TU Graz

SCIENCE
PASSION
TECHNOLOGY

Skin Image Classification Pipeline: Using SAM2 Encoder Features

Taisiya Parkhomenko (01650051)

Bachelor Thesis

Outline

Problem Definition & Motivation

- Skin lesion classification is key for early melanoma detection.
- Datasets are imbalanced: some classes have 20k+, others <500.
- OOD risk: flowers, textures, artifacts misclassified as lesions.
- Goal: Build a robust multi-stage pipeline that handles imbalance and rejects OOD.

Add dataset histogram (class imbalance)

Add motivating example: lesion vs flower misclassification

Pipeline Design

- Frozen SAM encoder extracts 256-dim features.
- Three downstream MLPs:
  - MLP1: Skin vs Not-Skin
  - MLP2: 8 Lesion classes
  - MLP3: Benign vs Malignant
- Compared Parallel vs Multihead setups.
  - Parallel = independent MLPs

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

Data Preparation & Balancing

- Exp1–5: Original ISIC splits (imbalanced).
- Exp6: Smart balancing strategy:
  - Cap NV/UNKNOWN at 5k, BCC/MEL at 3k.
  - Augment rare classes to 1k.
- Augmentations: flips, rotations, jitter.

Table: per-class counts before vs after balancing

Loss Functions & Training Strategies

- Weighted Cross-Entropy: Baseline.
- Focal Loss: From literature, but unstable.
- LMF (LDAM + Focal Mix): Improves minority class margins.
- Balanced MixUp: Soft-label augmentation.

Equation slide: CE, Focal, LMF side by side

Bar chart: lesion macro-F1 by loss type

Results & Key Insights

- Parallel > Multihead for lesion classification.
- Best lesion: Parallel + LMF (Exp1) $\sim$65% acc, 0.60 macro-F1.
- Skin detection near-perfect ($\sim$99%).
- Benign/Malignant stable ($\sim$75–79% acc).

Table: Parallel vs Multihead results

Confusion matrix for lesion task highlighting minority errors

OOD Detection with ODIN

- Manual test: flower misclassified as melanoma.
- Implemented ODIN (temperature scaling + input perturbation).
- Evaluated with Places365 dataset as OOD.
- Results: High AUROC/AUPR $\Rightarrow$ better safety.

ROC/AUPR curves comparing ODIN vs baseline

Diagram: ODIN workflow on top of MLP1

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

Extra Experiment: Cropped Lesion Masks

- From S2R2: Segment to Recognize (Janoušková et al.).
- Motivation: Remove background noise (skin tone, artifacts, rulers).
- Plan: Train with SAM-cropped lesion patches.
- Status: Not done yet $\rightarrow$ future extension.

Image: original vs cropped lesion mask overlay

Best Model (So Far)

- Parallel + LMF (Exp1) is best so far:
  - Skin: 0.999 / 0.999
  - Lesion: 0.645 / 0.599
  - Benign/Malignant: 0.795 / 0.790
- Strong balance across tasks compared to others.

Table: summary of best model metrics

Future Directions

- Use MedSAM / UniSeg for medical segmentation features.
- Test cropped vs full images.
- Ensemble models (Parallel + Multihead).
- Try higher resolutions (512 vs 1024).
- Semi-supervised learning with unlabeled dermoscopy.
- Stronger OOD rejection (energy-based, Mahalanobis).

Bullet slide: future research directions with icons

Thank You

Thank you for your attention!

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