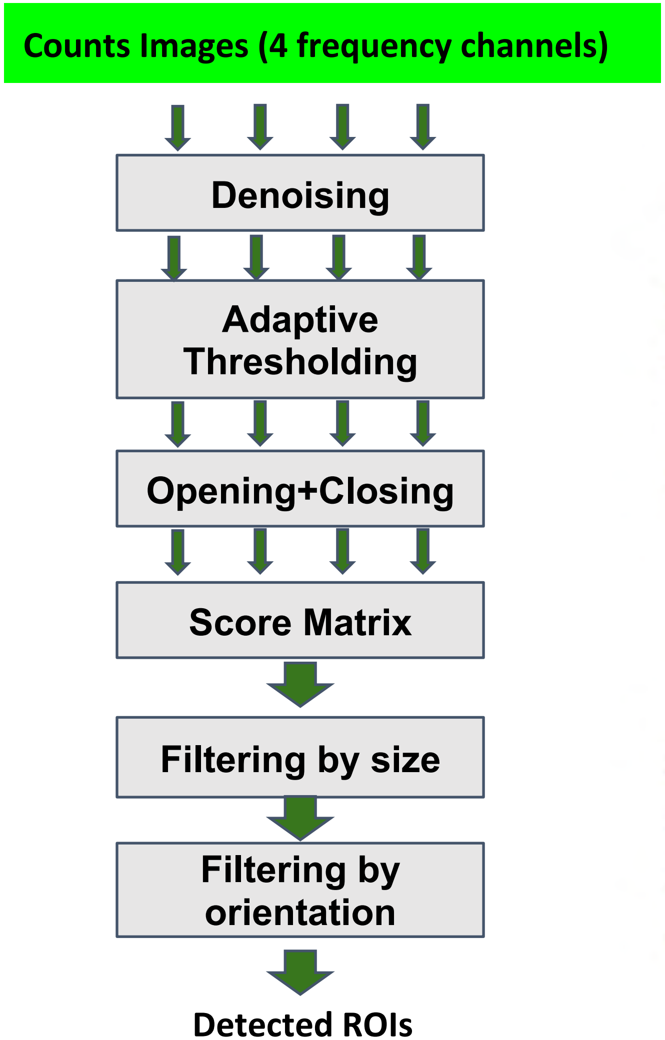


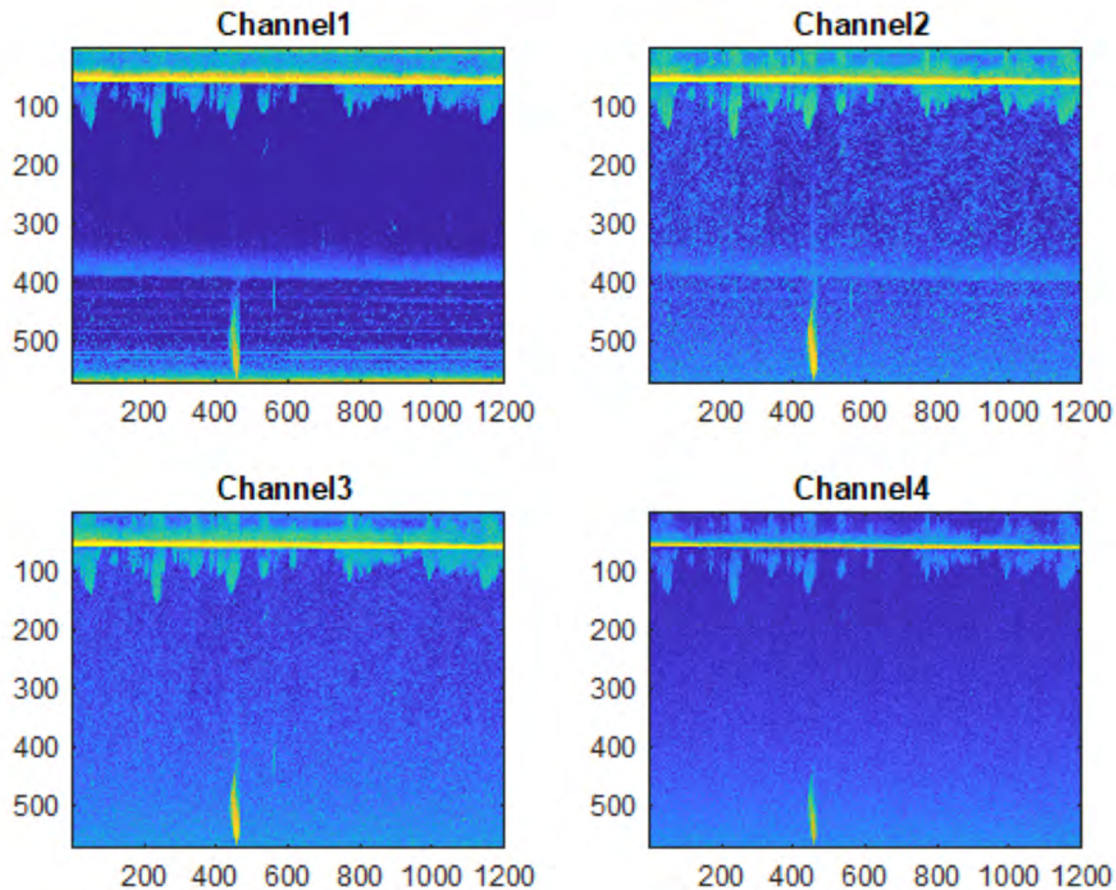
Proposed Method: Overview

***Inference* Phase**

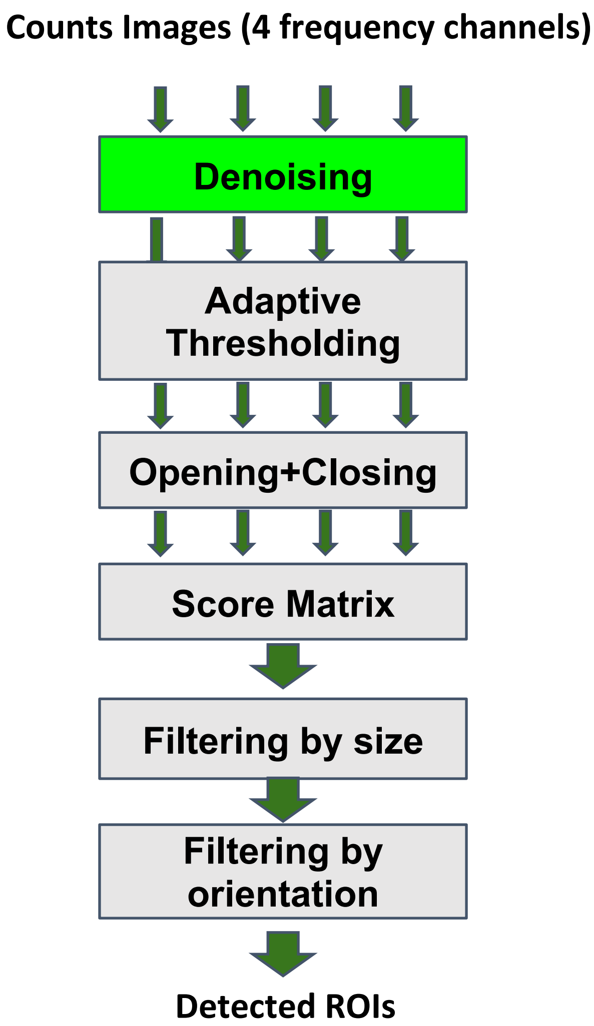


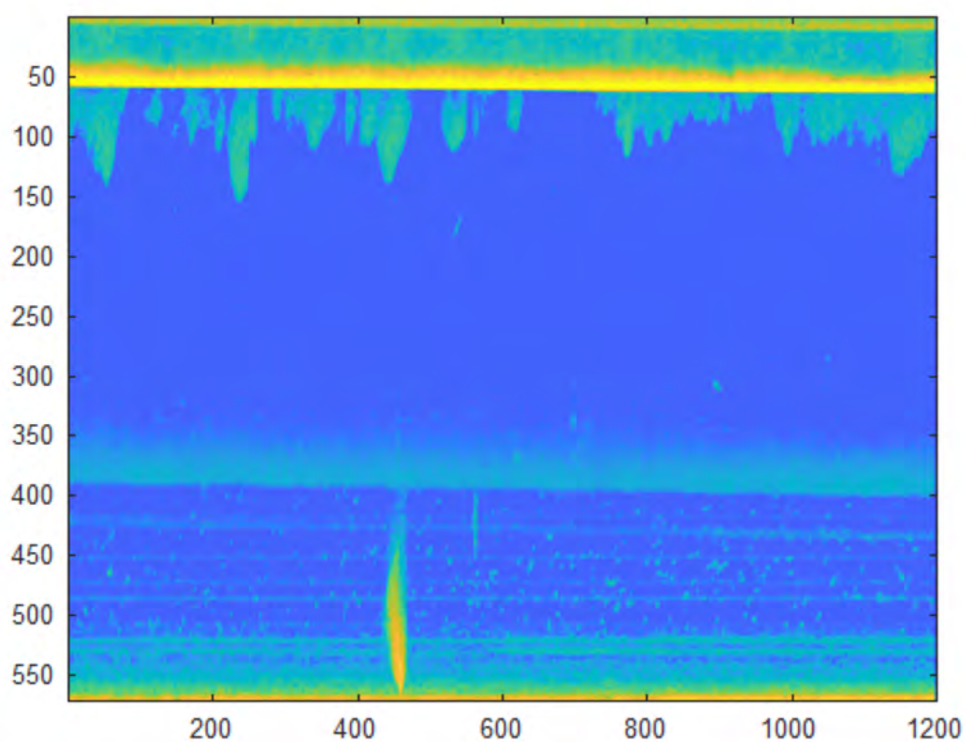
ROI Extraction





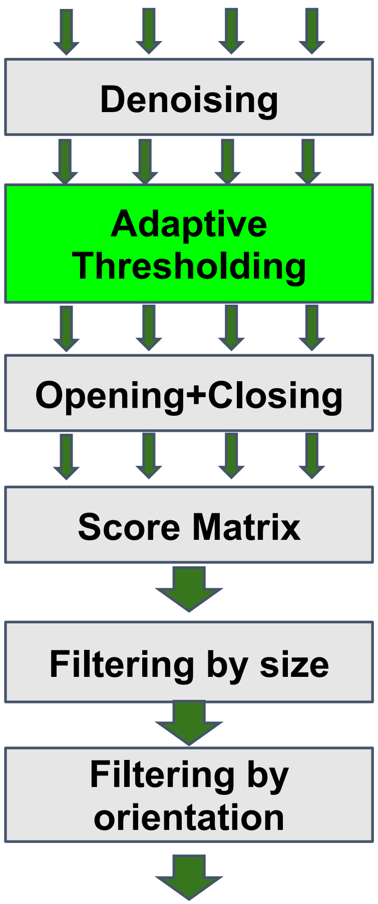
ROI Extraction



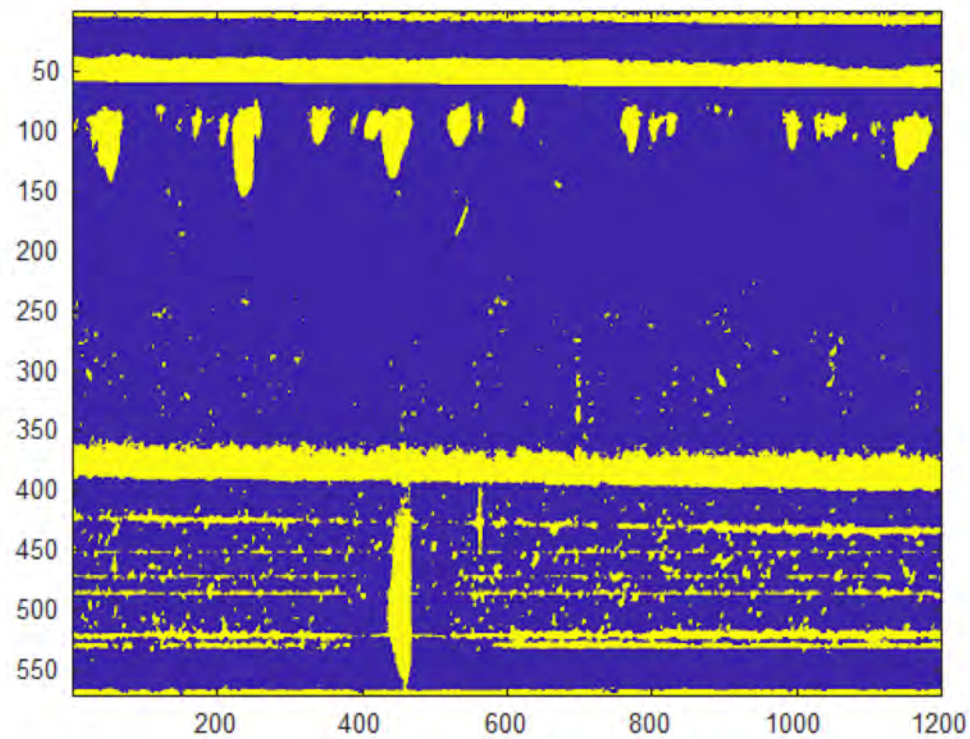


ROI Extraction

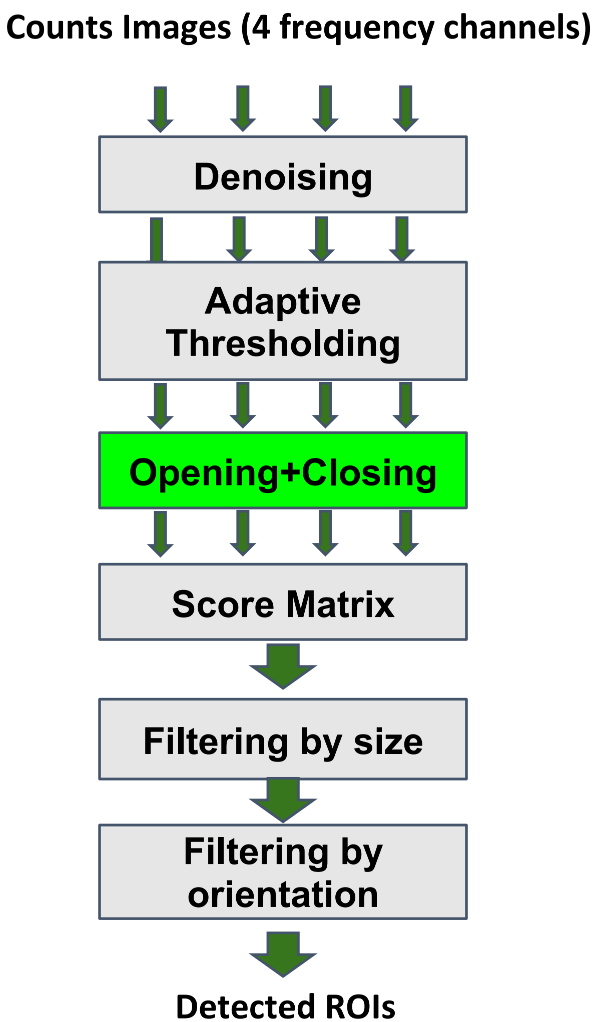
Counts Images (4 frequency channels)

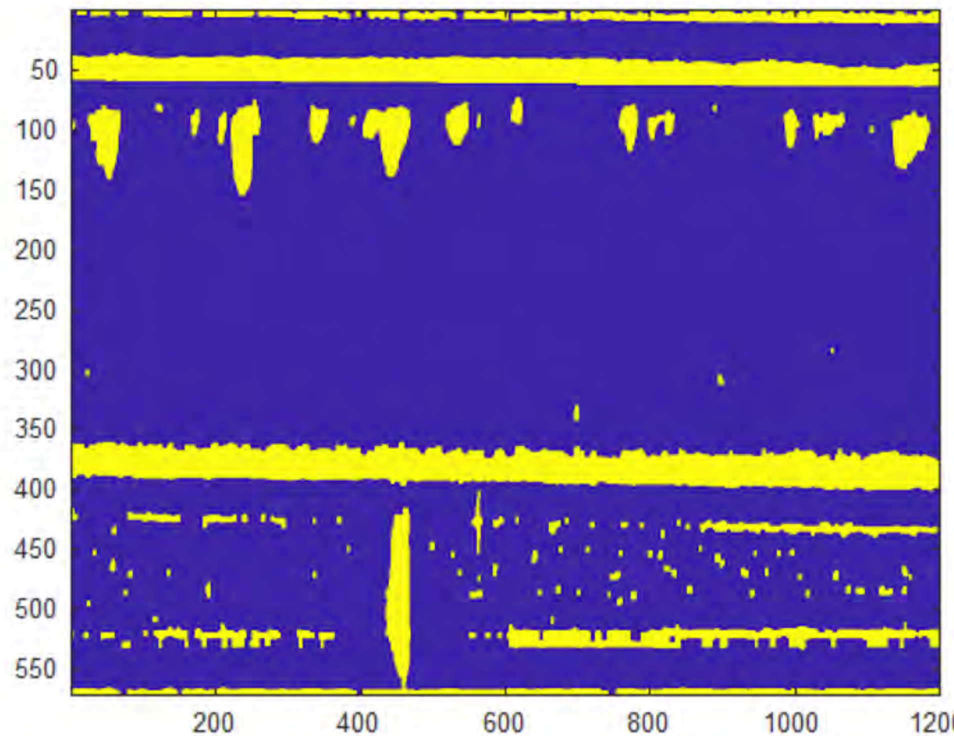


Detected ROIs



ROI Extraction





Computer Vision-Based Detection of Schools of Herring

Counts Images (4 frequency channels)

↓ ↓ ↓ ↓

Denoising

↓ ↓ ↓ ↓

Adaptive Thresholding

↓ ↓ ↓ ↓

Opening+Closing

↓ ↓ ↓ ↓

Score Matrix

↓

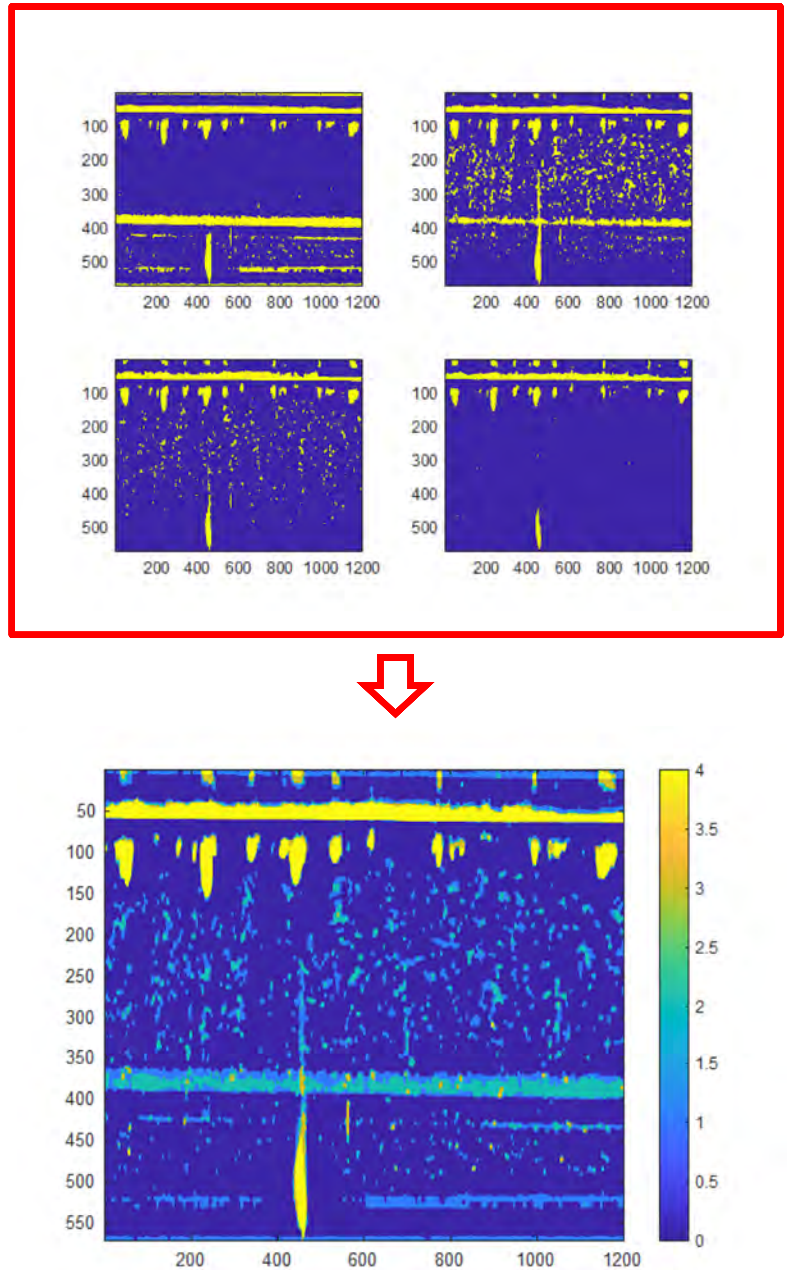
Filtering by size

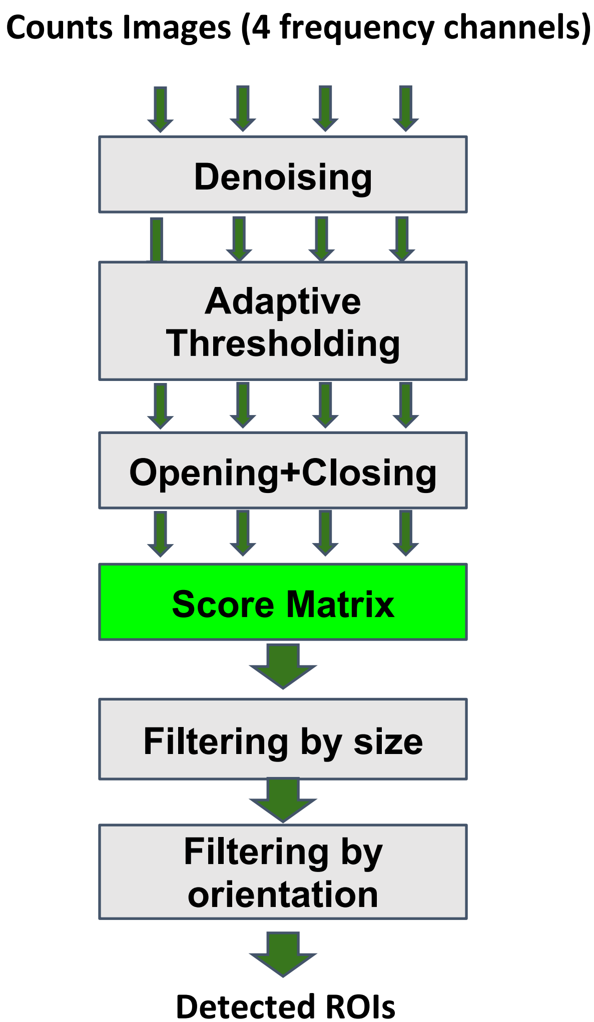
↓

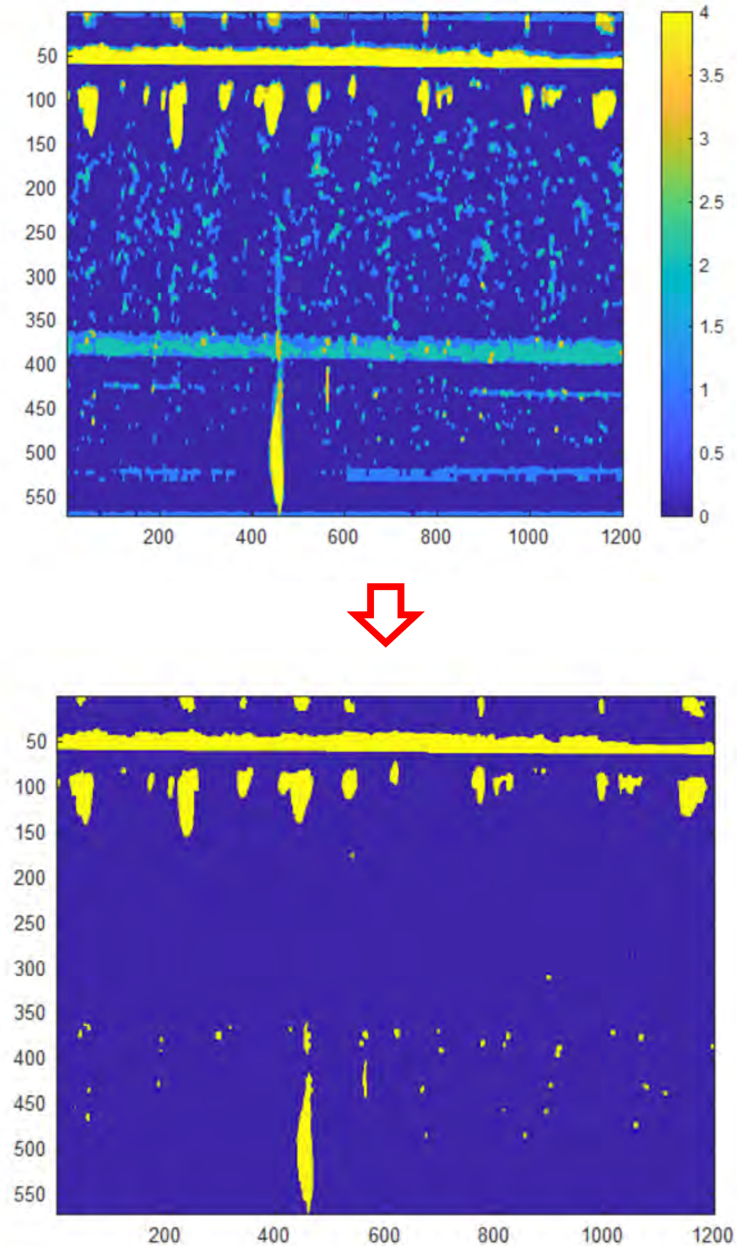
Filtering by orientation

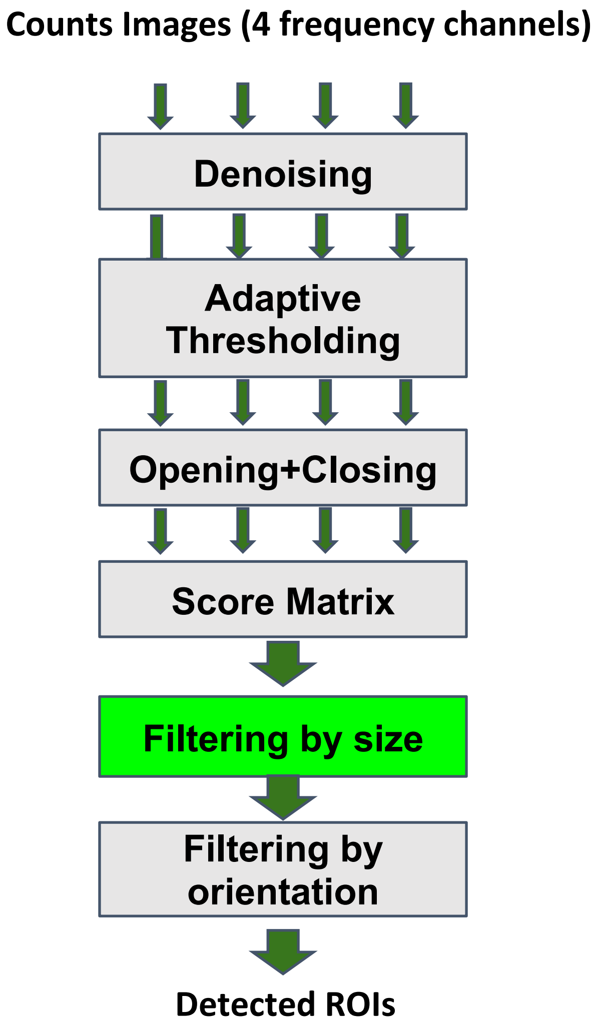
↓

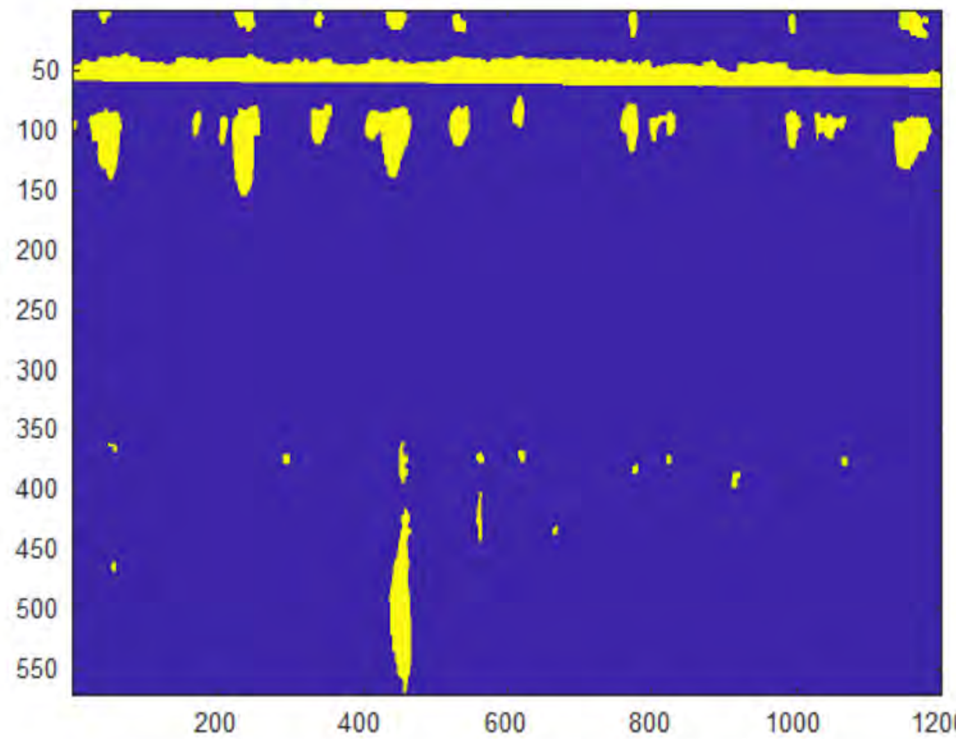
Detected ROIs




Counts Images (4 frequency channels)
Denoising
Adaptive Thresholding
Opening+Closing
Score Matrix
Filtering by size
Filtering by orientation
Detected ROIs


50
100
150
200
250
300
350
400
450
500
550
200
400
600
800
1000
1200
4
3.5
3
2.5
2
1.5
1
0.5
0


Counts Images (4 frequency channels)
Denoising
Adaptive Thresholding
Opening+Closing
Score Matrix
Filtering by size
Filtering by orientation
Detected ROIs


50
100
150
200
250
300
350
400
450
500
550
200
400
600
800
1000
1200

Counts Images (4 frequency channels)

Denoising

Adaptive Thresholding

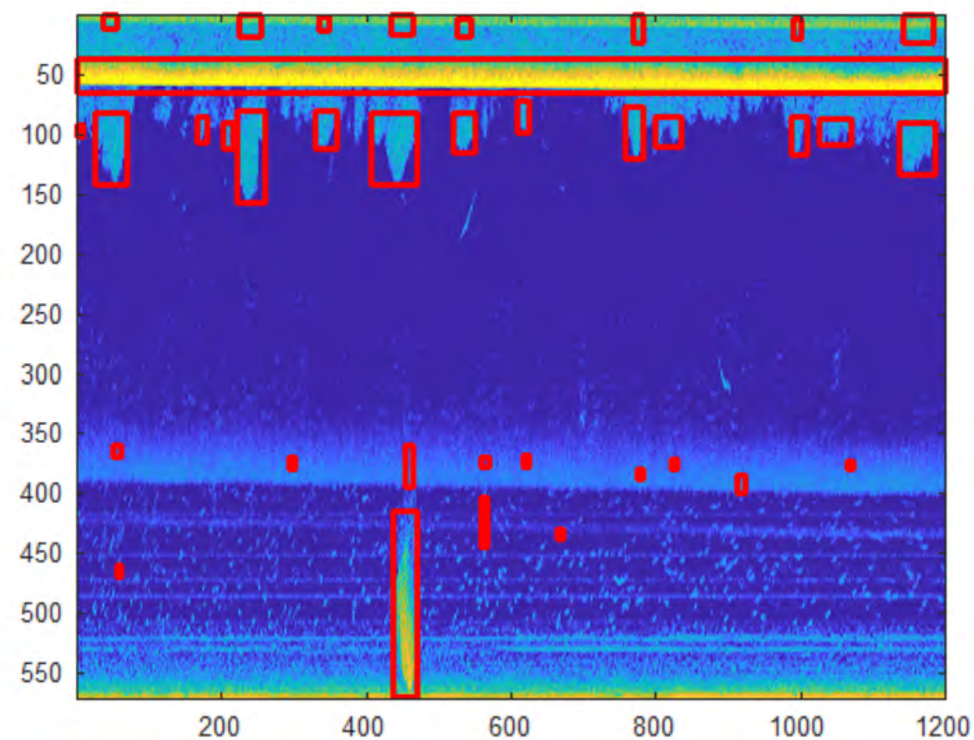
Opening+Closing

Score Matrix

Filtering by size

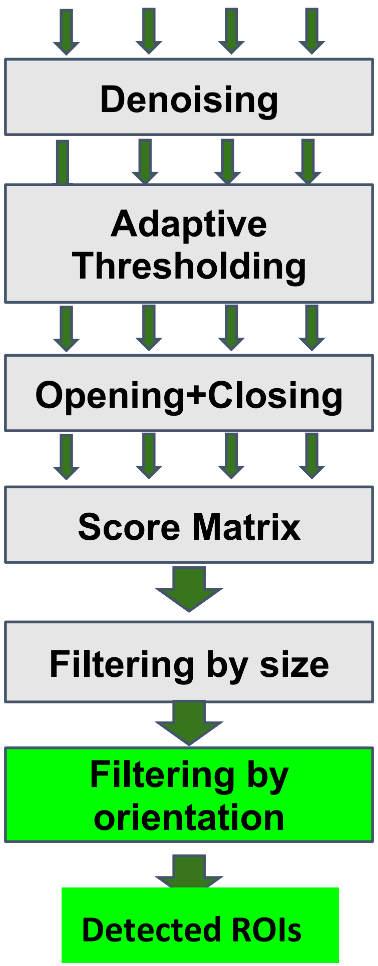
Filtering by orientation

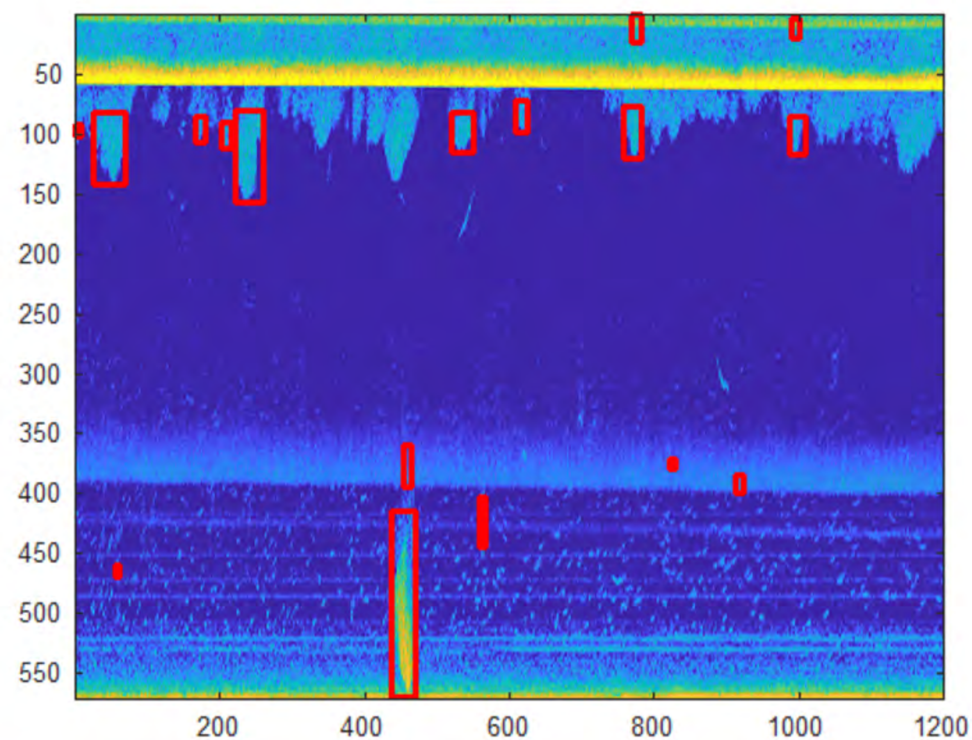
Detected ROIs



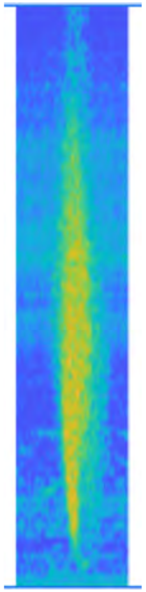
Computer Vision-Based Detection of Schools of Herring

Counts Images (4 frequency channels)





Classical ML: Classification



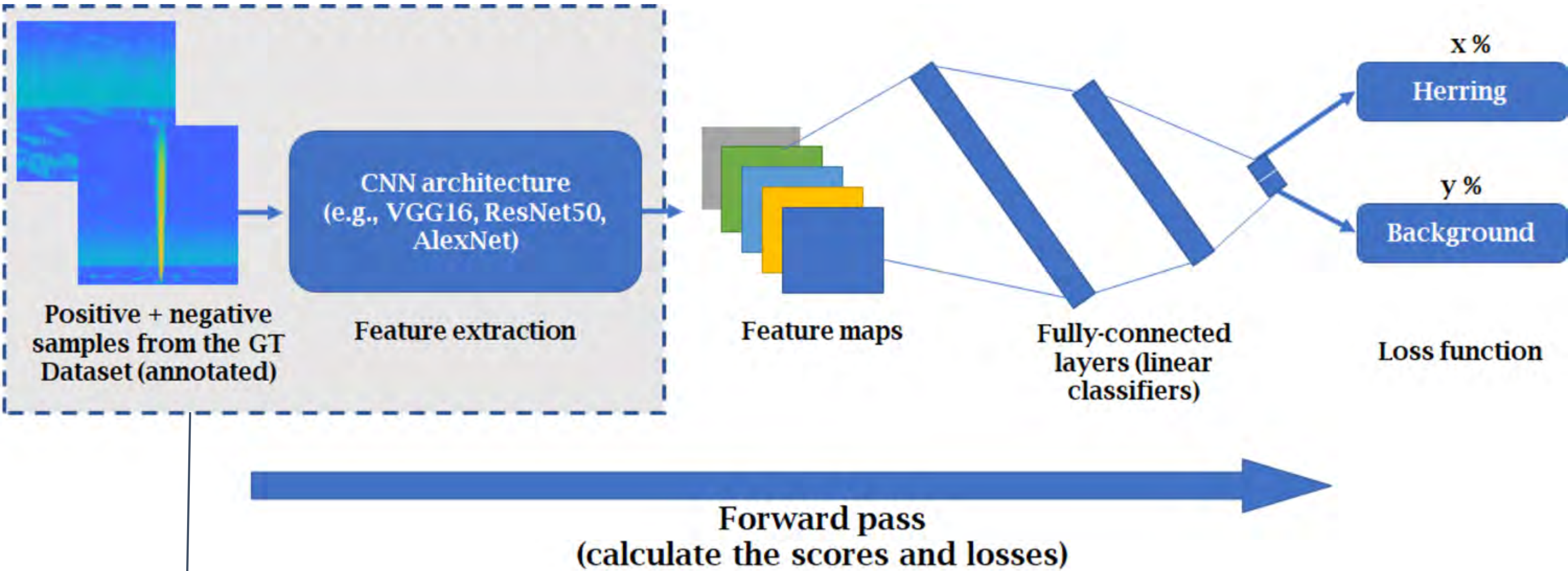
- Features:
  - The objective is to engineer the best set of features based on contextual information of the schools
  - Features should reflect the appearance and geometry of the schools
  - Selected features are:
    - Mean intensity of regions
    - Ratio between the minor axis to the major axis of an ellipse that has the same normalized second central moments as the region
    - Eccentricity: how much the center of mass differs from the center of the circumscribed circle
    - Circularity: specifies the roundness of object
- Classifier:
  - The still popular Support Vector Machines (SVM) classifier with linear kernel is utilized

Deep-Learning Classification

*The use of deep learning frameworks can automate the classification task by computing discriminant features, **regardless of object class.***

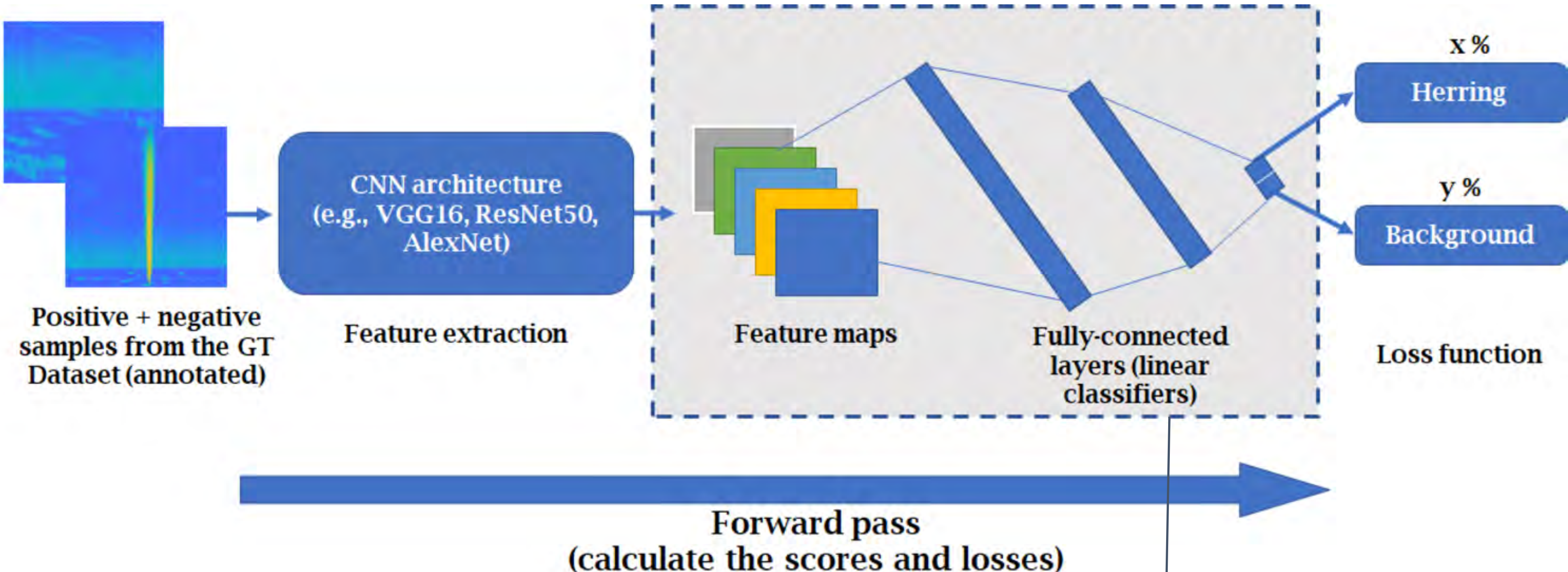
- 1. Use a Convolutional Neural Network (CNN)-based architectures for the automatic extraction of features
- 2. Use the extracted features as inputs of a fully connected network (FCN) that generates *predictions*
- 3. Calculate the loss based on the ground truth data
- 4. Use backpropagation to update network parameters, yielding better predictions

Deep-Learning Classification



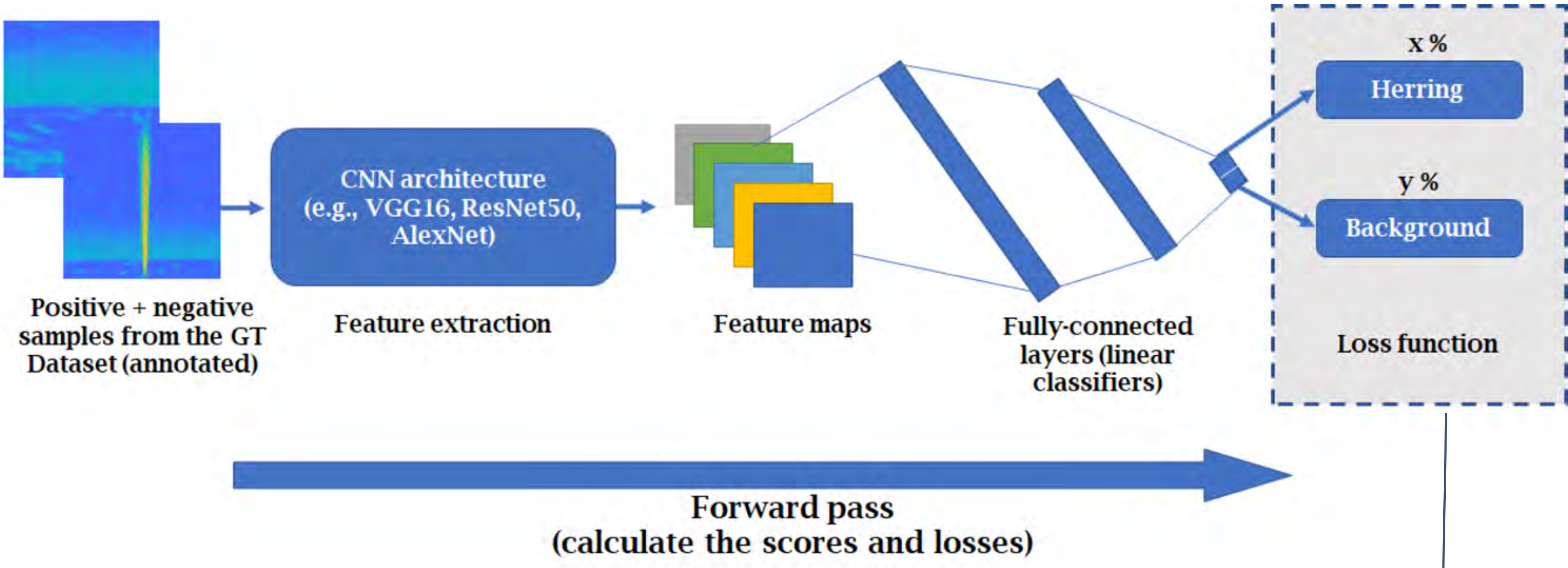
1. Use a Convolutional Neural Network (CNN)-based architectures for the automatic extraction of features.

Deep-Learning Classification



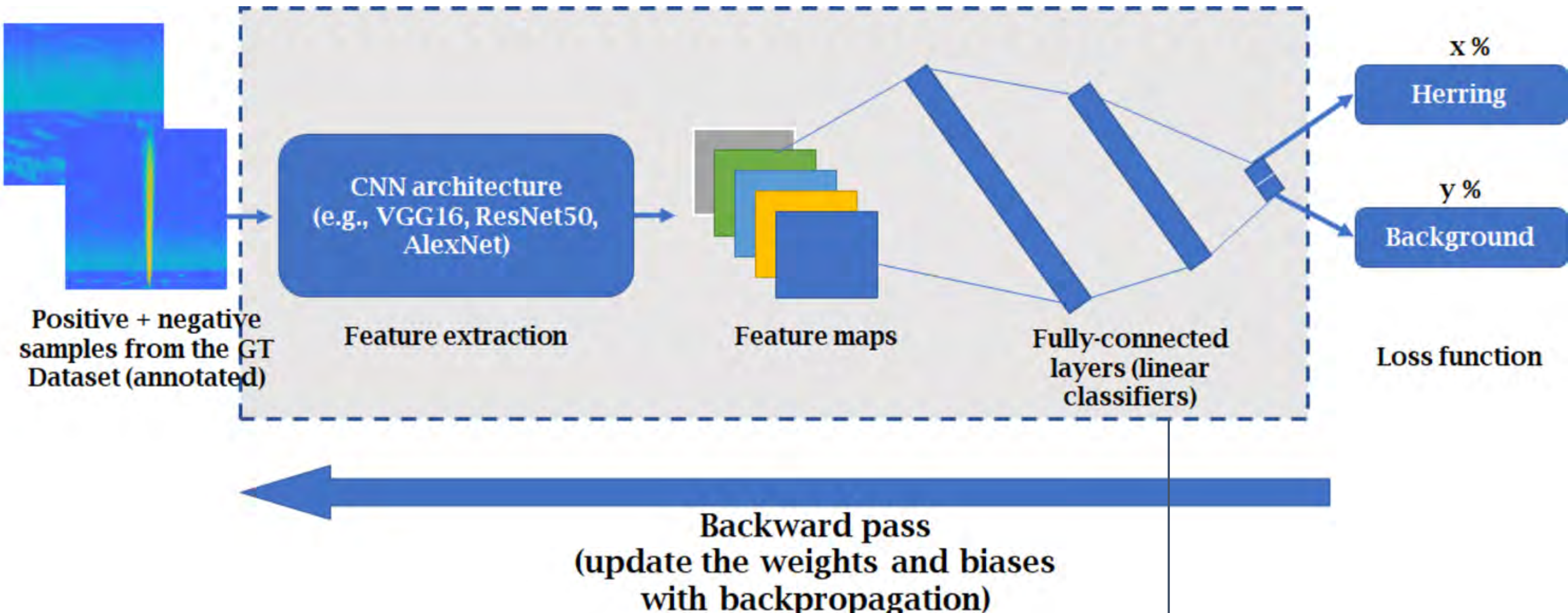
2. Use the extracted features as inputs of a fully connected network (FCN) that generates *predictions*.

Deep-Learning Classification



3. Calculate the loss based on ground truth data.

Deep-Learning Classification



4. Use backpropagation to update network parameters, yielding better predictions.

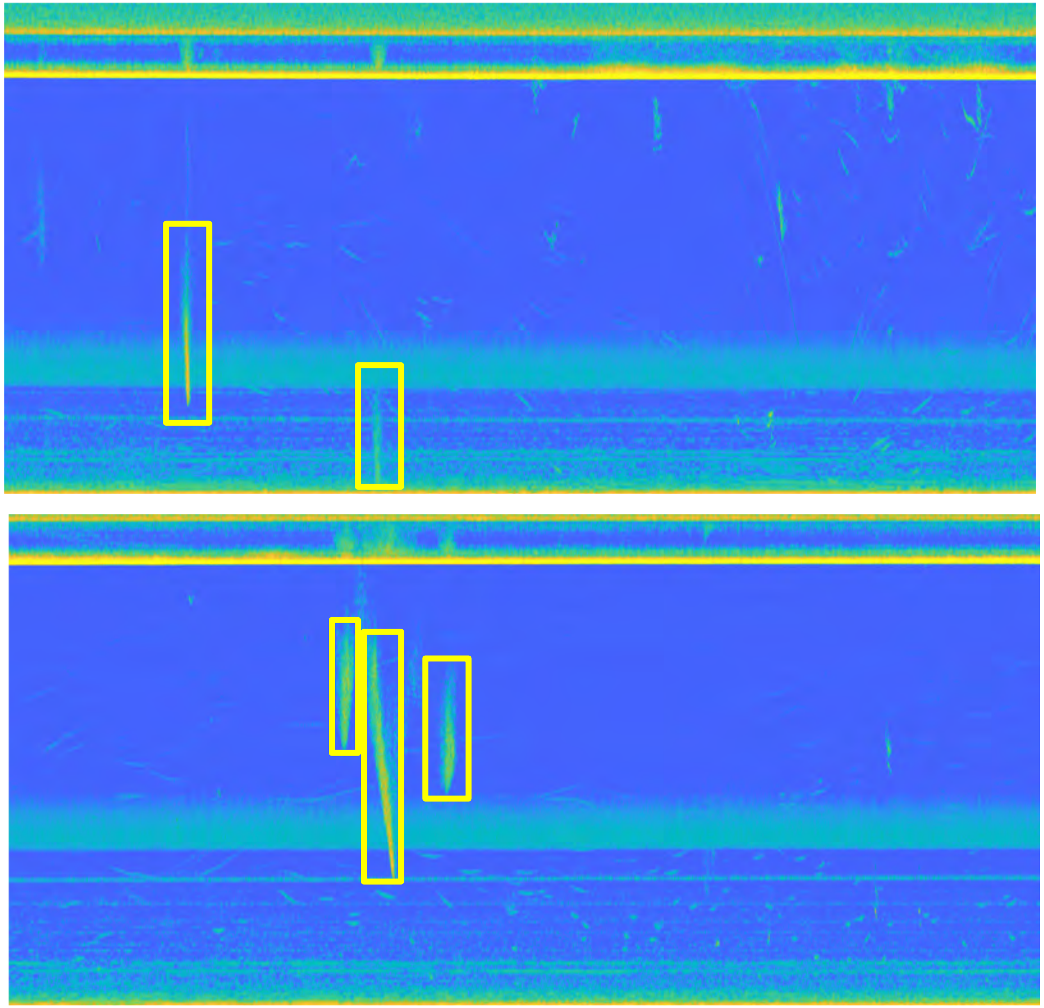
Experimental Results - Dataset

Ground truth dataset

100 echograms

145 samples of schools of herrings

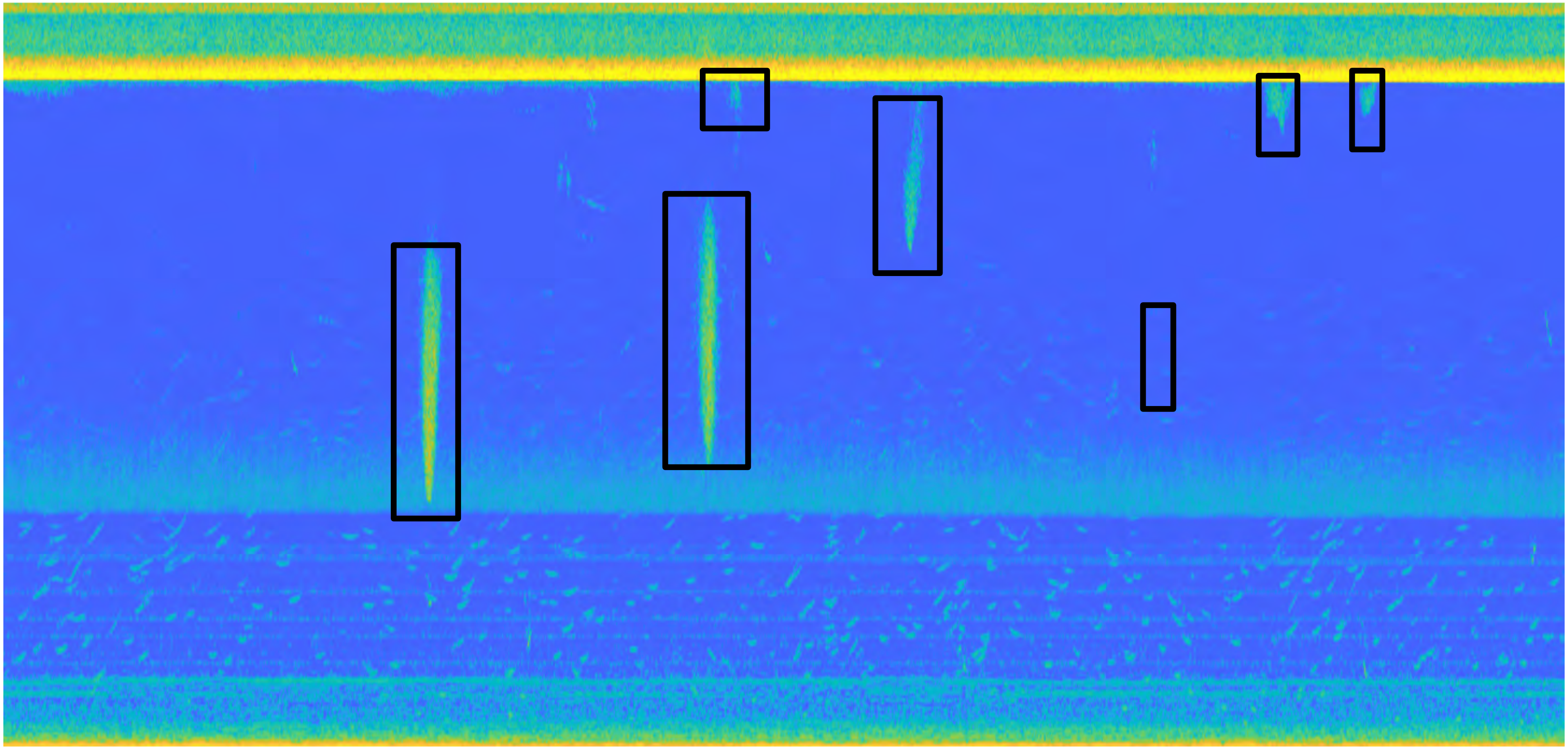
Samples are used for the extraction of hand-crafted features (SVM) and the training of the deep learning-based classifier.



Echograms with annotated samples (yellow bounding boxes)

Experimental Results - Evaluation

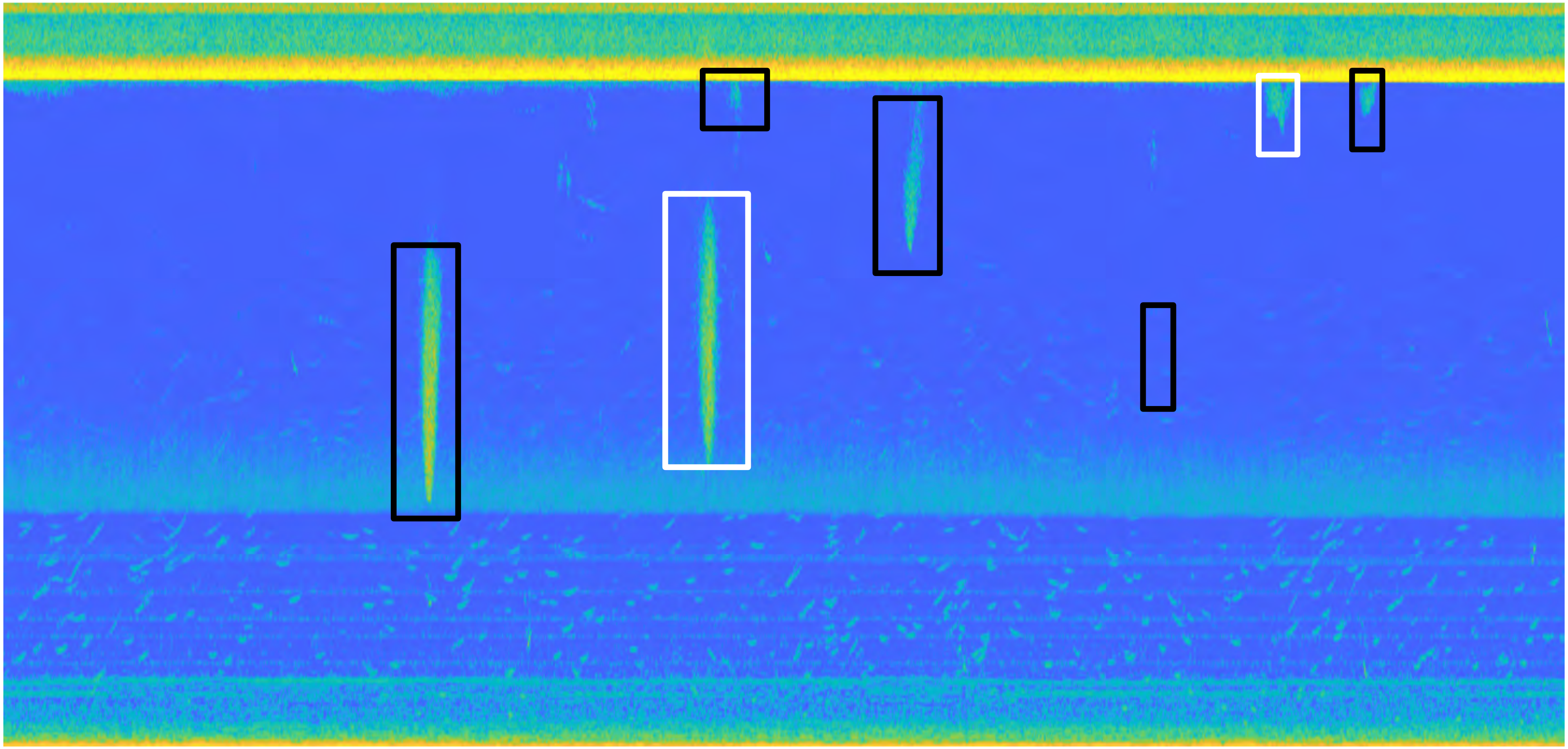
How to determine if an ROI is a true positive?



1. Regions of Interest (ROI extractor output): Black bounding boxes

Experimental Results - Evaluation

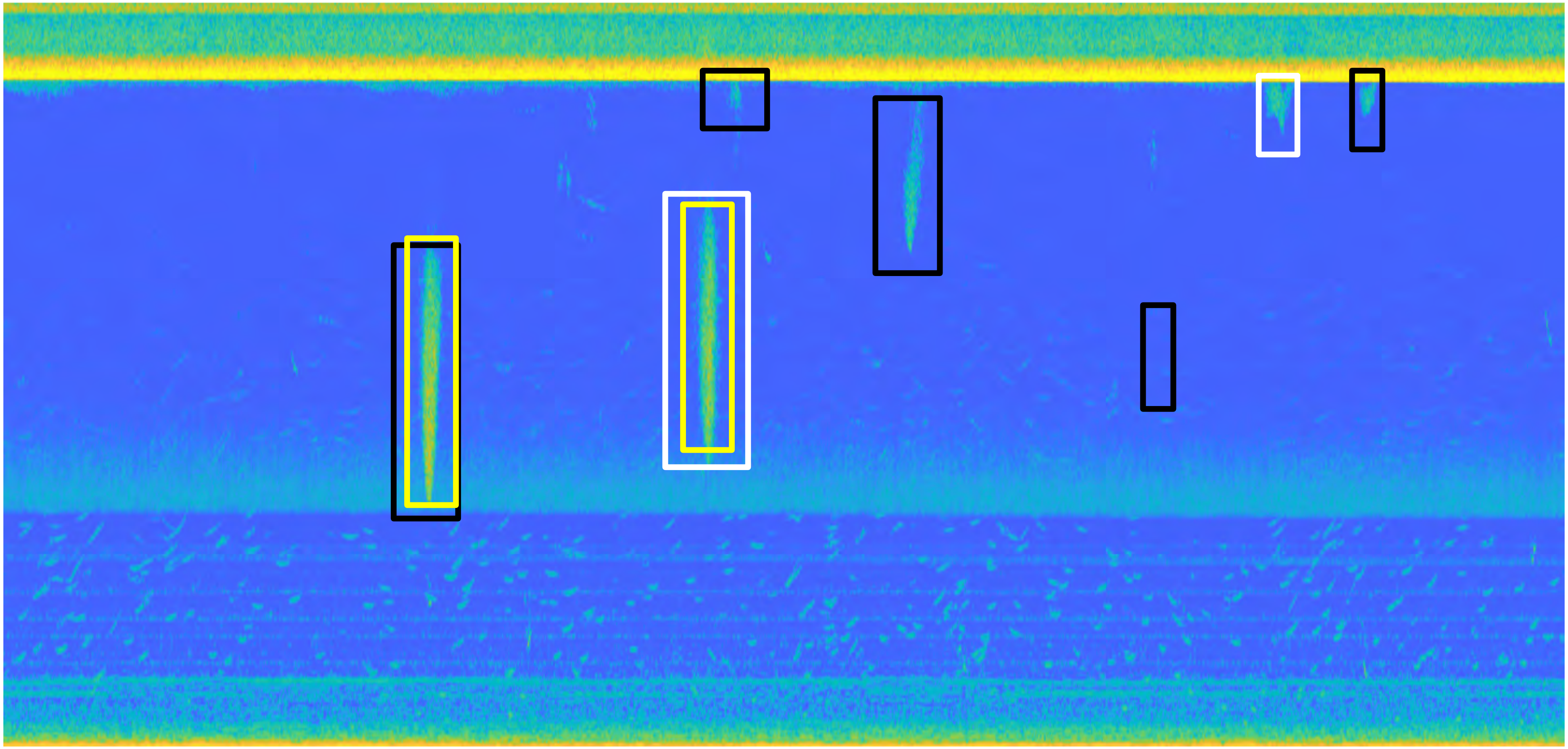
How to determine if an ROI is a true positive?



2. Use SVM (handcrafted-based features) or deep learning-based approach to classify each ROI: white bounding boxes represent prediction of *schools*

Experimental Results - Evaluation

How to determine if an ROI is a true positive?



3. Compare detection with the ground truth: yellow bounding boxes

Experimental Results - Evaluation

How to determine if an ROI is a true positive?

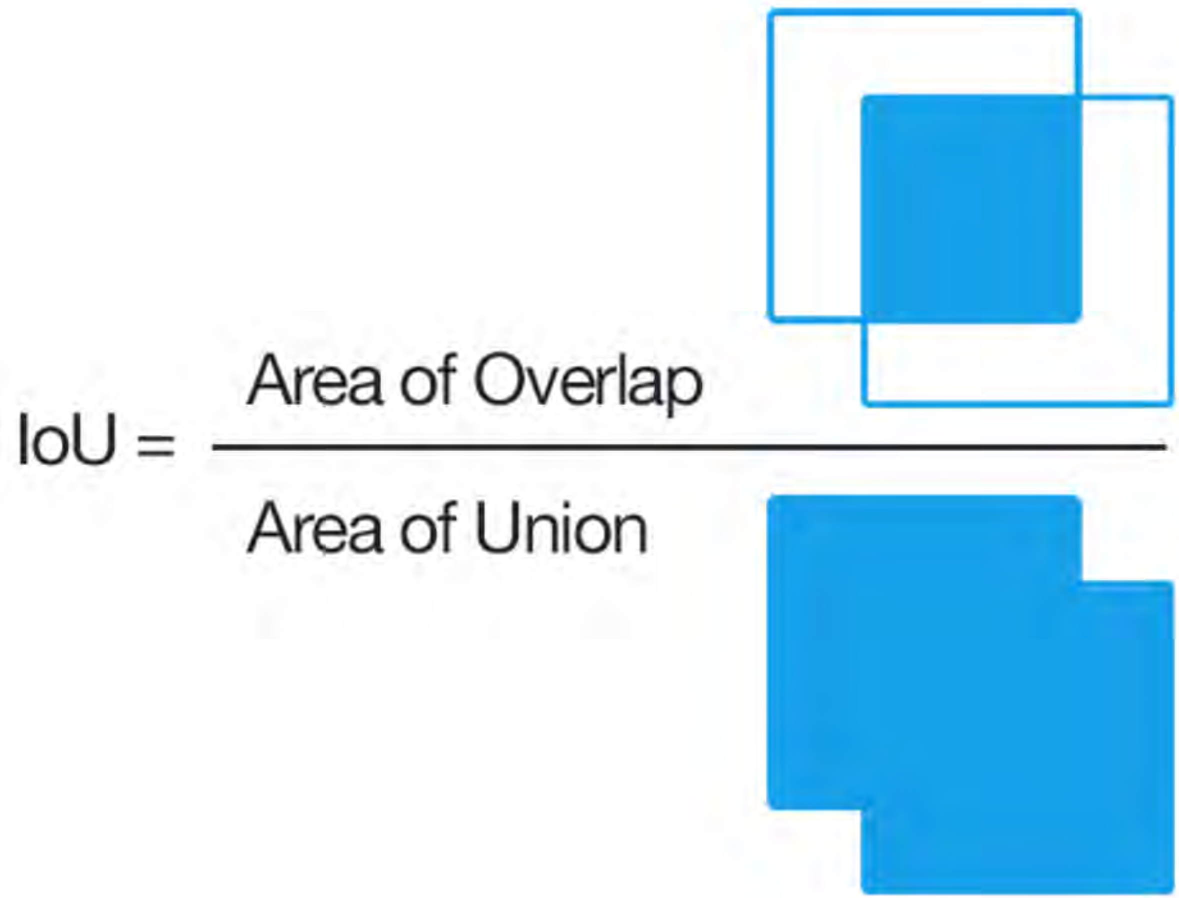
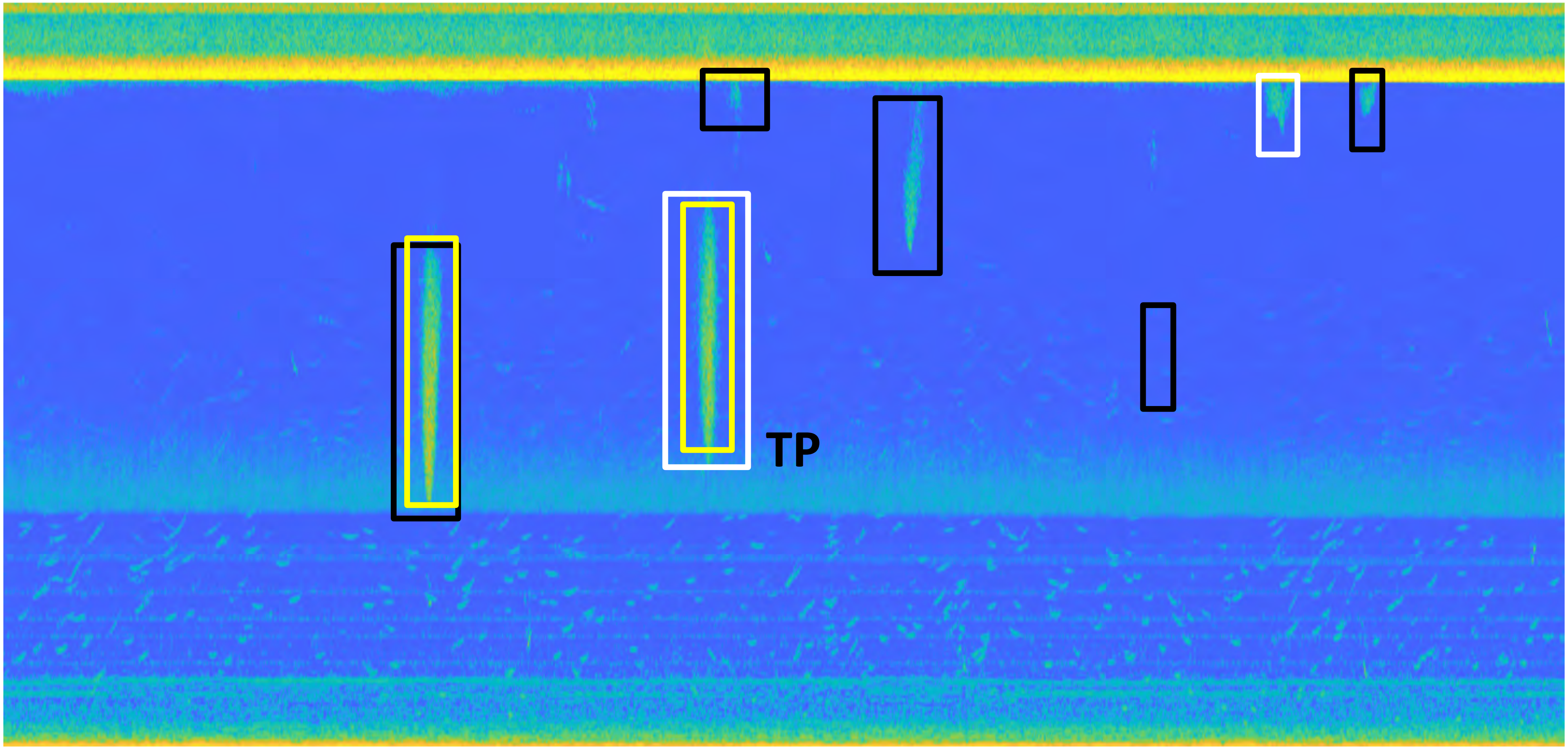


Image retrieved from
https://tinyurl.com/y3gn7mtl

If a detection has an IoU > threshold: **true positive**.

Experimental Results - Evaluation

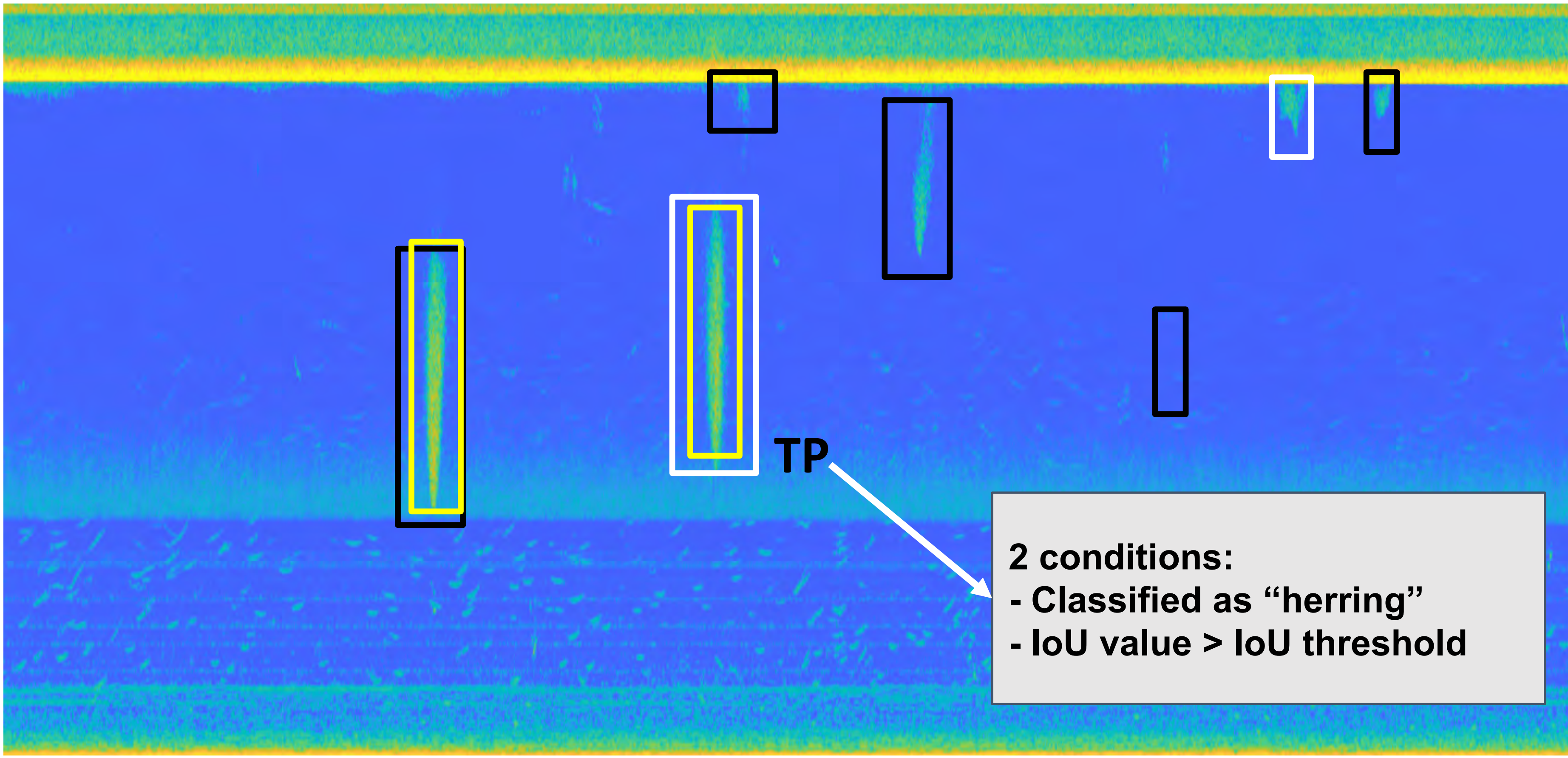
How to determine if an ROI is a true positive?



3. Compare detection with the ground truth: yellow bounding boxes

Experimental Results - Evaluation

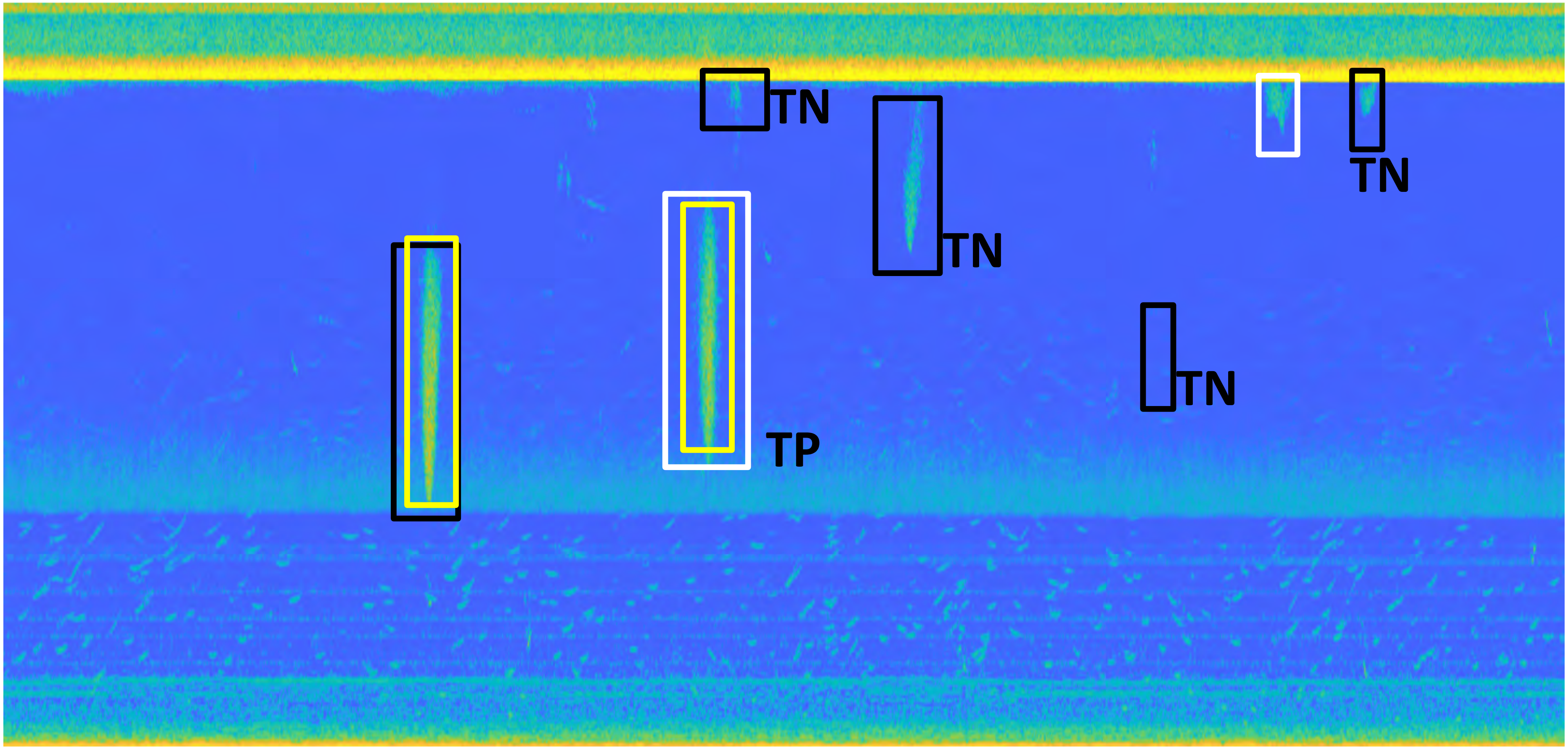
How to determine if an ROI is a true positive?



3. Compare detection with the ground truth: yellow bounding boxes

Experimental Results - Evaluation

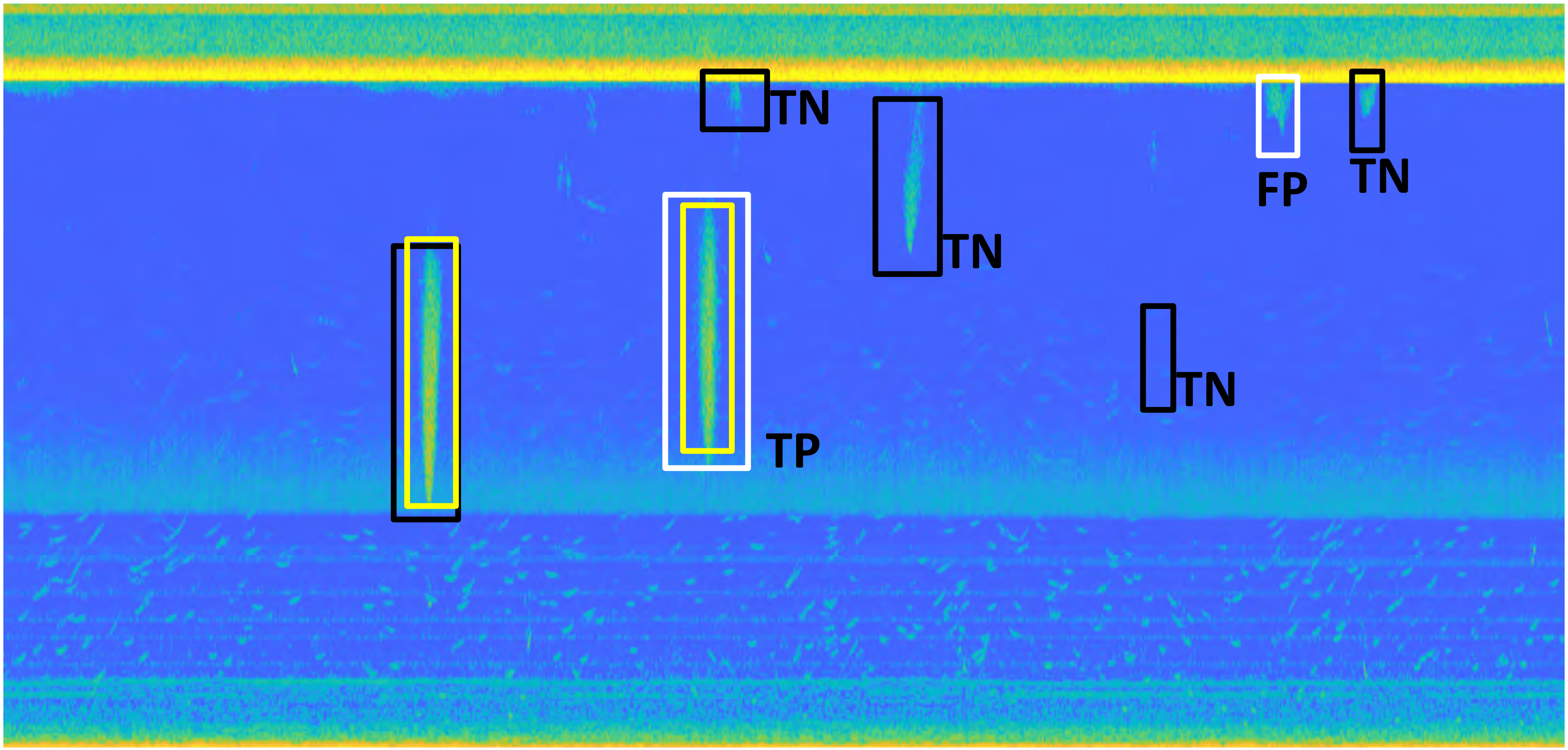
How to determine if an ROI is a true positive?



3. Compare detection with the ground truth: yellow bounding boxes

Experimental Results - Evaluation

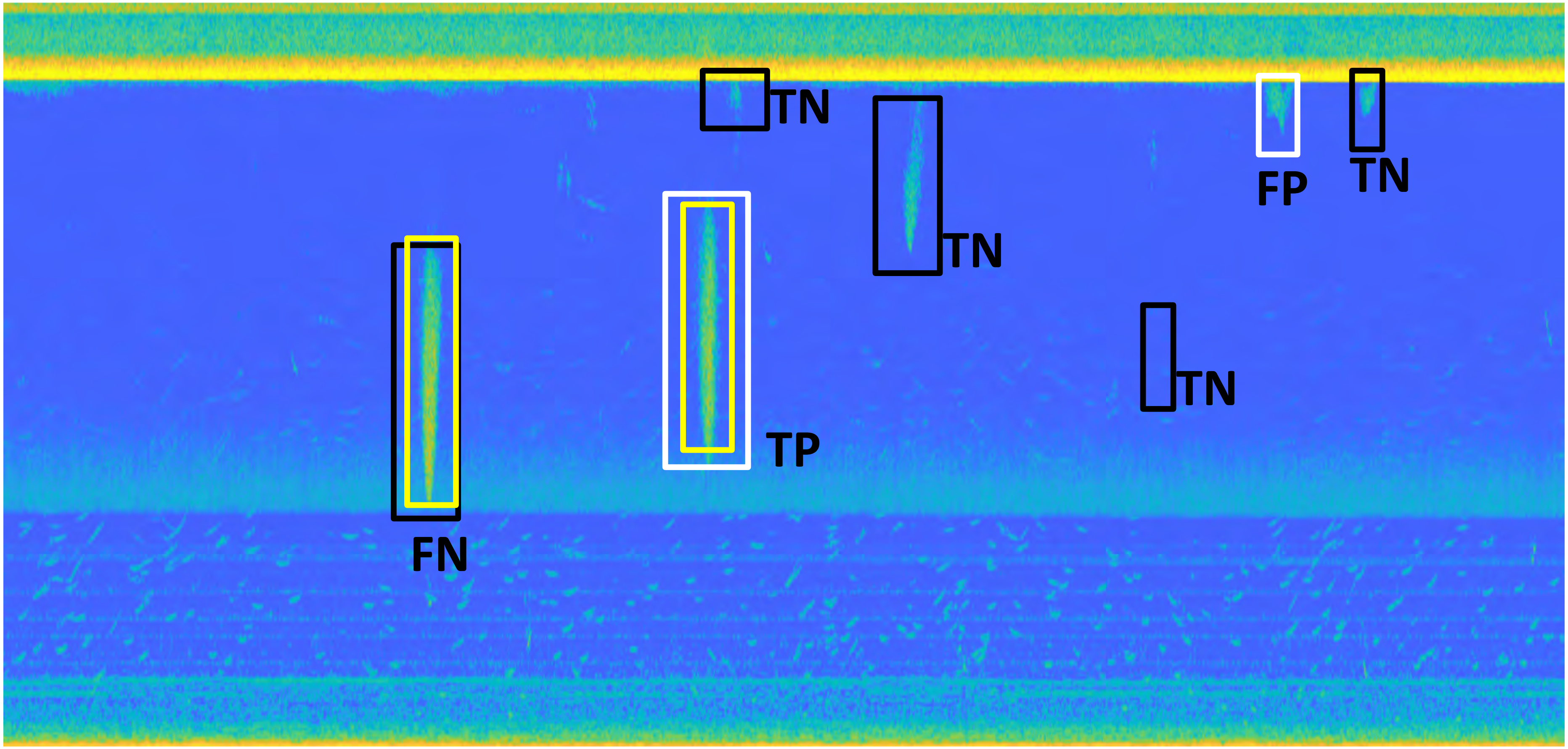
How to determine if an ROI is a true positive?



3. Compare detection with the ground truth: yellow bounding boxes

Experimental Results - Evaluation

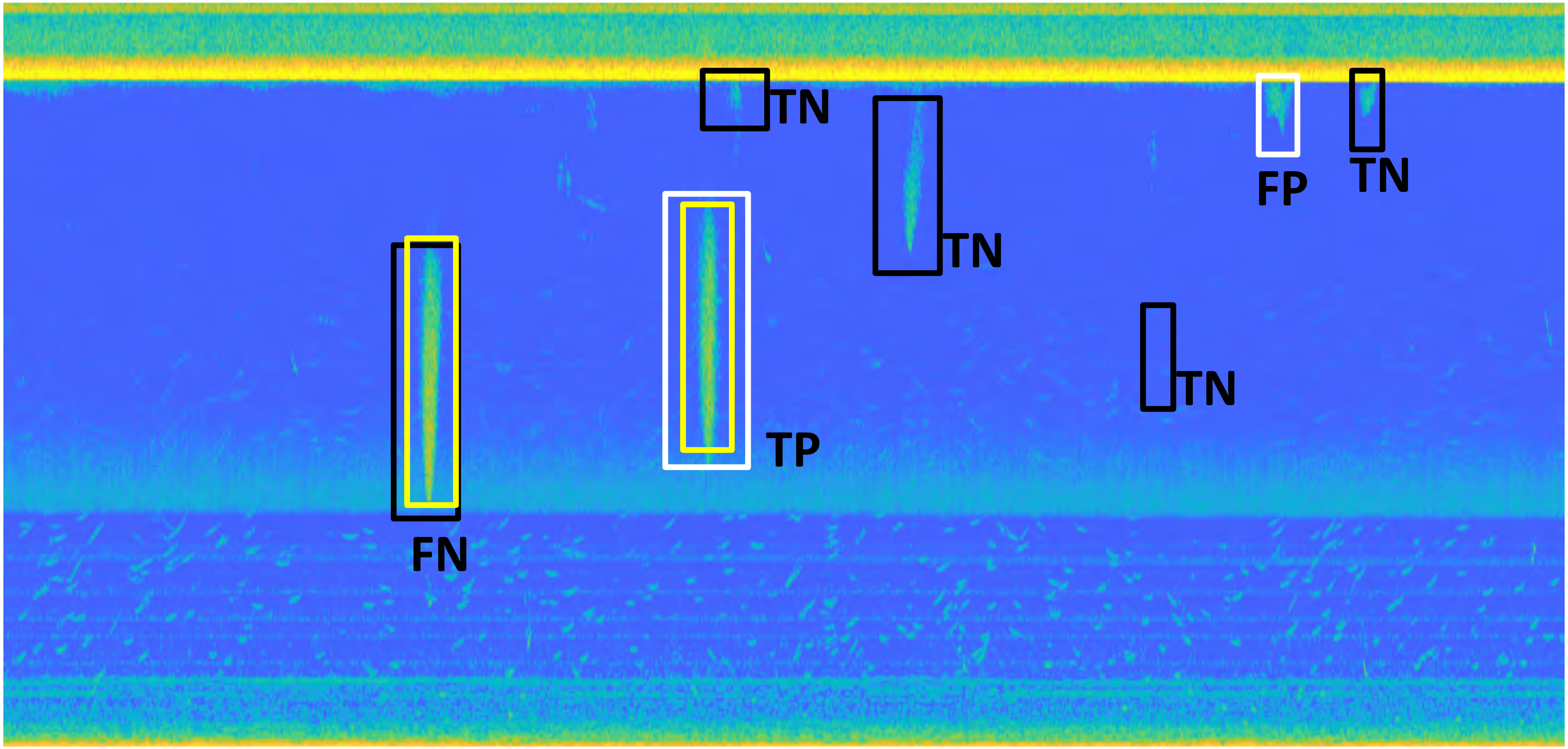
How to determine if an ROI is a true positive?



3. Compare detection with the ground truth: yellow bounding boxes

Experimental Results - Evaluation

How to determine if an ROI is a true positive?



4. Calculate that for all samples in the dataset:

- **100 samples**
- **145 instances of schools of herring**

TP, FP, TN, FN

Experimental Results: Quantitative

- ROI Extractor Evaluation

Precision	Precision	Recall	F1-Score
0.0	0.173	**0.931**	0.292
0.2	0.171	0.917	0.288
0.4	0.155	0.834	0.262

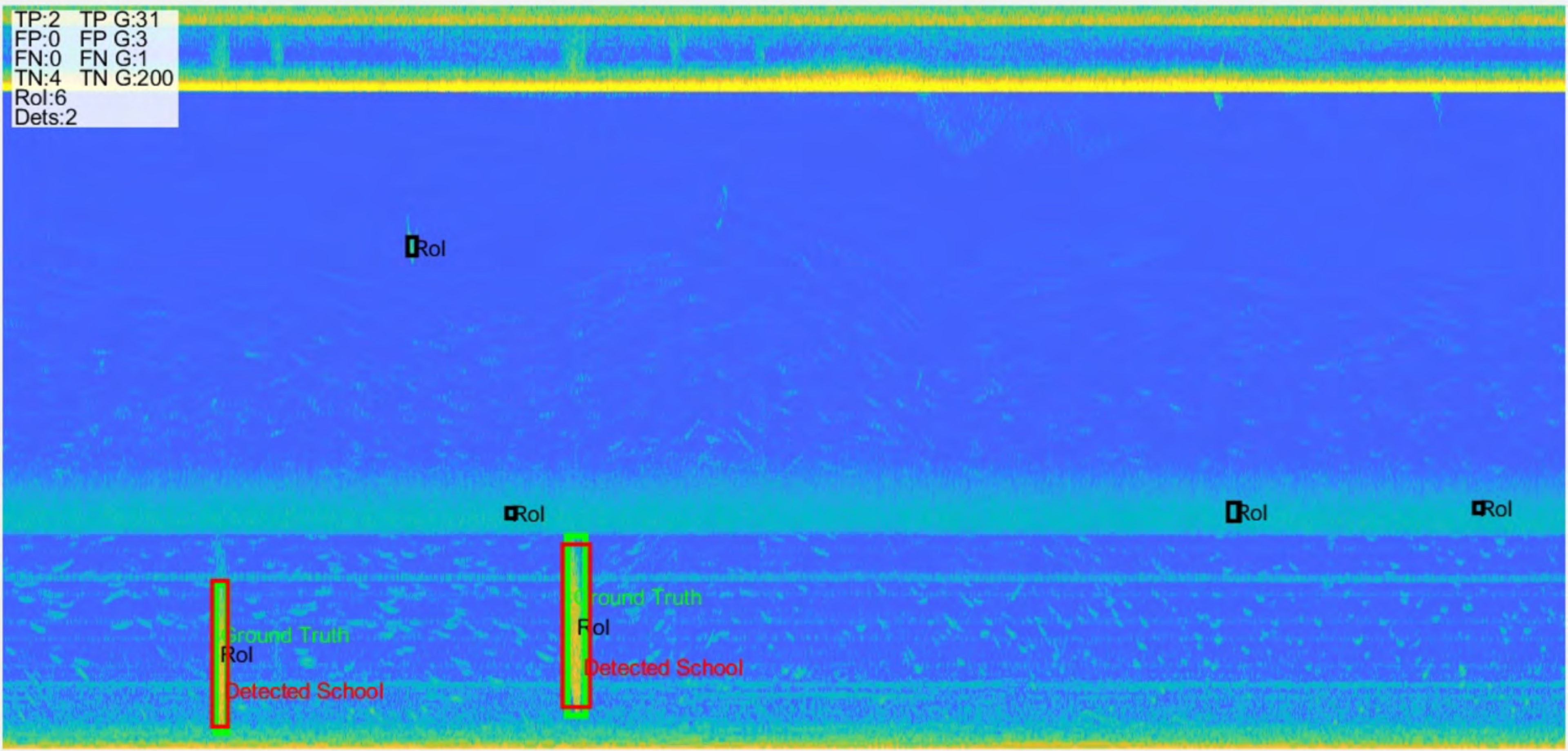
- Entire Framework Evaluation (IoU = 0.4)

Architecture	Precision	Recall	F1-Score
ResNet50	0.77	**0.85**	0.81
DenseNet201	0.78	**0.85**	**0.82**
InceptionNet	**0.81**	0.81	0.81
Baseline (SVM)	0.51	0.78	0.62

Experimental Results: Qualitative

SVM (IoU threshold 0.4)

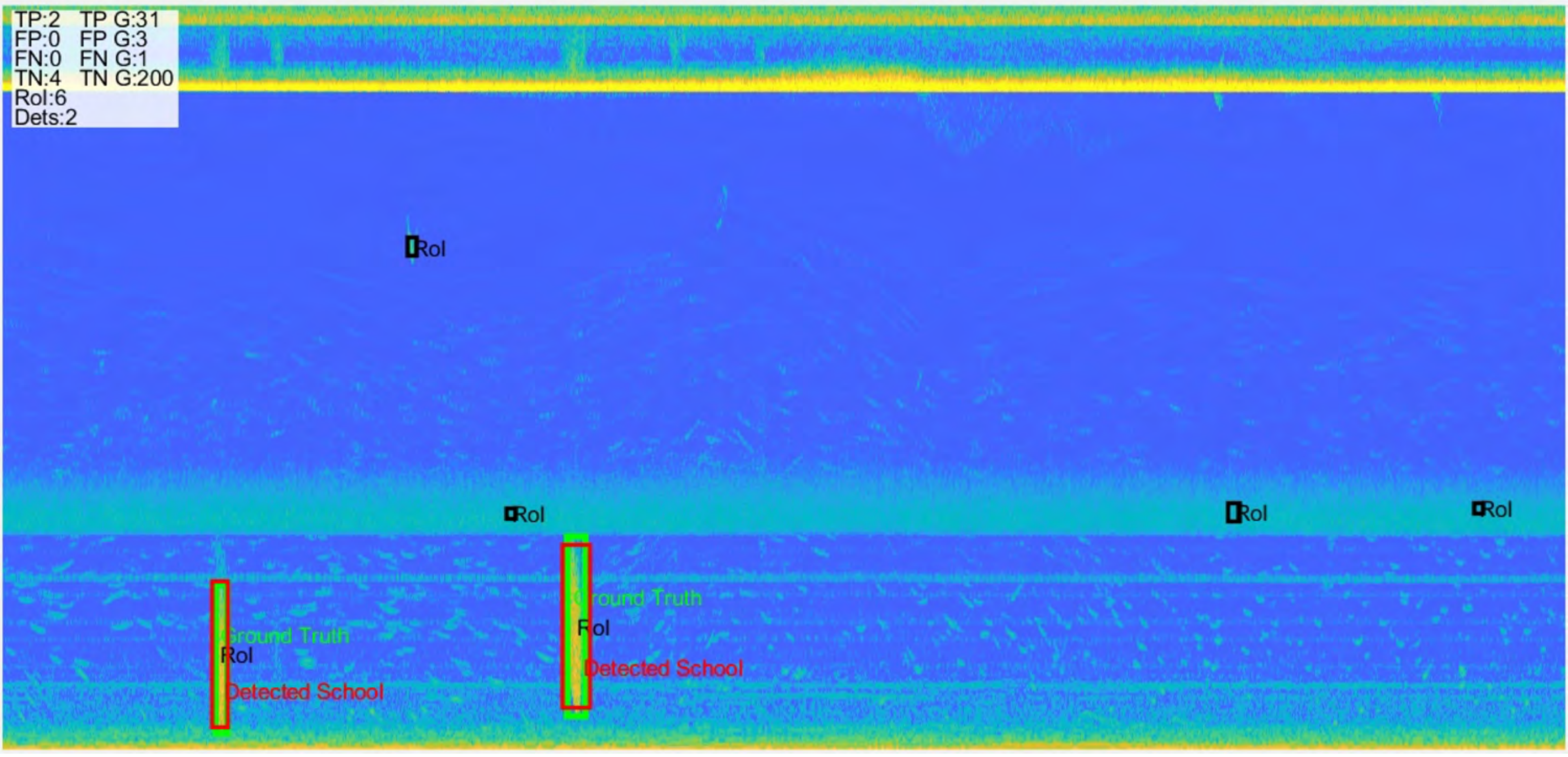
Correct detections



Experimental Results: Qualitative

Deep Learning: ResNet-50 (IoU threshold 0.4)

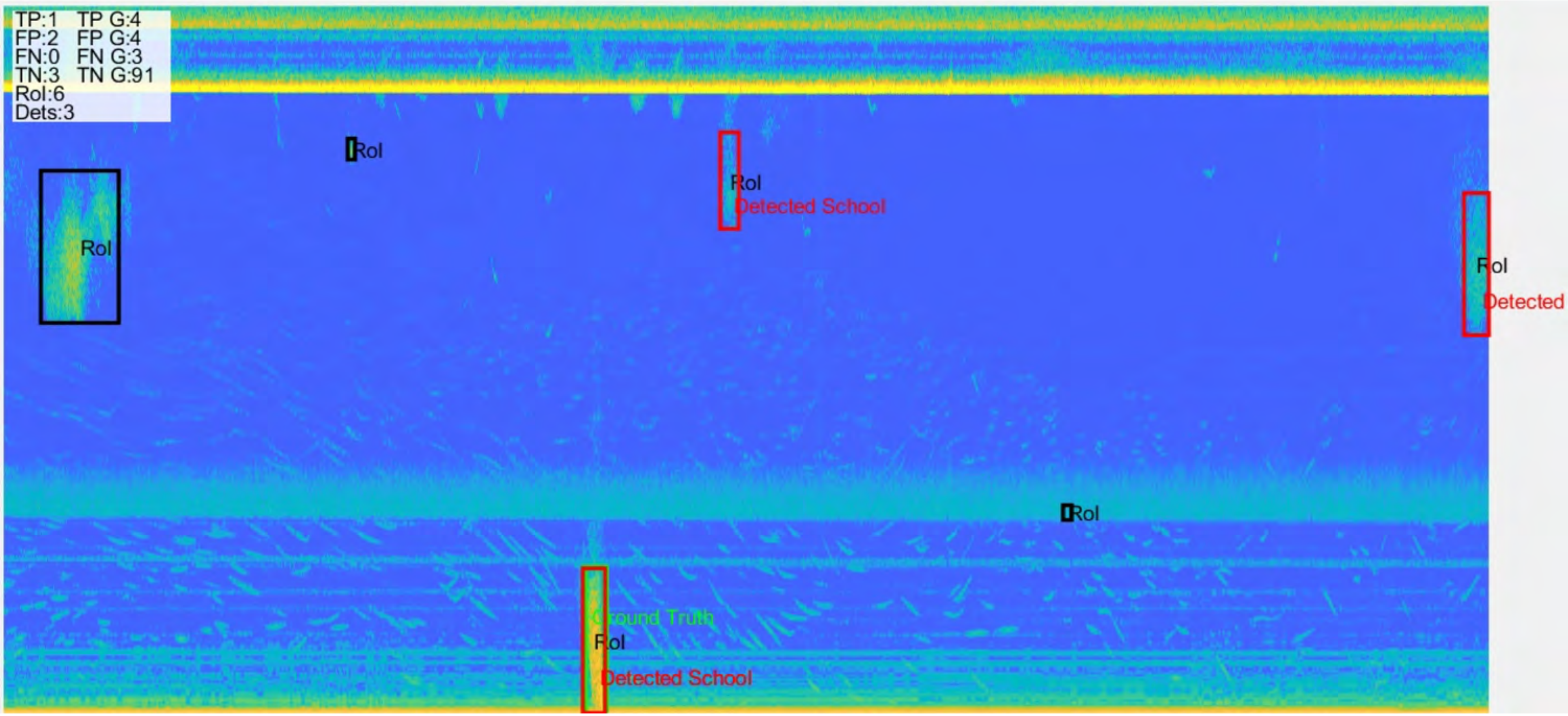
Correct detections



Experimental Results: Qualitative

SVM (IoU threshold 0.4)

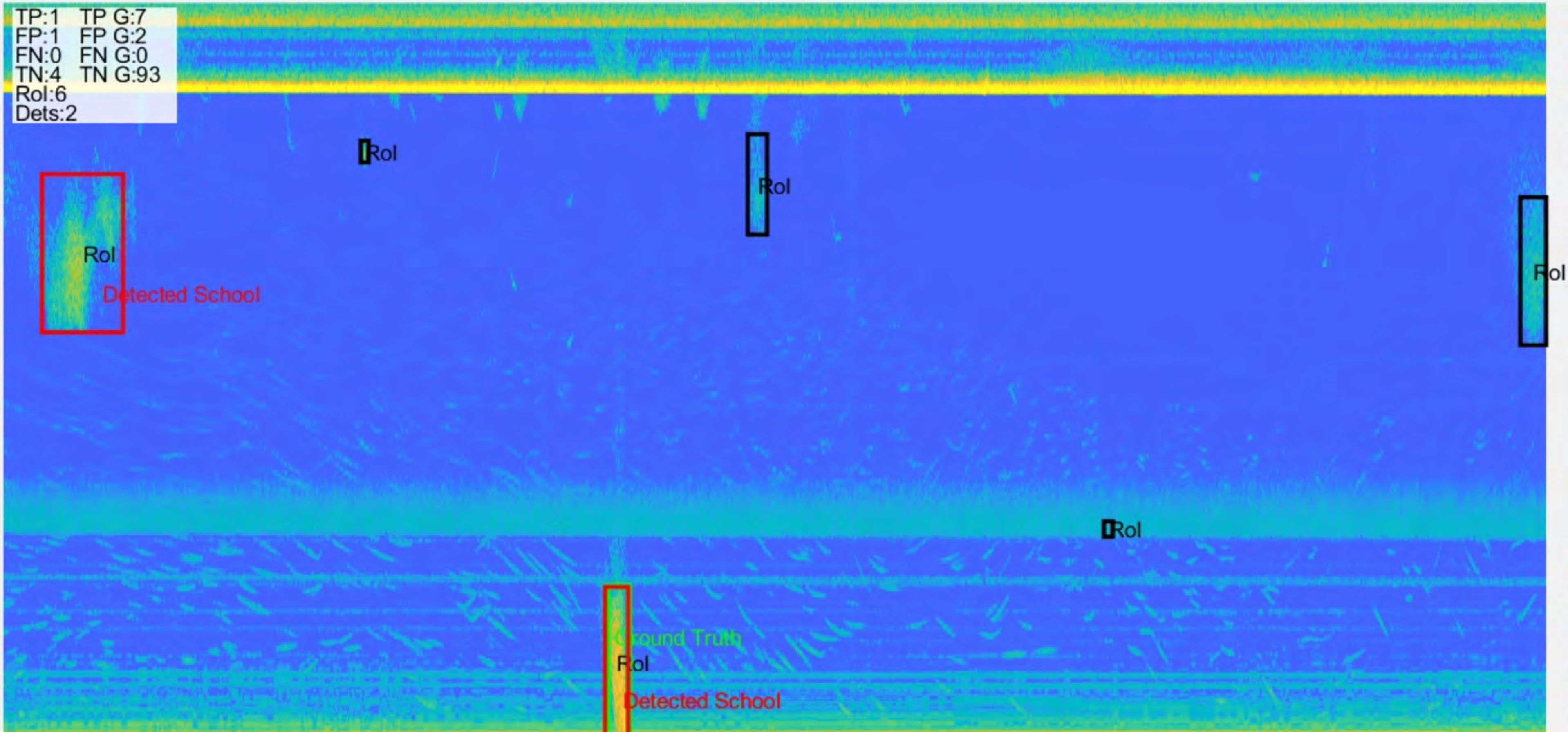
False detections (FP)



Experimental Results: Qualitative

Deep Learning: ResNet-50 (IoU threshold 0.4)

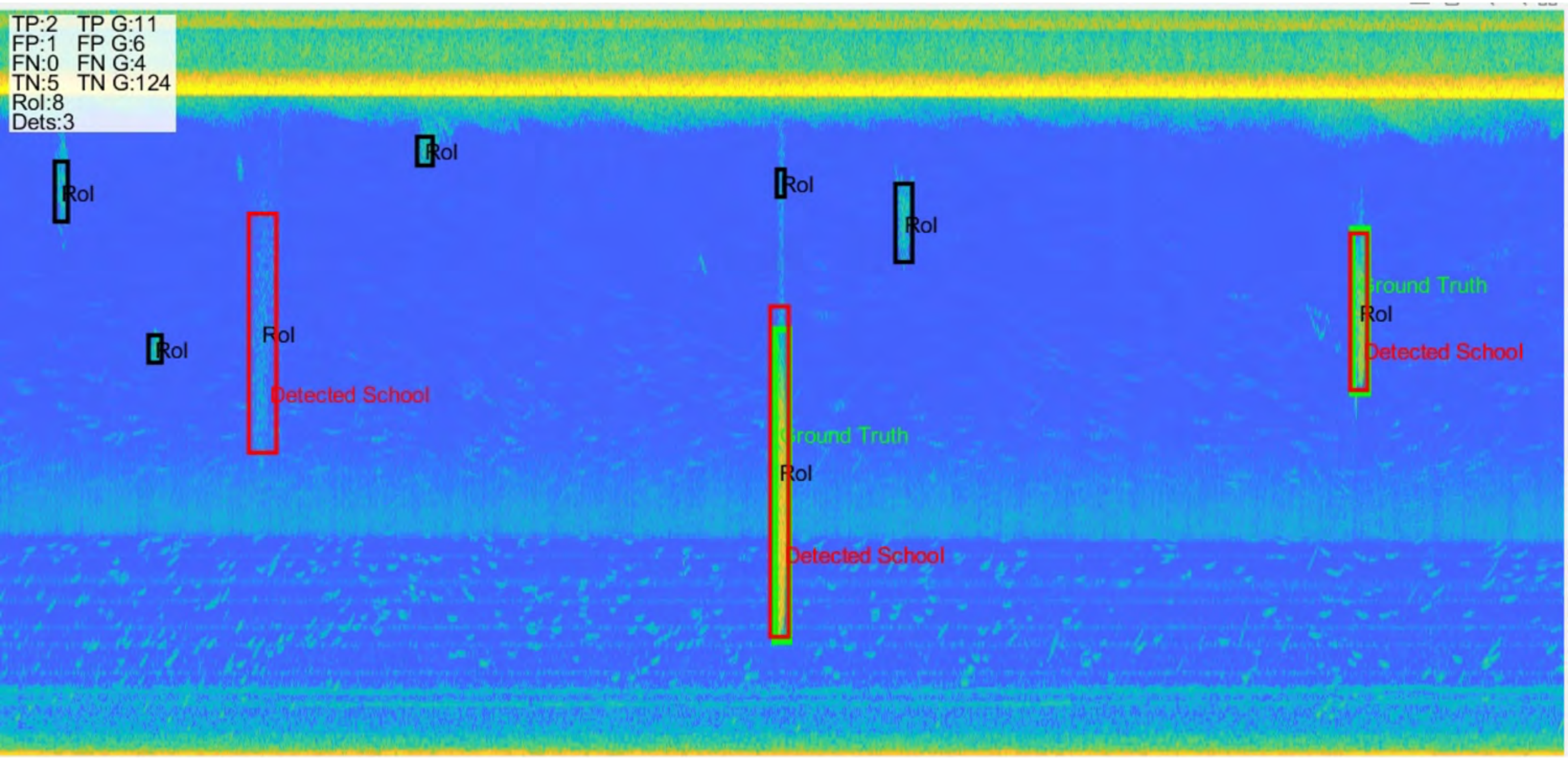
False detection are now TN
A new FP



Experimental Results: Qualitative

SVM (IoU threshold 0.4)

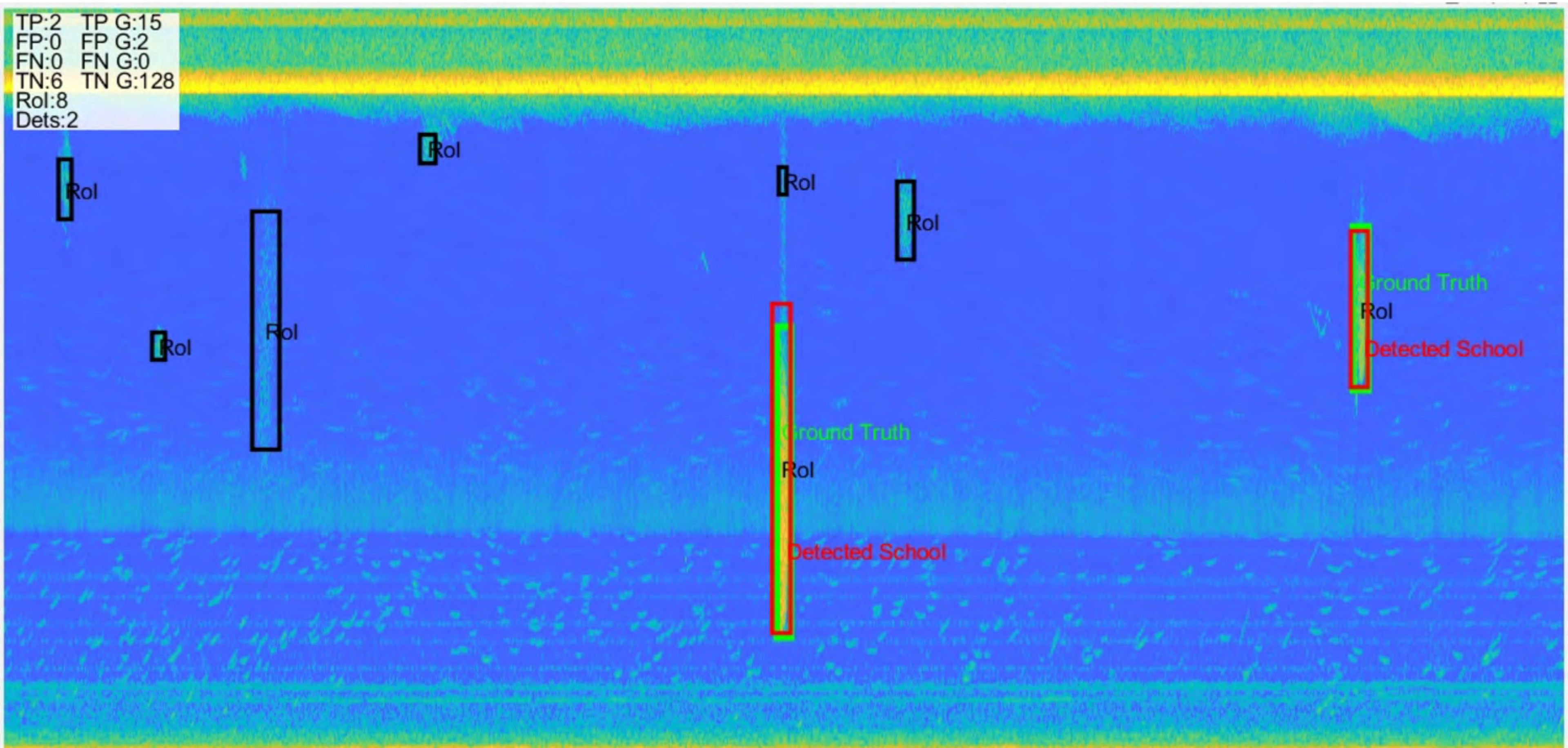
False detections (FP)



Experimental Results: Qualitative

Deep Learning: ResNet-50 (IoU threshold 0.4)

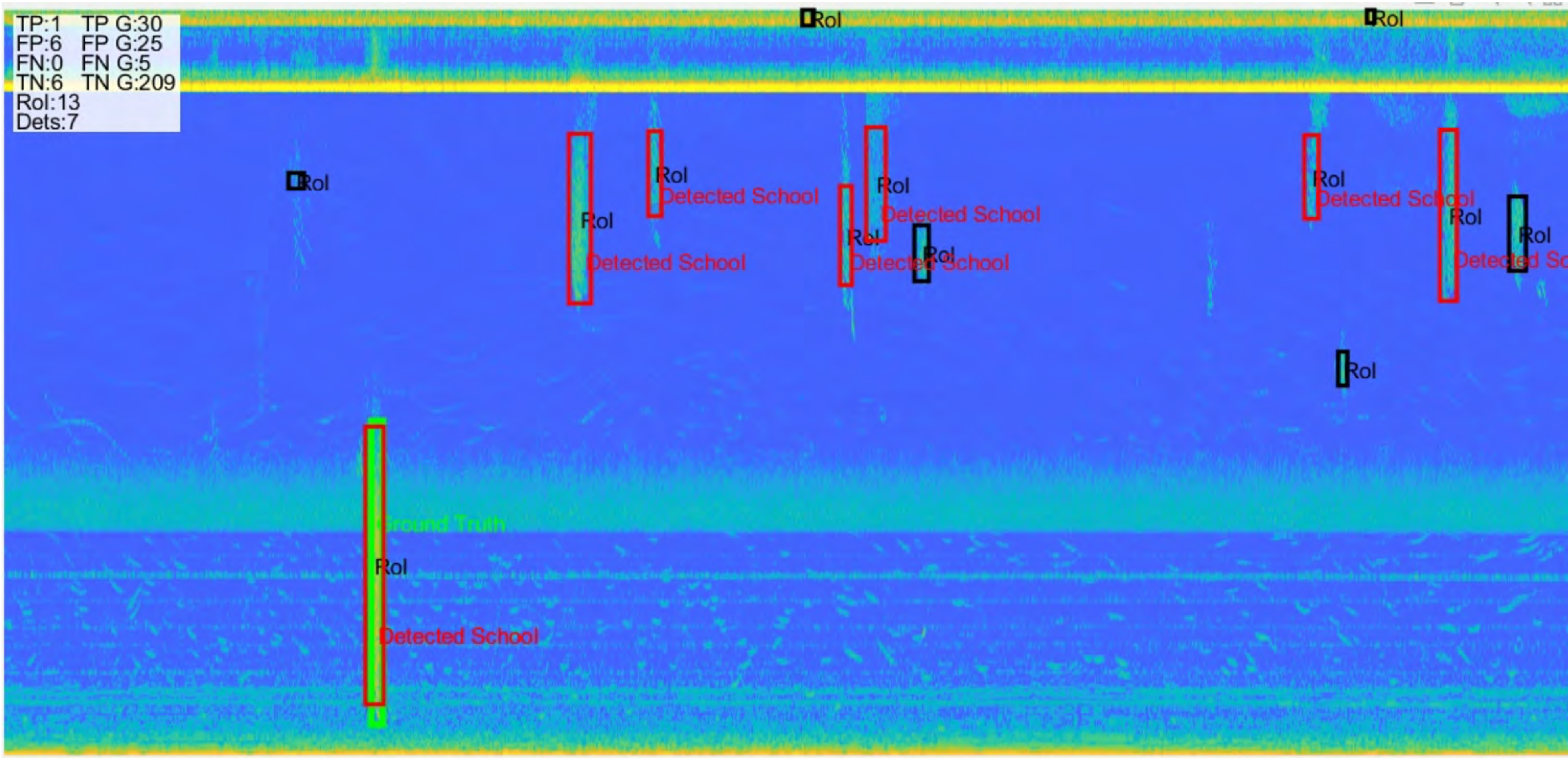
False detection are now TN



Experimental Results: Qualitative

SVM (IoU threshold 0.4)

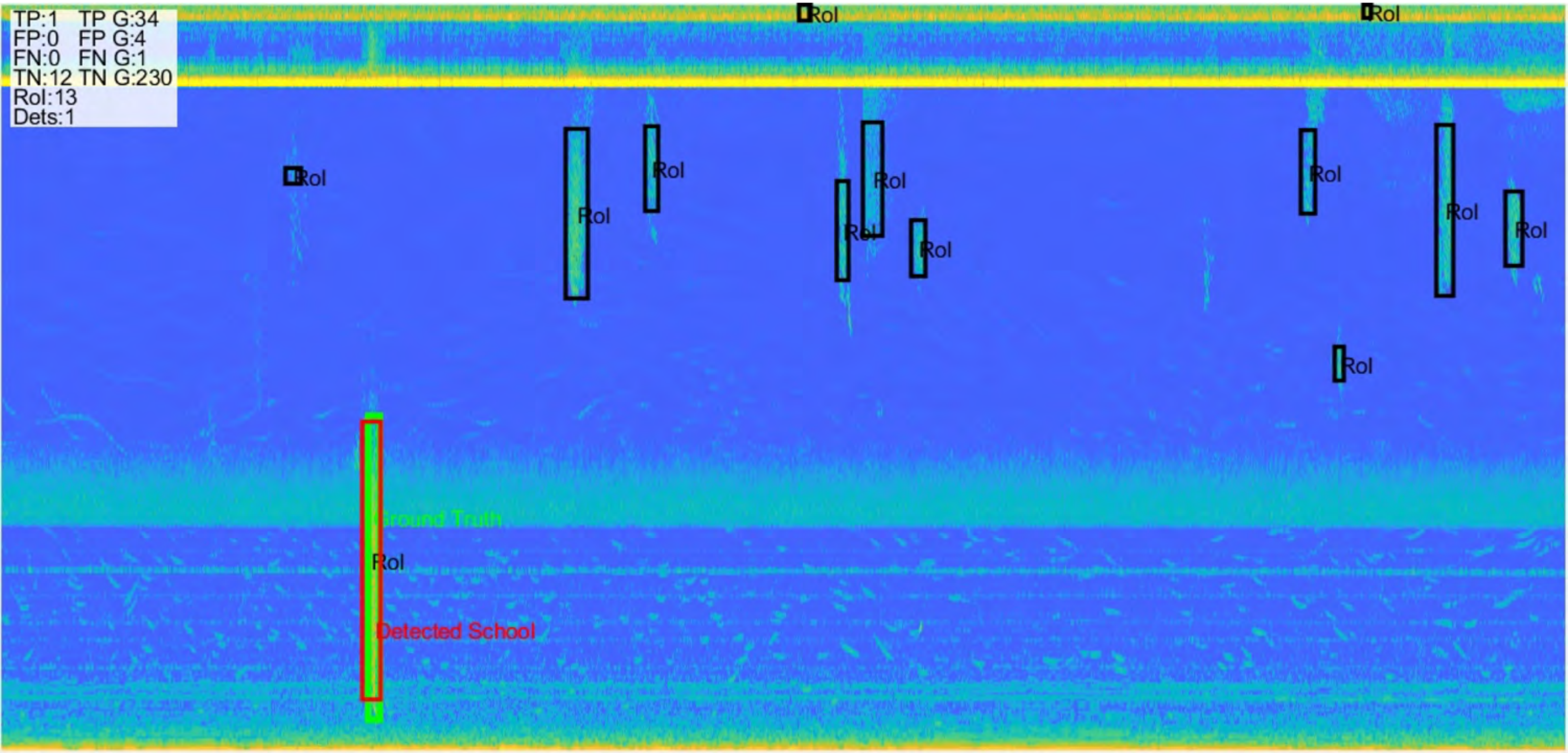
False detections (FP)



Experimental Results: Qualitative

Deep Learning: ResNet-50 (IoU threshold 0.4)

False detection are now TN

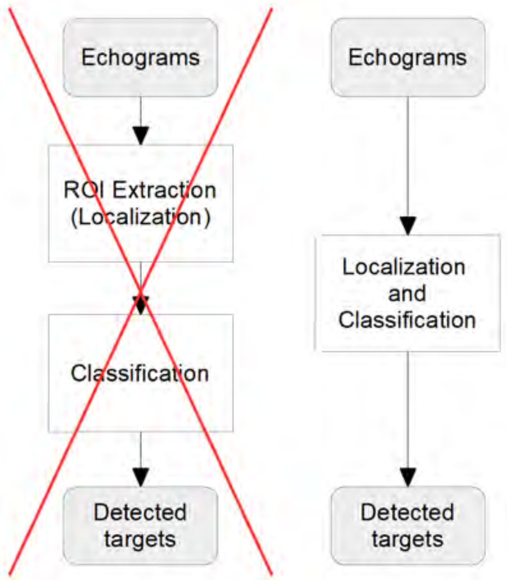


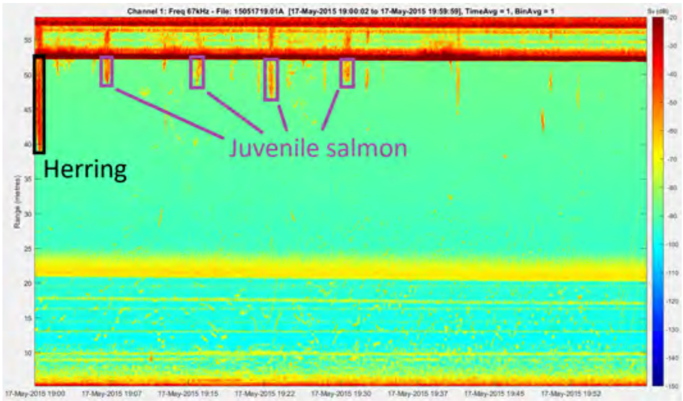
Conclusion

- We explored machine learning approaches for the automatic detection of schools of herring from echograms created from AZFP data
- We proposed and compared two different methods to classify regions of interests:
  - hand-crafted features + support vector machine
  - features automatically extracted and classified by CNNs
- Both methods yielded good results, but CNNs performed best (F1-score: 0.82), even though the dataset was small
- Limitation: performance of ROI extraction, as classifiers can only classify extracted ROIs

Current/Future Work

- Our collaborative project continues!
  - Single deep learning detection pipeline:
    - Single network to perform *localization* and *classification*
    - More scalable approach
  - Extension to other species, structures, and phenomena that can be monitored with echosounders:
    - Current: salmon, zooplankton
    - Future: suspended sediments, ocean turbulence, etc.





References and Acknowledgment

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273-297.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (pp. 770-778), IEEE.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (pp. 4700-4708), IEEE.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (pp. 1-9), IEEE.

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). Inception-v4, Inception-ResNet and the impact of residual connections on learning. In *Thirty-First AAAI Conference on Artificial Intelligence (AAAI)*, (pp. 4278-4284).

NSERC
CRSNG