UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

AImage Lab
University of Modena and Reggio Emilia

ICIAP 2025 ROME

TesticulUS

# Enhancing Testicular US Classification Through Synthetic Data and Pre-Training

N. Morelli, K. Marchesini, L. Lumetti,
D. Santi, C. Grana, F. Bolelli

# The role of Testicular Ultrasound (TUS)

- Male-related factors account for approximately **30–50%** of infertility cases in couples
- **TUS** is a key, non-invasive tool in the assessment of male infertility and related diseases, often undetected by routine clinical exams (low testicular volume, **inhomogeneity**, microlithiasis, varicocele…)
- Despite its diagnostic potential, AI and computer vision applications for testicular ultrasound remain extremely limited, few studies exist, **datasets are not publicly available**, and **the field is largely unexplored**

# Introduction & Motivations

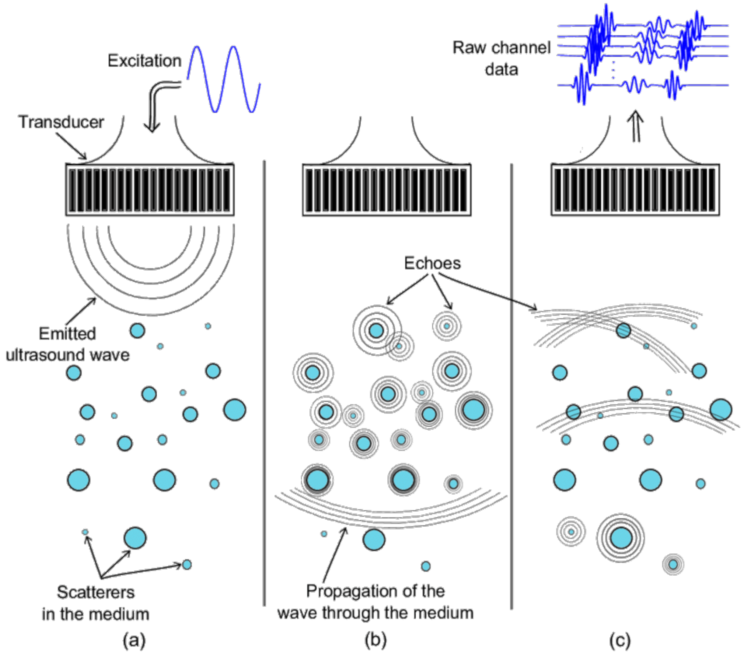
## Challenges

- Noisy labels
- Subjective interpretation
- Lack of public datasets
- Artificial markers

## Key Contibutions

- A heuristic approach to reduce label noise
- Application of diffusion models for synthetic data generation
- Systematic evaluation of pre-training strategies

# How Ultrasound Works




## Modality Attributes

- Single Channel
- Values between [0, 255]
- Affected by noise and artifacts (like speckle)

# Datasets

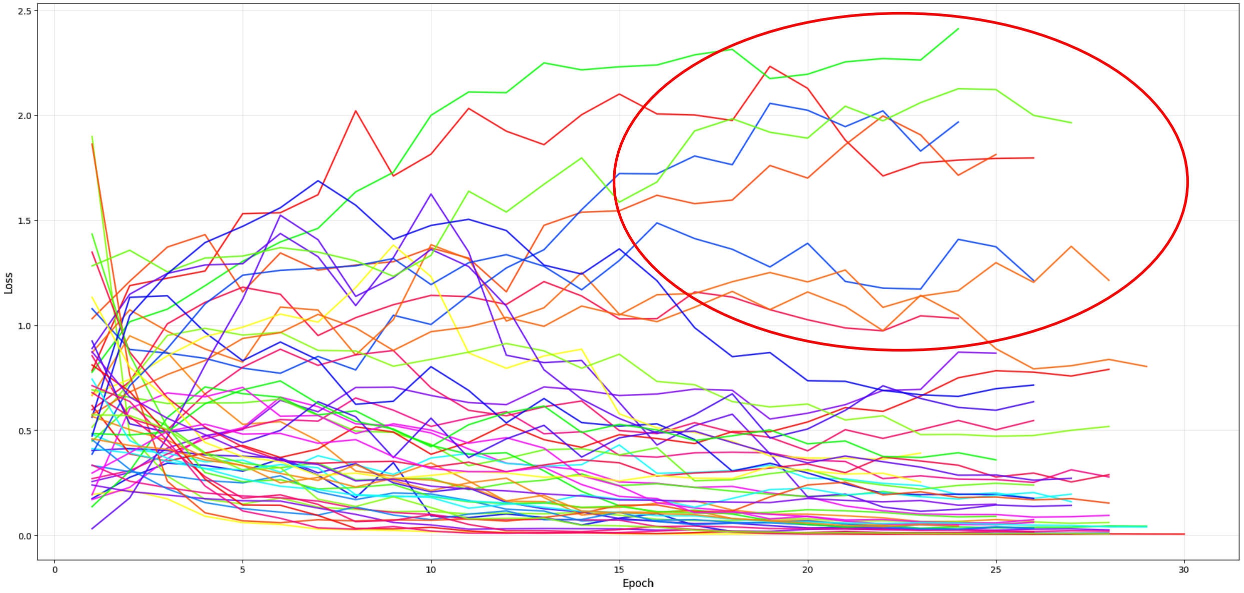
## Labeled Dataset

- **220 Patients**
- **~880 Testicular Images**
  2 images per patient (left/right)
  2 views per image
- **Annotations of tissue homogeneity**
  20% inhomogeneous

## Unlabeled Dataset

- **About 25k ultrasound unlabeled images**
- **Testicular and Thyroids scans**
  1.6k Testis, 23.4k Thyroids
- **No Label**

# Labeled Dataset Problems - Noisy Labels




Cross-entropy loss per sample: most converge, but some persistently high-loss samples (circled) indicate mislabeled or hard cases

# Labeled Dataset Problems - Noisy Labels

## Heuristic Filtering Algorithm

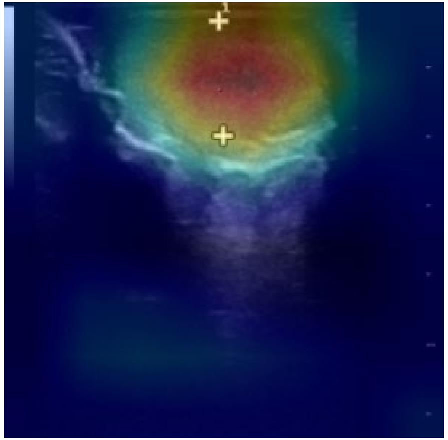
- Samples with loss spikes higher then 1.0 for at least three times are marked as suspicious
- Suspicious samples have been flipped rather then removed

## How did filtering impact Accuracy and F1 Score?

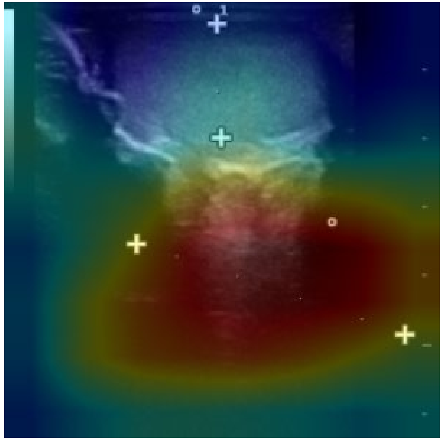
| Dataset | Accuracy (↑) | F1-Score (↑) |
|---|---|---|
| Complete | $81.51 \pm 2.78$ | $55.72 \pm 4.37$ |
| Filtered | $\mathbf{88.15 \pm 1.94}$ | $68.59 \pm 3.30$ |
| Flipped | $86.78 \pm 2.21$ | $\mathbf{73.17 \pm 1.55}$ |

# Labeled Dataset Problems - Markers

**Grad-CAM on Original Image**

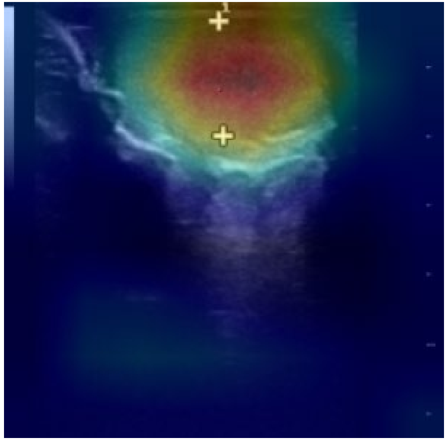


**No Marker Invariance Training**

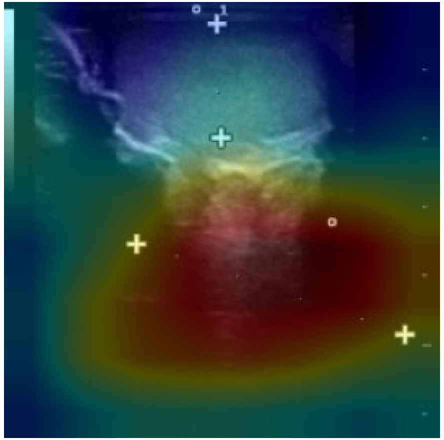


# Labeled Dataset Problems - Markers

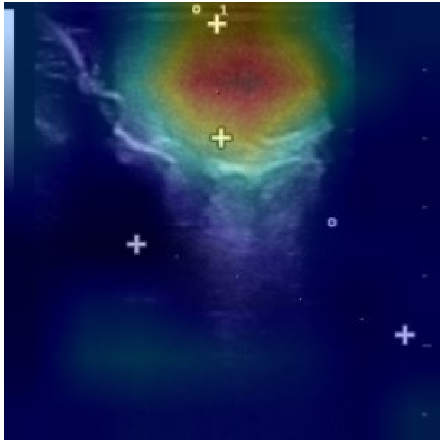
**Grad-CAM on Original Image**



**No Marker Invariance Training**

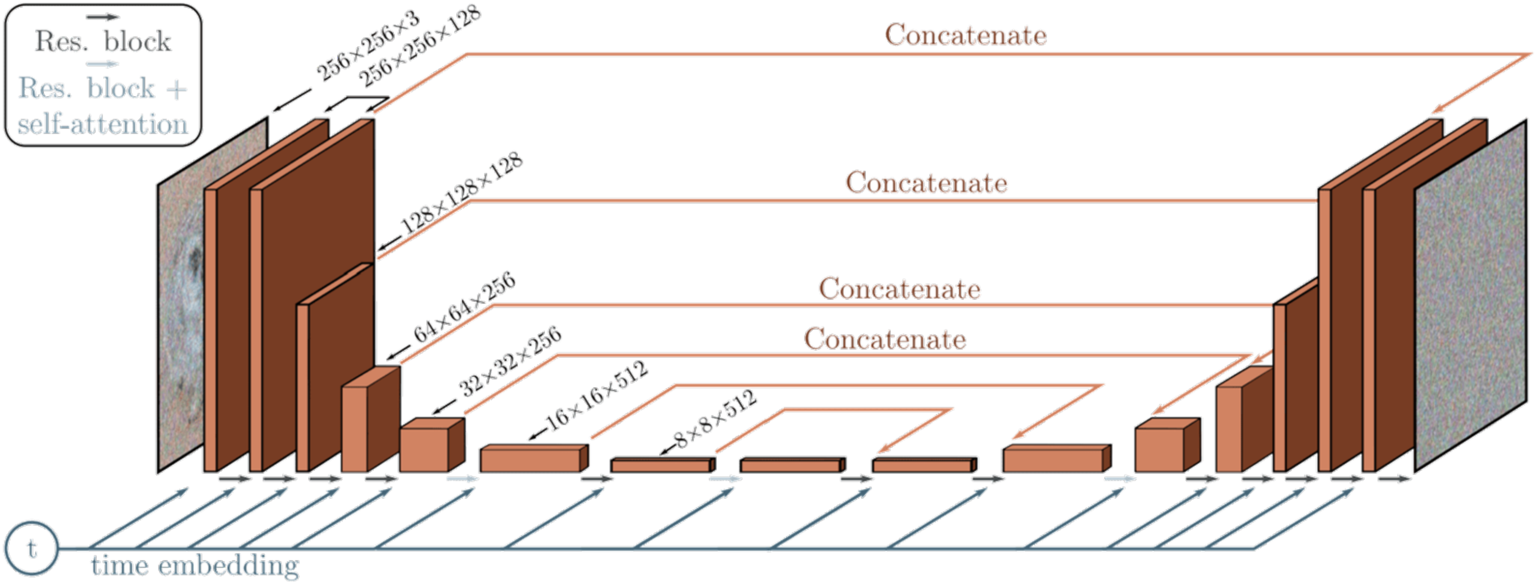


**Trained for Marker-Invariance**



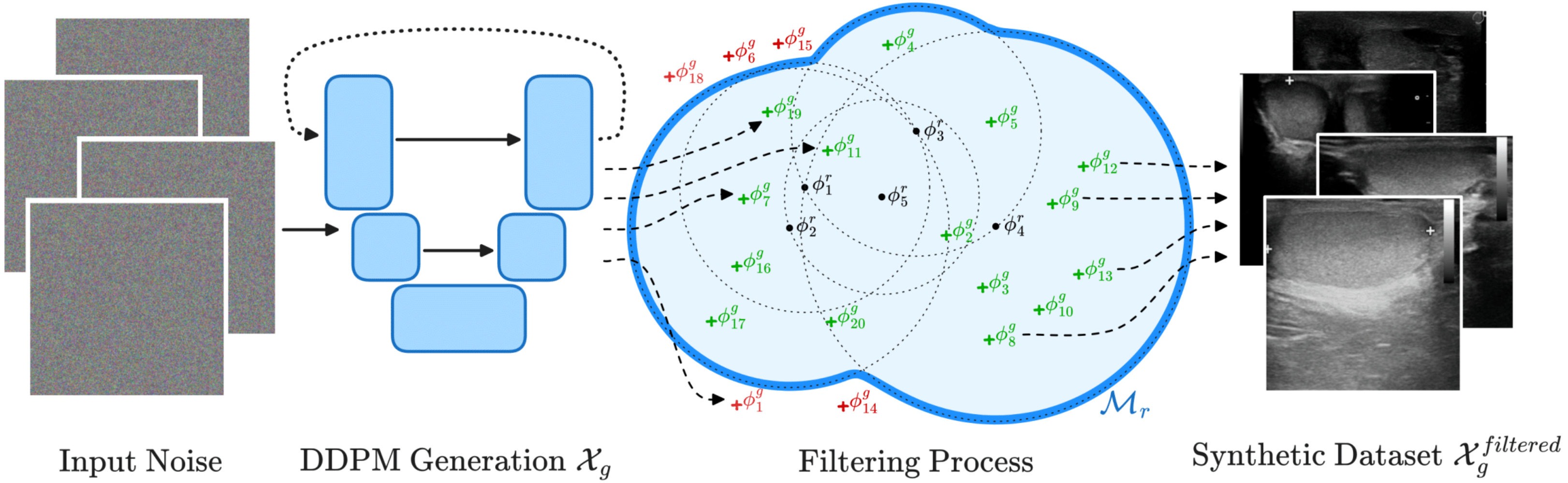
Without augmentation, models confuse artificial markers with real features. Marker-invariant training restores correct focus

# Synthetic Unlabeled Dataset



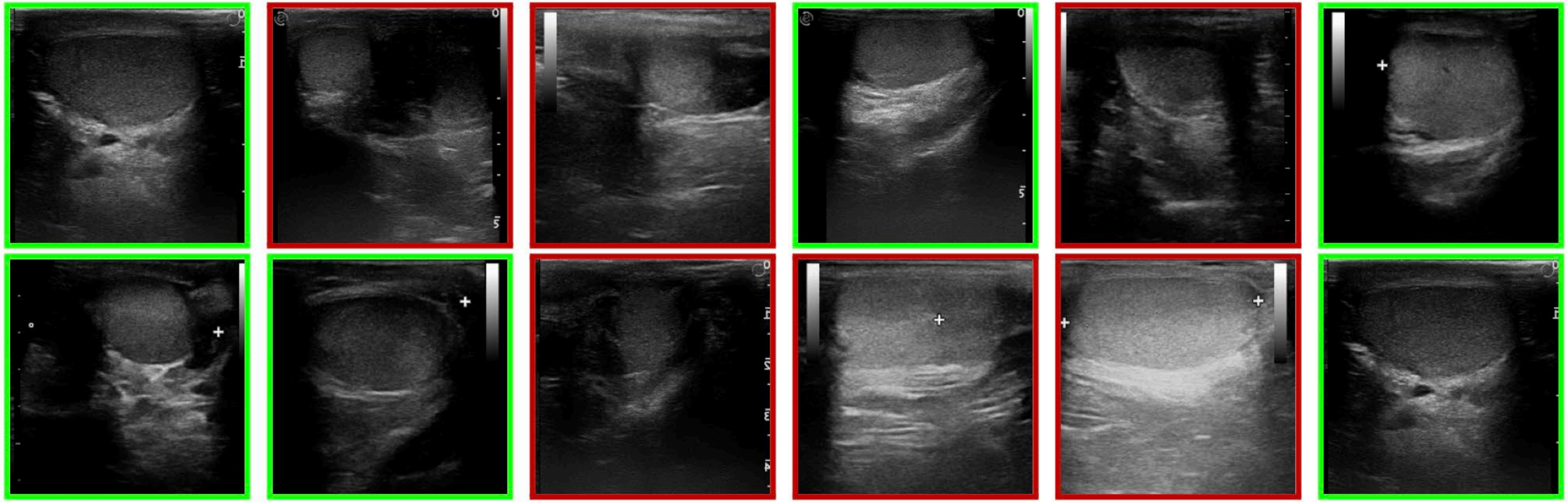

20k images generated through improved DDPM model

# Synthetic Generation & Filtering

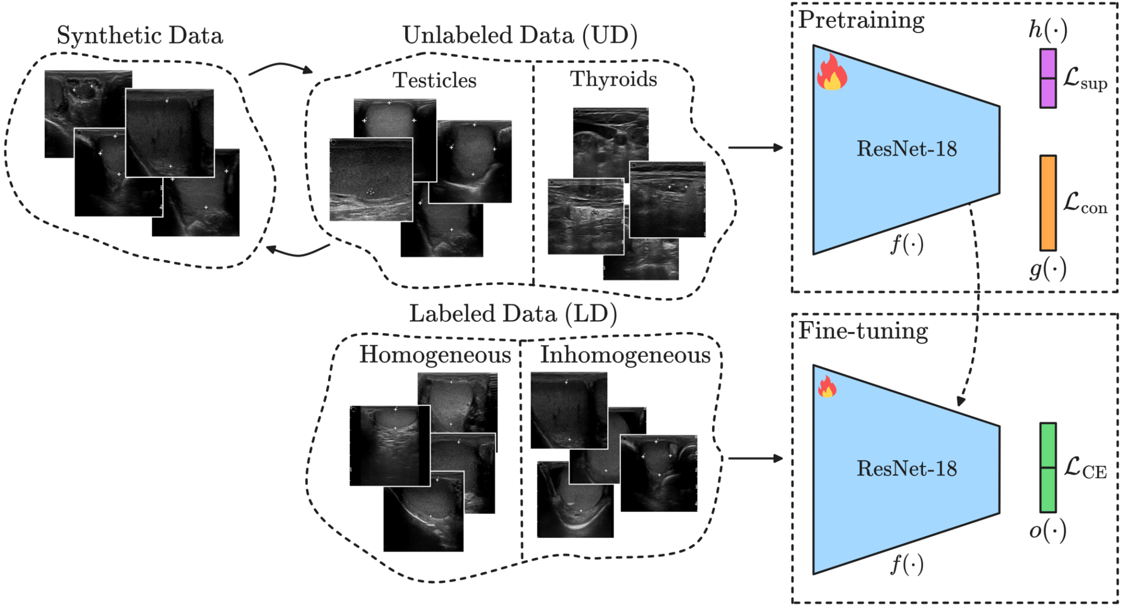



Proposed pipeline for synthetic generation and filtering of unlabeled datasets

# Synthetic Generation & Filtering



# Pretraining Strategies




- Semi-supervised SimCLR pre-training framework
- Pre-training supervised loss for classifying different organs
- Synthetic data proved to be a valid substitute for real data during pre-training

# Results

**Impact of Label Noise Filtering**:

Classification performance boosts the F1-score from 55.72% to 73.17%

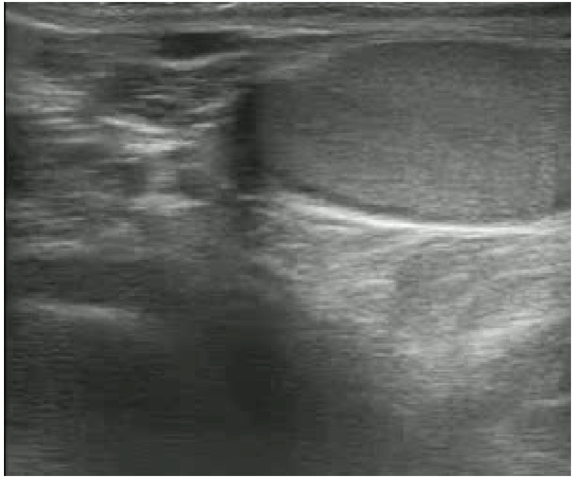
**Utility of (Filtered) Synthetic Data:**

Pretraining on synthetic filtered data achieves classification performance comparable to real data, overcoming privacy concerns

**Effectiveness of Pretraining Strategies:**

Combined supervised and unsupervised pre-training on diverse ultrasound images (testicular and thyroid) delivered the best overall performance by achieving 86.78% Accuracy and 73.17% F1-score in classifying testicular inhomogeneity

# Future Directions



- Incorporate **dynamic ultrasound videos** to offer richer contextual information for classification

- Develop **label-conditioned** synthetic image generation to produce datasets suitable for both pretraining and fine-tuning

Nicola Morelli

Kevin Marchesini

Luca Lumetti

Daniele Santi

Costantino Grana

Federico Bolelli

Thanks for the attention.