TU Graz

SCIENCE
PASSION
TECHNOLOGY

# Skin Image Classification Pipeline: Using SAM2 Encoder Features

Taisiya Parkhomenko (01650051)

Bachelor Thesis

# Outline





Taisiya Parkhomenko (01650051), Institute of Visual Computing (IVC)

# Introduction

- Two-stage classification pipeline using frozen SAM2 encoder features:
  - Stage 1: Skin vs Not Skin Classification
  - Stage 2: Lesion Group and Malignancy Classification
- Large-scale dataset combination:
  - ISIC 2019: 157,610 skin lesion images (augmented from 33,569 original)
  - DTD: 28,200 texture images (augmented from 5,640 original)

# Objectives

- Implement a skin lesion detection pipeline using foundation models
- Explore the applicability of SAM2 for medical imaging
- Reduce labeling costs by automating classification steps
- Combine datasets from different domains for improved robustness
- Evaluate transfer learning capabilities for skin lesion detection

# Pipeline Overview

- Two-stage classification using SAM2 encoder features:
  - Stage 1: Skin vs Not-Skin (MLP1)
  - Stage 2: Lesion Group (MLP2) and Malignancy (MLP3)
- Feature extraction:
  - SAM2 ViT encoder (frozen)
  - 256-dimensional embeddings
- Linear probing to leverage pre-trained features

# Datasets

- ISIC 2019 Dataset:
  - 157,610 augmented images (33,569 original)
  - 5 lesion groups:
    - melanocytic: 106,095 images (21,219 original)
    - non-melanocytic carcinoma: 25,455 images (5,091 original)
    - keratosis: 22,625 images (4,525 original)
    - vascular: 1,785 images (357 original)
    - fibrous: 1,650 images (330 original)
  - Malignancy labels: benign (96,705), malignant (60,905)

# Data Processing

- Unified preprocessing pipeline:
  - Resize all images to 1024x1024 (SAM format)
  - 5x augmentation per image:
    - Original
    - Horizontal flip
    - Rotations (90°, 180°, 270°)
- Dataset splits (total 196,045 images):
  - Train: 137,122 images (117,490 skin / 19,632 non-skin)

# Feature Extraction and Linear Probing

- SAM2 Feature Extractor:
  - Frozen ViT encoder
  - 256-dimensional embeddings
  - GPU-accelerated batch processing
- Linear Probing Architecture:
  - Input: 256-dim SAM features
  - Hidden layers: [512, 256] with ReLU & Dropout(0.3)
  - Output: task-specific classes

# Segmentation Results

- Qualitative mask visualization
- SAM2 performance on different skin conditions
- Mask selection effectiveness

Add side-by-side comparison of original images and masks

# Classification Performance

- Skin vs Not-Skin (MLP1): 99.38% validation accuracy
  - Balanced dataset: 2,000 samples per class
  - FocalLoss with class weights [1.0, 5.93]
- Lesion Type (MLP2): 65.8% validation accuracy
  - Max 2,000 samples per majority class
  - F1 scores: 0.59-0.72 across 5 classes
  - FocalLoss with class weights [1.5, 1.5, 1.5, 1.2, 1.2]

# Ablation Study Results

- Class imbalance handling (MLP3):
  - Balanced sampling vs. original distribution
  - Balanced (4k samples): 72.55% accuracy, F1 score 0.7255
  - Original (22k samples): 71.29% accuracy, F1 score 0.7045
- Training time comparison:
  - Balanced model: 1.5 hours
  - Original distribution: 4.5 hours (3x slower)

# Connection to Objectives

- Foundation model applicability:
  - SAM2 features effective for medical imaging tasks
  - Linear probing demonstrates strong transfer learning capabilities
- Multi-domain dataset integration:
  - Successfully combined DTD textures with ISIC skin lesions
  - Hierarchical pipeline provides modular classification

# Key Findings & Future Work

- Key Findings:
  - SAM2 encoder features highly effective for skin lesion classification
  - Linear probing with balanced sampling outperforms larger imbalanced datasets
  - Multi-stage pipeline effective for hierarchical classification
- Future Work:
  - Investigate prompt point generation using grayscale extrema

# Thank You

## Thank you for your attention!

- Supervisor: Dr. Marc Masana Castrillo
- Institute of Visual Computing (IVC)
- Technical University of Graz