GS-TransUNet: Integrated 2D Gaussian Splatting and Transformer UNet for Accurate Skin Lesion Analysis

Anand Kumar, Kavinder Roghit Kanthen and Josna John
University of California, San Diego

INTRODUCTION

Early skin cancer detection is crucial, with melanoma accounting for 75% of skin cancer deaths. We present GS-TransUNet, a novel framework that unifies classification and segmentation tasks by integrating 2D Gaussian splatting [1] with Transformer-Unet [2] architecture. With a dual task consistency loss, our approach demonstrates **improved accuracy** over existing methods while maintaining computational efficiency.

A unified framework for skin classification and segmentation improves performance for both.

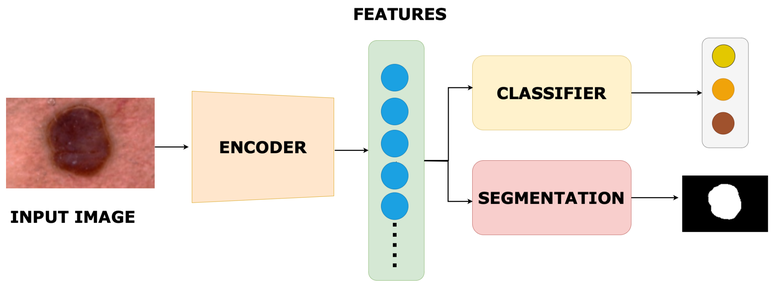

Fig. 1 An overview of our approach to unify classification and segmentation by using common features where the encoder performs a dual task.

DATASET

Our experiments utilize two comprehensive dermoscopic image datasets: ISIC-2017 [3] with 2,750 images and PH2 [4] with 200 images. Both datasets include segmentation masks and classification labels for skin lesions, for melanoma, seborrheic keratosis and common nevus.

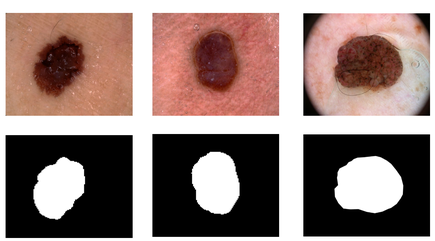

Fig. 2 Three dermoscopic samples from our dataset with the input image at the top row and their corresponding segmentation masks at the bottom. The classes are melanoma, keratosis and common nevus from left to right.

NETWORK ARCHITECTURE

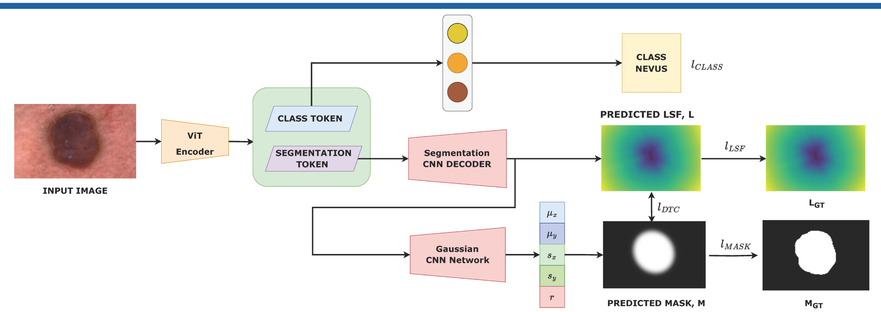

Fig. 3 Architecture of GS - TransUNet model, combining Vision Transformer (ViT) with UNet for segmentation and classification. The input image is passed through a pre-trained ViT encoder to obtain the class (global) and segment (local) tokens and are used for simultaneous classification and segmentation. The classification is done using an MLP, and the segmentation mask is predicted using two parallel networks: (i) CNN Decoder to obtain level set function and (ii) CNN Gaussian Network to obtain binary mask.

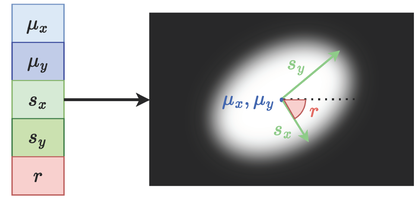

Fig. 4 Illustrative diagram for generating binary masks using 2D Gaussian Splatting. The Gaussian Splats are generated using 6 features: μ_x, μ_y , s_x, s_y , r

RESULTS

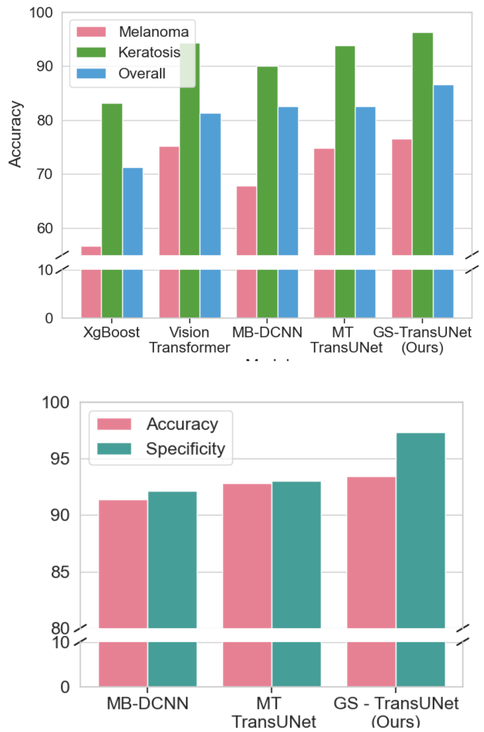

Fig 5. Graphs showing performance of models in classification (top) and segmentation (bottom) tasks.

Fig 6. Visualization of the segmentation masks. The pictures are input image, ground truth mask, MT-TransUNet mask and GS-TransUNet mask from left to right.

Gaussian Splats exploit the spherical shape of lesions

CONCLUSION

1. **We introduce a novel approach that generates segmentation masks through two parallel networks using Gaussian splatting and signed distance fields. This dual-path design ensures that the generated masks are consistent and robust to noise.**

2. **GS - TransUNet achieved a 2.5% improvement in accuracy, setting new benchmarks in classification and segmentation tasks**

REFERENCES

1. 3D Gaussian Splatting for Real-Time Radiance Field Rendering
2. TransUNet: Rethinking the U-Net architecture design for medical image …
3. Skin Lesion Analysis toward Melanoma Detection
4. PH² - A dermoscopic image database for research and benchmarking,