Multi-scale Attention-based Multiple Instance Learning for Breast Cancer Diagnosis

MICCAI 2025 Daejeon REPUBLIC OF KOREA

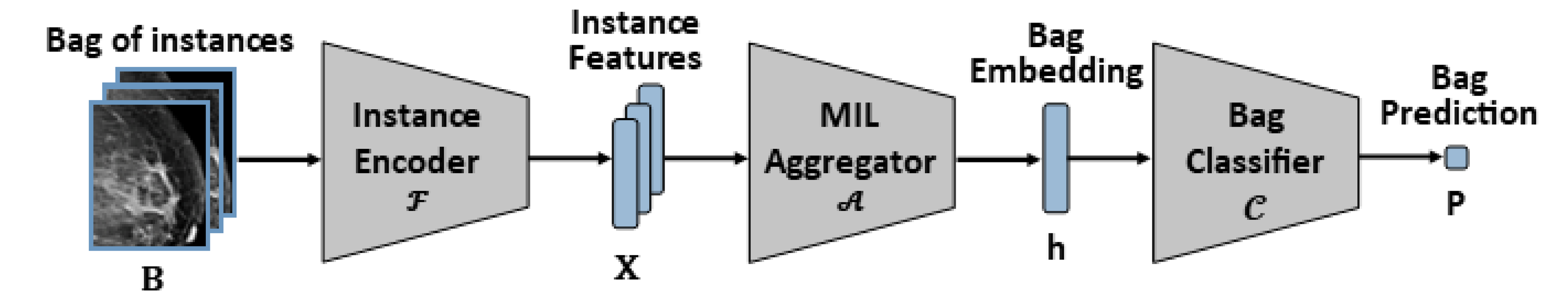
Institute for Systems and Robotics | LISBOA

TÉCNICO LISBOA

Mariana Mourão, Jacinto C. Nascimento, Carlos Santiago, Margarida Silveira
Institute for Systems and Robotics, LARSyS, Instituto Superior Técnico, Portugal

Introduction

- **Mammography** is the gold standard for early breast cancer detection.
- **Multiple Instance Learning (MIL)** enables weakly supervised learning from image-level labels, supporting image-level classification and instance-level (e.g., patches) detection on high-resolution images.



Challenge: MIL models in mammography overlook **instance interactions & multi-scale lesions**, limiting robustness across lesion types and sizes.

Contributions

Proposed a **Feature Pyramid Network (FPN)-based MIL** framework for weakly supervised classification and detection of breast lesions.

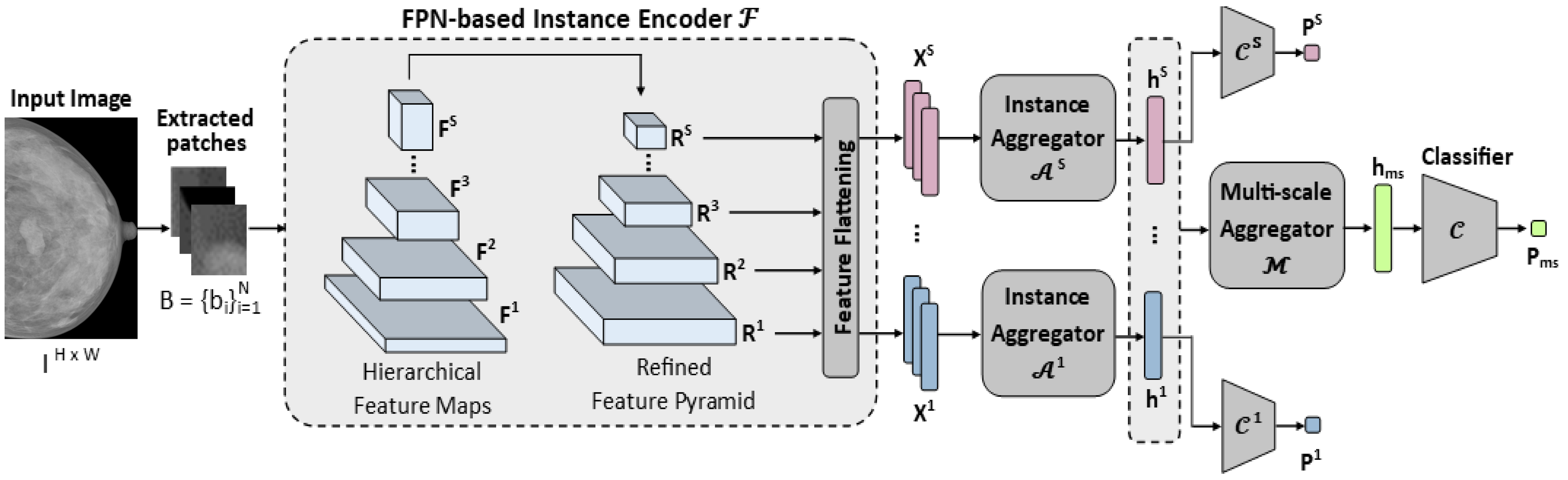
- **Multi-scale analysis** across receptive-field granularities.
- **Flexible & interpretable aggregation** using localized or context-aware attention.
- **Adaptive scale fusion** to handle lesion size variability.
- **Thorough evaluation** against baselines & SoTA models + ablation studies across different lesion types and sizes.

Code & Paper

Scan Me!

Methods

The proposed **FPN-MIL framework** has three main modules.



1. FPN-based Instance Encoder

Extracts fine-to-coarse pixel-level instances X^s at different scales s from a refined feature pyramid.

2. Instance Aggregators

At each scale, aggregate the instance features X^s into an image embedding h^s, also producing interpretable scale-specific heatmaps.

Attention-based MIL (AbMIL)	Set Transformer (SetTrans)
Localized attention	Efficient context-aware attention

3. Multi-scale Aggregator

Adaptively combines the scales, producing:

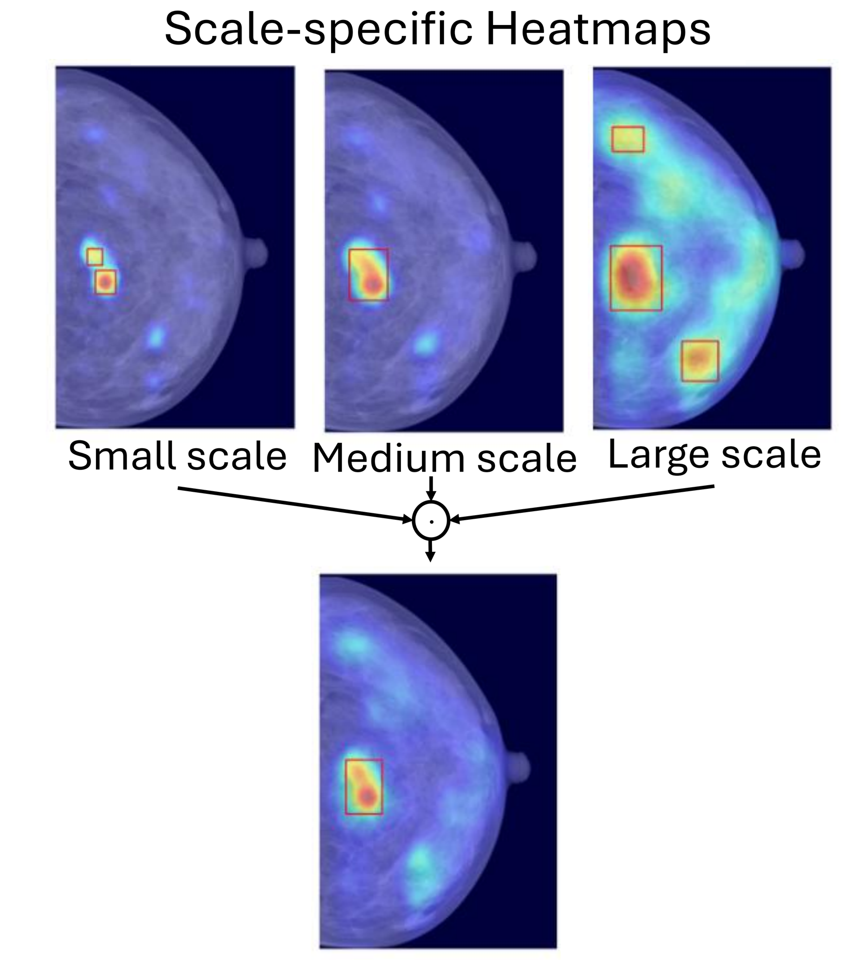
- multi-scale image embedding h_{ms} for image classification
- multi-scale aggregated heatmap for lesion detection

Model Outputs

- **Image Classification**

Prediction P_{ms} of lesion presence

- **Lesion Detection**



Experimental Results

Experiments were performed on the **Vindr-Mammo** dataset, assessing performance on **image classification (AUC)** and **lesion detection (mAP).**

→ **Comparison with baselines & SoTA models**

- **FPN-MIL outperforms Single-scale Patch-based (SSP)-MIL baselines**
- **Our best models are lesion-dependent**

  FPN-SetTrans better for calcifications (distributed point-like lesions)

  FPN-AbMIL better for masses (isolated volume-like lesions)
- **Outperform or competitive against FSC, FSOD & WSOD SoTA models**

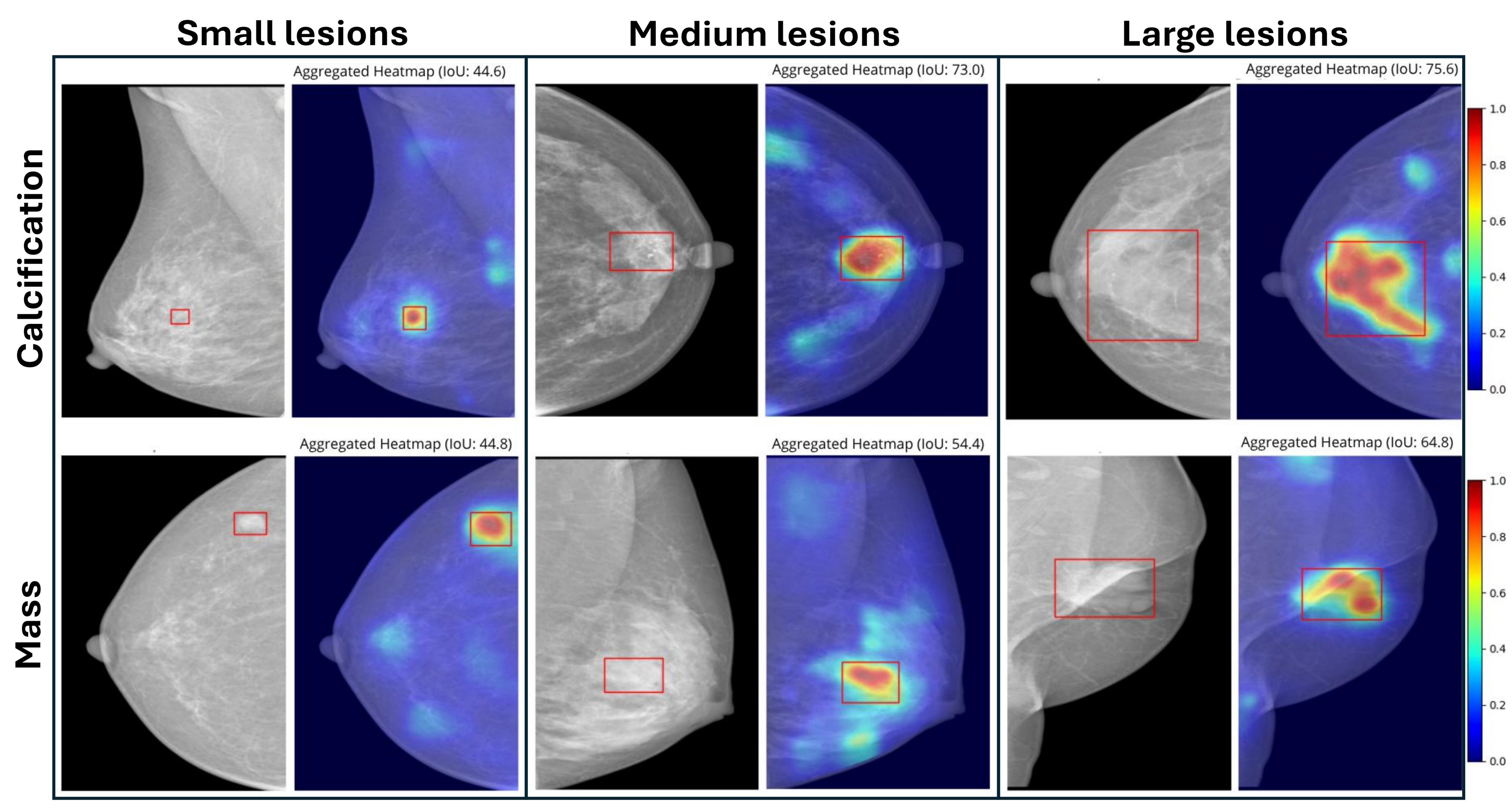
Type	Model	Calcification					Mass				
		AUC	mAP	mAP_s	mAP_m	mAP_l	AUC	mAP	mAP_s	mAP_m	mAP_l
FSC	EN-B2	92.0	-	-	-	-	73.0	-	-	-	-
FSOD	RetinaNet	-	17.0	-	-	-	-	37.0	-	-	-
WSOD	Mammo-FActOR	-	20.0	-	-	-	-	**38.0**	-	-	-
SSP-MIL	AbMIL	90.5	15.9	0.0	26.6	52.1	75.8	14.7	0.0	18.8	61.0
SSP-MIL	SetTrans	88.9	18.4	0.1	29.4	57.6	73.2	5.8	0.0	9.1	22.0
FPN-MIL	(Our) FPN-AbMIL	93.5	32.0	9.1	34.8	57.5	**79.2**	28.2	**4.7**	**32.1**	66.2
FPN-MIL	(Our) FPN-SetTrans	**94.2**	**37.4**	**18.8**	**39.5**	**62.2**	77.4	24.3	3.0	28.0	**73.2**

→ **Ablation Studies**

✓ **FPN > Multi-Scale Patches (MSP)** given its finer receptive-field granularity.

✓ **Multi-scale attention aggregator** improves robustness to lesion size variability.

Inst-Enc	MS-Aggr	Calcification					Mass				
		AUC	mAP	mAP_s	mAP_m	mAP_l	AUC	mAP	mAP_s	mAP_m	mAP_l
MSP	Attention	91.3↓	18.5↓	0.3↓	22.8↓	54.9↓	77.1↓	9.5↓	0.0↓	9.5↓	46.6↓
FPN	w/o	93.8↓	33.0↓	8.5↓	35.7↓	61.6↓	78.8↓	25.2↓	5.0↑	30.7↓	56.0↓
FPN	Concat	92.2↓	28.8↓	12.6↓	17.2↓	59.4↓	76.9↓	19.4↓	**7.0**↑	**32.6**↑	26.4↓
FPN	Attention	**94.2**	**37.4**	**18.8**	**39.5**	**62.2**	**79.2**	**28.2**	4.7	32.1	**66.2**



Interpretability: Our multi-scale aggregated heatmaps highlight clinically relevant breast lesions across different types and sizes.

Conclusions

- The proposed FPN-MIL framework has a modular and adaptable design, robust across different lesion types and sizes.
- Our best models outperform or are competitive against SoTA in breast lesion classification & detection while more label-efficient.
- **Future work:** Explore more MIL aggregators and jointly analyze multi-view mammograms.

fct Fundação para a Ciência e a Tecnologia

Acknowledgements

This work was supported by FCT projects MIA-BREAST and AI-Radiologist (10.54499/2022.04485.PTDC and 10.54499/2024.07344.IACDC).