Automatic Identification of Regions of Interest in Dermoscopy Images Using Vision Transformers and Weakly Supervised Learning

D. Araújo, C. Barata, A. Bissoto, C. Santiago
ISR / IST / LARSyS

Abstract

- Skin cancer is a growing public health concern. Early detection of the lesion plays a crucial role in ensuring successful treatment.
- In this paper, we propose a novel approach that combines **Vision Transformers** and **Multiple Instance Learning (MIL)**.
- Our method consists of two branches: 1) the **Vision Transformer** branch; and 2) the **deep-instance MIL** branch.
- This combination enables accurate image and patch classification, which facilitates ROI identification.

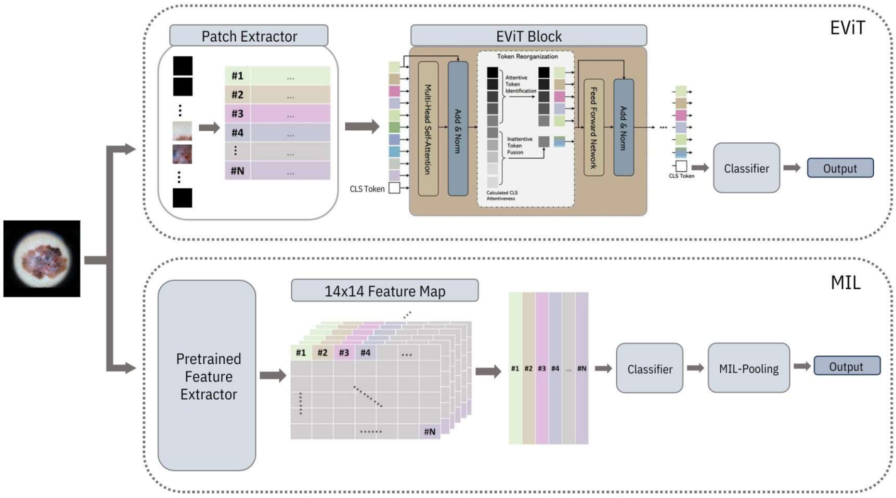

Figure 1: Proposed model architecture. The model is composed by two branches. The first branch used a variant of the Vision Transformer [4]. The second is comprised by a Pretrained Feature Extractor and MIL classifier [5].

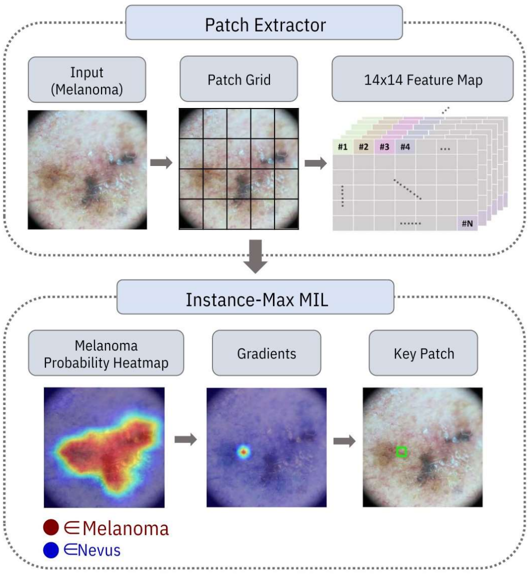

Figure 2: ROIs identification process illustration using an Instance-level MIL classifier with Max-Pooling aggregation function. The region in red represents the patches that were classified as melanoma. The region in blue represents the patches that were classified as nevus.

Table 1: Performances with different approaches: best results on ISIC 2019 [1], along with their corresponding results in PH2 [2], and Derm7pt [3] test sets. The best results for each architecture are highlighted in bold.

	Models		ISIC 2019			PH2			Derm7pt		
			BA	R-MEL	R-NV	BA	R-MEL	R-NV	BA	R-MEL	R-NV
Baseline	RN-18		83.1	73.9	92.3	71.9	45.0	98.7	70.6	51.2	90.1
	DEiT-S		**90.6**	**85.6**	**95.6**	**85.9**	**72.5**	**99.4**	76.5	57.9	**95.1**
	EViT-S		87.9	82.5	92.0	84.6	70.0	98.8	**78.8**	**65.1**	90.4
MIL-RN-18	Instance	Max	86.2	84.7	87.7	82.5	72.5	**92.5**	74.5	**65.1**	83.8
		Avg	86.1	82.7	89.4	**87.7**	**85.0**	89.4	**77.0**	62.3	**91.7**
		Topk	**88.3**	**85.2**	**91.5**	79.7	70.0	89.4	72.6	59.9	89.4
	Embedding	Max	88.0	**83.6**	92.4	79.7	80.0	86.9	74.5	61.5	**87.5**
		Avg	87.8	83.3	92.3	74.0	80.0	80.0	**75.2**	**65.5**	84.9
		Topk	**88.6**	84.3	**92.9**	**85.0**	**82.5**	**87.5**	75.0	63.9	86.1
MIL-EViT-S	Instance	Max	90.6	86.7	94.4	81.2	**72.5**	90.0	73.6	54.0	93.2
		Avg	**91.5**	86.9	**95.7**	**84.1**	70.0	**98.1**	73.4	52.4	94.4
		Topk	91.1	**87.1**	95.1	82.5	70.0	95.0	**74.7**	**56.7**	**96.7**
	Embedding	Max	90.9	86.0	95.8	82.8	70.0	**95.6**	74.2	54.4	**94.1**
		Avg	**91.3**	86.6	**96.0**	80.6	70.0	91.9	73.4	54.4	92.5
		Topk	91.1	**87.7**	94.5	**86.6**	**80.0**	91.3	**76.3**	**60.7**	91.8

References

[1]. Combalia, M., et al.: BCN20000: Dermoscopic Lesions in the Wild. arXiv preprint arXiv:1908.02288 (8 2019).

[2] Mendonca, T., et al.: Ph2 - a dermoscopic image database for research and benchmarking. Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS (2013).

[3].Kawahara, J., et al.: Seven-Point Checklist and Skin Lesion Classification Using Multitask Multimodal Neural Nets. IEEE Journal of Biomedical and Health Informatics 23 (2019).

[4] Y. Liang, C. Ge, Z. Tong, Y. Song, J. Wang, and P. Xie, "Not all patches are what you need: Expediting vision transformers via token reorganizations," 2022. [Online]. Available: https://arxiv.org/abs/2202.07800

[5] M. Ilse, J. M. Tomczak, and M. Welling, "Attention-based deep multiple instance learning," CoRR, vol. abs/1802.04712, 2018. [Online]. Available: http://arxiv.org/abs/1802.04712.

LARSyS Annual Meeting | 17 & 18 July 2023