A New Approach to Enhance Bruise Analysis with Limited Labelled Data

Qiming Wang, Paul L. Rosin, Xianfang Sun

School of Computer Science and Informatics, Cardiff University, UK

Motivations

Bruising is a key form of forensic evidence in domestic abuse cases. However, current analysis methods are subjective and rely on expert judgement. Due to the lack of annotated bruise datasets, supervised learning is impractical. This project explores alternative approaches for pixel-level bruise segmentation without relying on labelled data.

Challenges

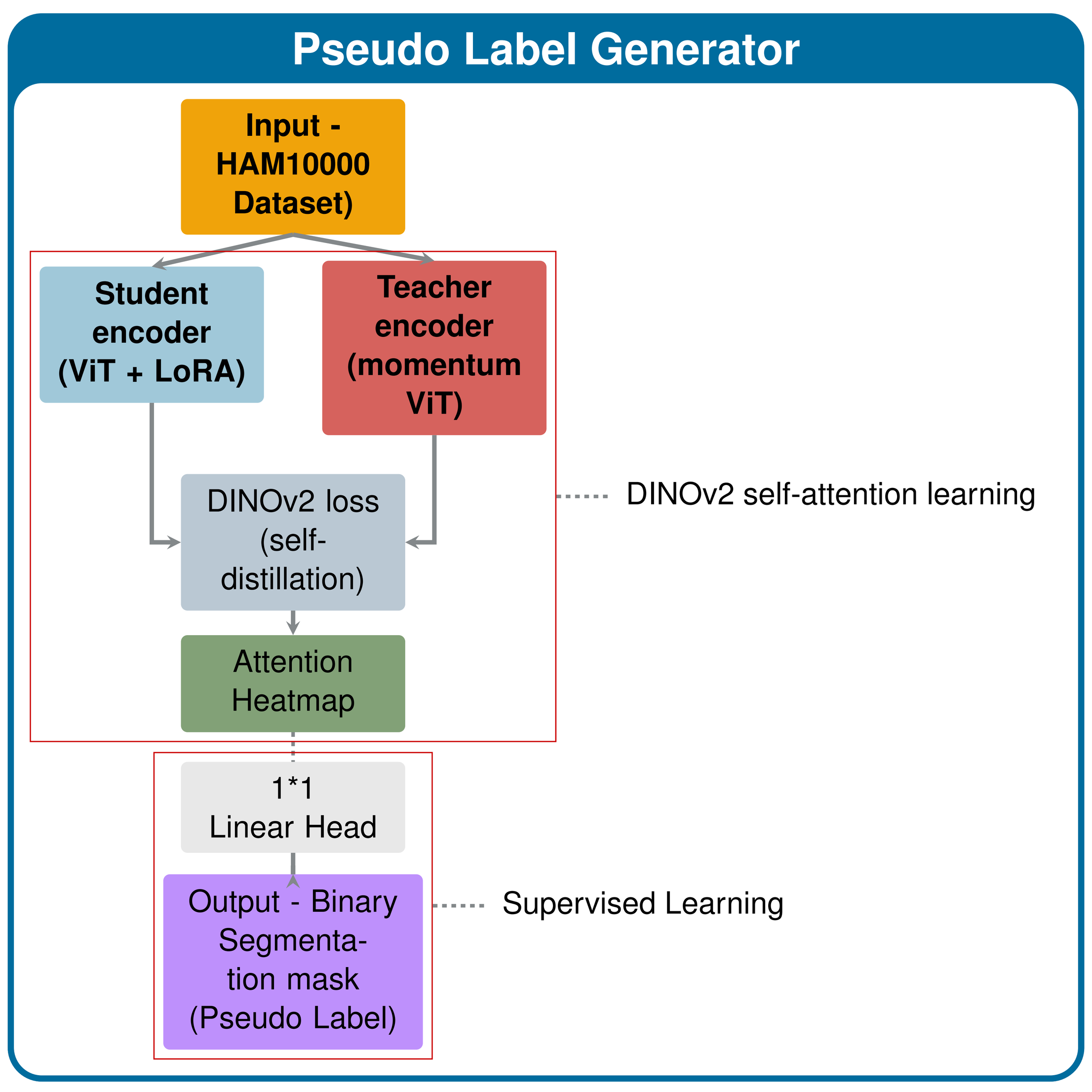
- Limited labelled datasets, even when including online sources, e.g. Roboflow open database, their annotation cannot be trusted since they are labelled by unkown sources
- The only reliable annotated dataset was originally labelled using two nested circles by forensics experts:
  - **Inner Circle:** confidently area with bruise only
  - **Outer Circle:** uncertain region mixed with background
  - **Outside Outer Circle:** pure background

  Since it only contains 54 images, it is only applied during evaluation. Hence, a new method for evaluation is required
- Identifying bruise boundaries from a single image is challenging for humans[1], limiting re-annotation of existing datasets.

Methods

- Using a Vision Transformer model (DINOv2) to generate pseudo labels
- Applying generated pseudo labels to supervised models, e.g. Florence-2
- A new evaluation method is designed and tailored to the double-circled annotations

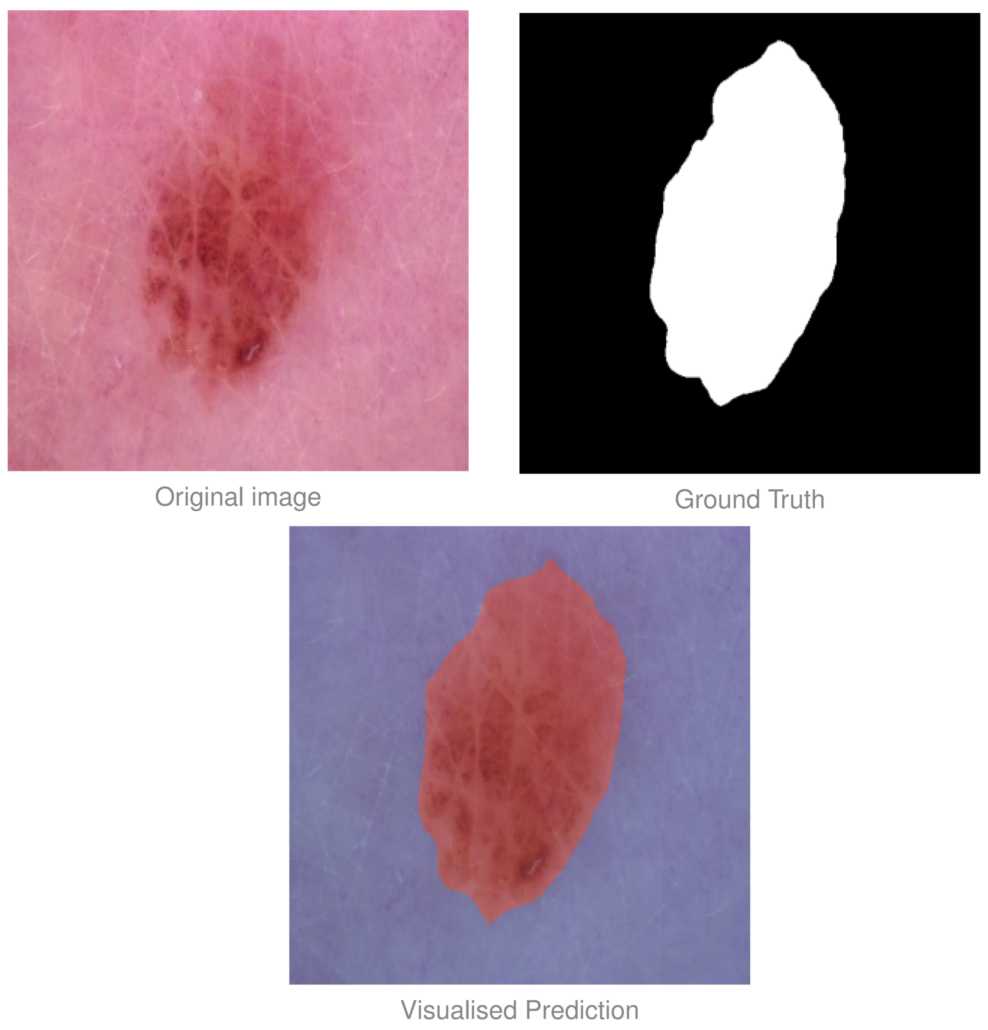
Pseudo Label Generator



Pseudo-label generation pipeline using DINOv2 with LoRA. Unlabelled skin lesion images from HAM10000 are processed by DINO to output an attention heatmap. A lightweight 1×1 linear head is trained on these features to output binary segmentation masks

Since DINO(v2) was trained on general dataset, lead to a difficulty when apply it directly for medical purposes. To use DINO as a pseudo-label generator for bruising, it needs to be trained on a dataset with similar characteristics. In this case, HAM10000 with 10015 images of skin lesions is used.

To test the result, 15% of the HAM10000 data was reserved for evaluation:



DINOv2-LoRA-based segmentation result on a HAM10000 image. Left: Original image; Middle: Ground-truth pixel-level segmentation mask; Right: Visualised prediction highlighting the lesion boundary.

The current DINOv2 with LoRA reaches the average-level accuracy compared to other state-of-the-art models:

Model	IoU (%)
DINOv2 + LoRA (Ours)	90.6
MFSNet [2]	90.6
ViT + SAM [3]	96.01
SkinSAM [4]	78.43
U-Net [5]	88.28

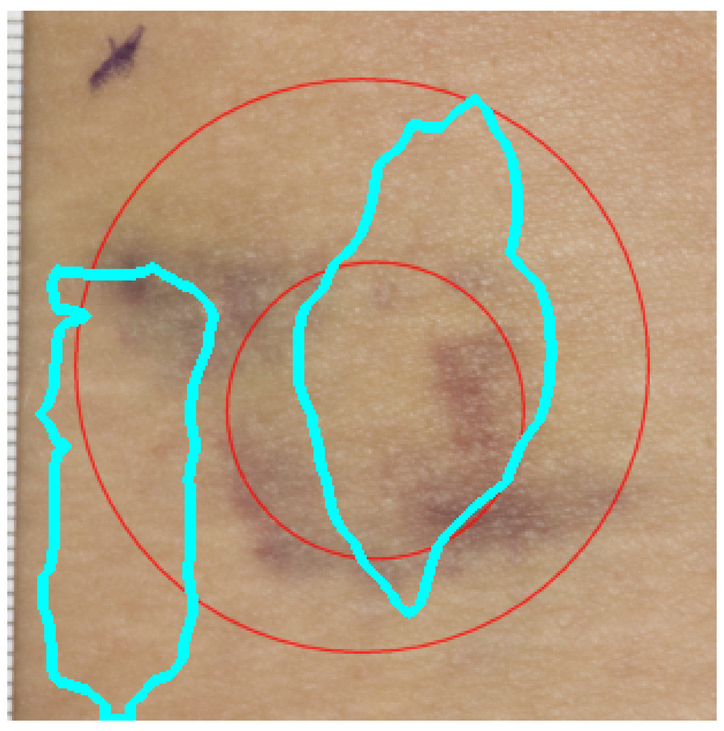
Our model (DINOv2 + LoRA + One Linear Layer) test on 15% HAM10000 images with the IoU result on the right side, compare other models on the ISIC 2018 competition

It is worth noting that there was another model that used a similar structure (vision transformer + segmentation head), which achieved over 96% IoU, but the segmentation head is a large Segment Anything Model (632M parameters).

Evaluation

To use the non-standard annotated dataset introduced earlier, we divide the annotation into three sections, all positives are inside of the inner circle, all negatives are outside of the outer circle. Since the area between two circles is mixed with bruises and background, it is excluded from evaluation. This ensures more reliable measurement of true positive (TP), false positive (FP), true negative (TN), and false negative (FN):

- **True Positives (TP):** Area covered by both the segmentation mask and the inner circle.
- **False Positives (FP):** Area covered by the segmentation mask but outside of the outer circle.



Example image with the double-circle annotation (red lines) with the prediction (cyan line). The area covered by both the prediction and the inner circle is true positive. All outside of the inner circle and inside of the outer circle is ignored, and the prediction outside of the outer circle is false positive, all else outside of the circle is true negatives

- **True Negatives (TN):** Area outside of both segmentation mask and outer circle.
- **False Negatives (FN):** Area inside of the inner circle but not covered by segmentation mask.

	Mean (%)	Min (%)	Max (%)
TP	94.59	10.73	100.00
FP	6.40	0.00	38.45
TN	5.41	0.00	89.27
FN	93.60	60.55	100.00
F1-Score	87.49	19.37	100.00

Mean, minimum, and maximum values for True Positives (TP), False Positives (FP), True Negatives (TN), False Negatives (FN), and the overall F1-Score. Higher TP, TN, and F1-Score values indicate better performance, while lower FP and FN values are preferable.

Conclusion

This study demonstrated that a Vision Transformer model, specifically DINOv2 and LoRA adaptation with a lightweight segmentation head, can effectively generate pseudo labels even with extremely limited unlabelled data for further training.

By applying the tailored evaluation method for the double-circle ground-truth annotation, a promising result was achieved for training the pseudo-label generator based on a proxy dataset with visual similarity to bruises. However, further validation will be necessary to confirm the effectiveness of this evaluation.

Further Work

- Convert the current double-circle to double-bounding box, which is a more usual method of annotation, to compare the performance
- Test and train the generator on various more complex and lower-quality public data
- Use the current pseudo-label generator to produce labels for the internet data and train on different supervised models, and evaluate the result.

To improve the pseudo-label generator, we also propose that:

- Apply larger zero-shot segmentation head like UNet, or test with SAM for its zero-shot characteristic.

References

[1] Sophie E Grossman et al. "Can we assess the age of bruises? An attempt to develop an objective technique". en. In: *Med. Sci. Law* 51.3 (July 2011), pp. 170–176.

[2] Hritam Basak, Rohit Kundu, and Ram Sarkar. "MFSNet: A multi Focus Segmentation Network for skin lesion segmentation". In: *arXiv [eess.IV]* (Mar. 2022).

[3] Galib Muhammad Shahriar Himel et al. "Skin cancer segmentation and classification using Vision Transformer for automatic analysis in dermatoscopy-based non-invasive digital system". In: *arXiv [eess.IV]* (Jan. 2024).

[4] Mingzhe Hu, Yuheng Li, and Xiaofeng Yang. "SkinSAM: Empowering skin cancer segmentation with Segment Anything Model". In: *arXiv [cs.CV]* (Apr. 2023).

[5] Su Myat Thwin and Hyun-Seok Park. "Enhanced skin lesion segmentation and classification through ensemble models". en. In: *Eng* 5.4 (Oct. 2024), pp. 2805–2820.

Contact: {WangQ79, RosinPL, SunX2}@cardiff.ac.uk