Can Self-Training Identify Suspicious Ugly Duckling Lesions?

Mohammadreza Mohseni[1, 2], Jordan Yap[2], William Yolland[2], Arash Koochek[4], M Stella Atkins[1, 2, 3]

[1] School of Computing Science
Simon Fraser University

[2] MetaOptima Technology Inc

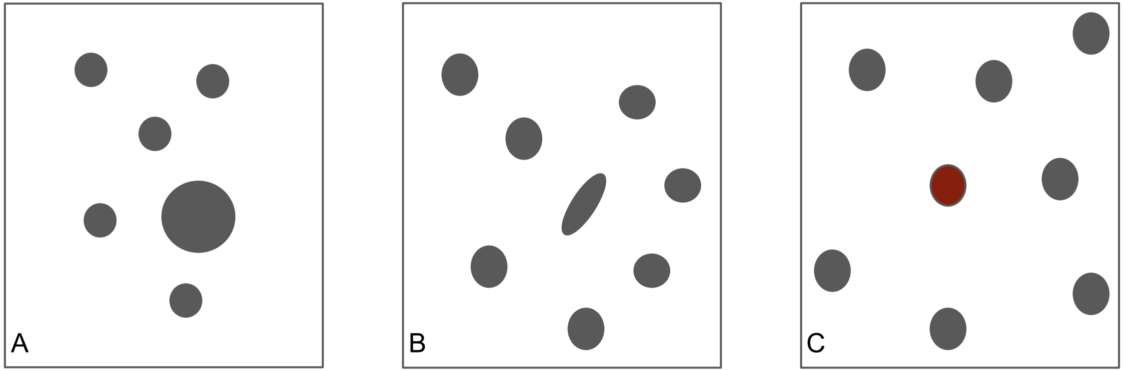
[3] Department of Skin Science and Dermatology,
University of British Columbia

[4] Banner Health

What is an Ugly Duckling?

Ugly Duckling (UD) is the lesion (or lesions) that looks different from other lesions in its proximity. The difference can be in:

A. Size
B. Shape
C. Color



Why is Ugly Duckling Sign Important?

Based on previous research:

Ugly Duckling Sign ⟷ Has Correlation With ⟷ Existence of Malignant Melanoma

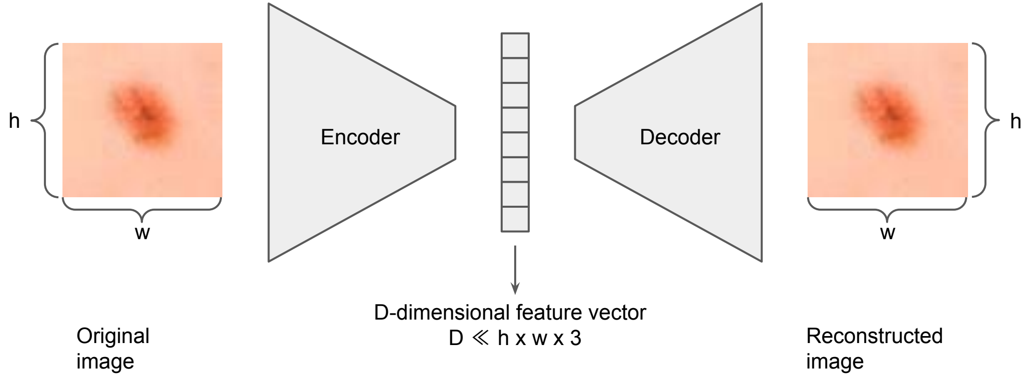
Allows faster decisions when making diagnosis from visual inspection of:

- Patients
- Photos

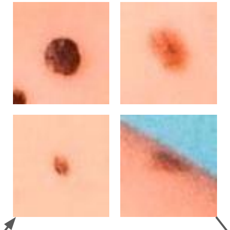
An Outlier Detection Problem

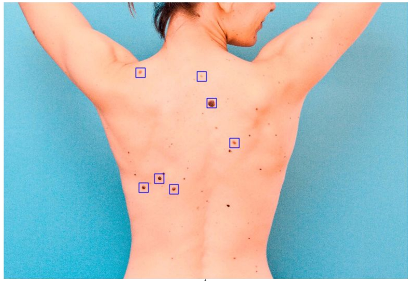
We can model ugly duckling detection problem as an outlier detection problem. Ugly duckling lesions are visually outliers by definition. We identify the outlier lesions, by incorporating a variational autoencoder (VAE).

Given an image of a single lesion as input, VAE is asked to reconstruct the same image in the output. Previous work has shown that VAEs can't reconstruct anomalous (outlier) samples well, leading to high reconstruction loss on those samples.



Our Initial Approach - Pipeline





Detection

ResNet-18 backbone with
Feature Pyramid Network

Extraction

Center-cropping 64x64
patches per each lesion

Outlier
Detection

Variational Autoencoder

Our Initial Approach - Ugly Duckling Detection

In order to detect Ugly duckling lesions, we initially:

1. Extracted all of the lesions from a Total Body Photography (TBP) image
2. Self-Trained a Variational Autoencoder on all of the extracted lesions
3. Selected items with high reconstruction loss as ugly duckling lesions

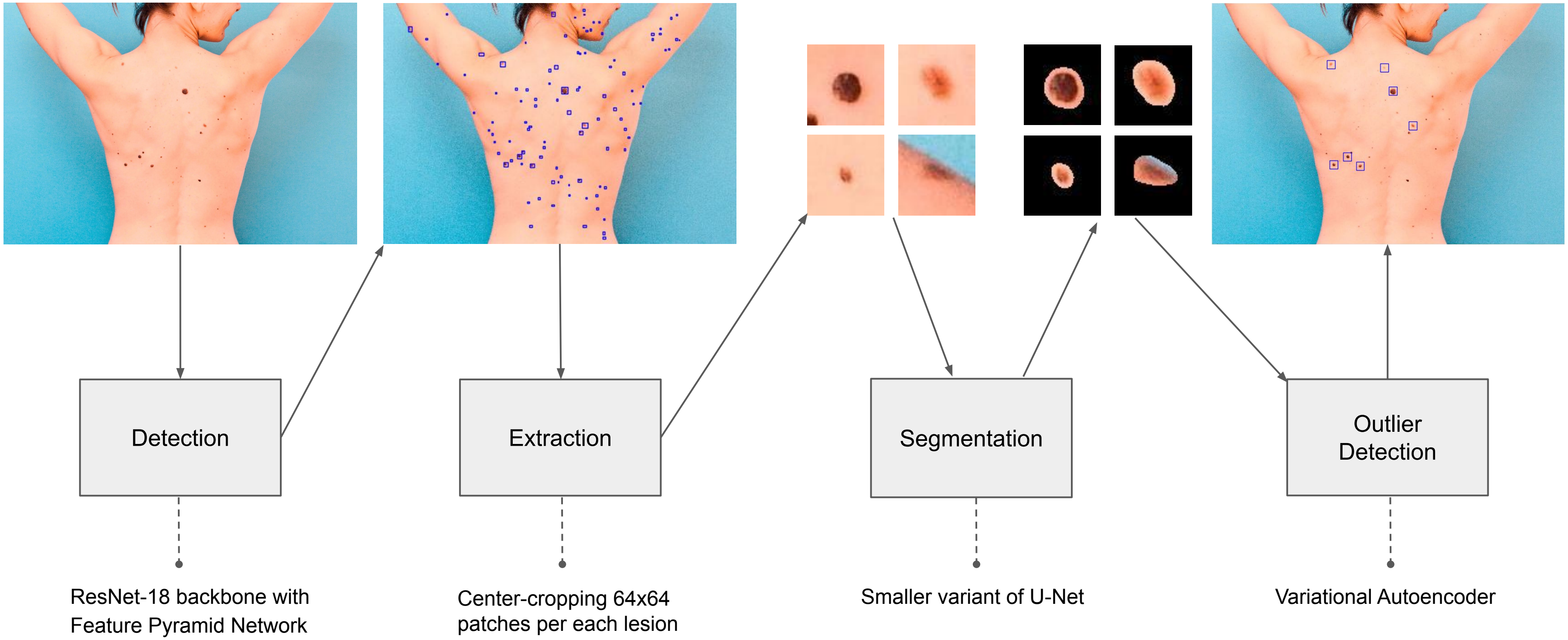
However, there were two major limitations of this approach:

- Lighting conditions and surrounding area of the lesions played a huge role in our predictions
- Training the variational autoencoder for each TBP image increased the response time for each query to 2 minutes.

Improving the Results

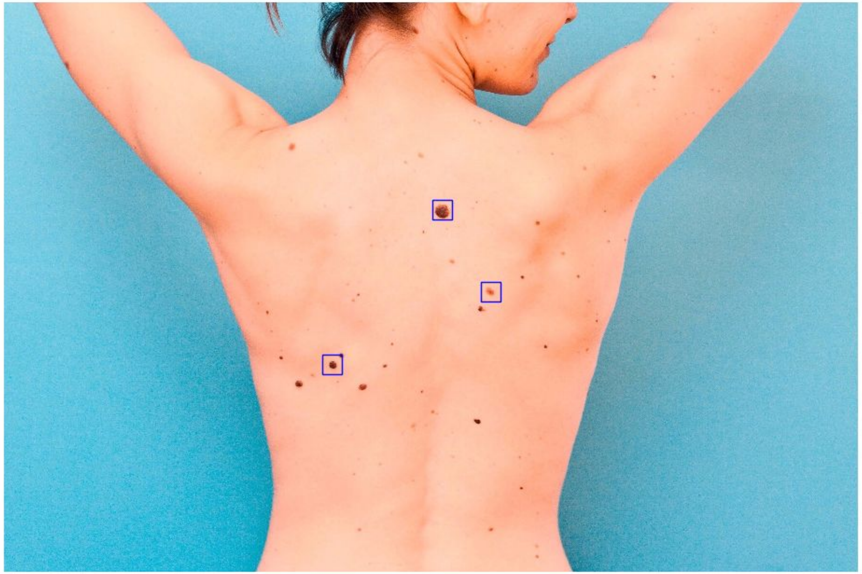
- We added a segmentation module to our pipeline, which greatly reduced the amount of noise in our result (e.g. shadow and lesion location noise)
- We found that by self-training the VAE on 300 unlabelled TBP images, we can greatly reduce the training time.
- Additionally, we found that by using feature-vector distances instead of reconstruction loss, produced better results for the algorithm.

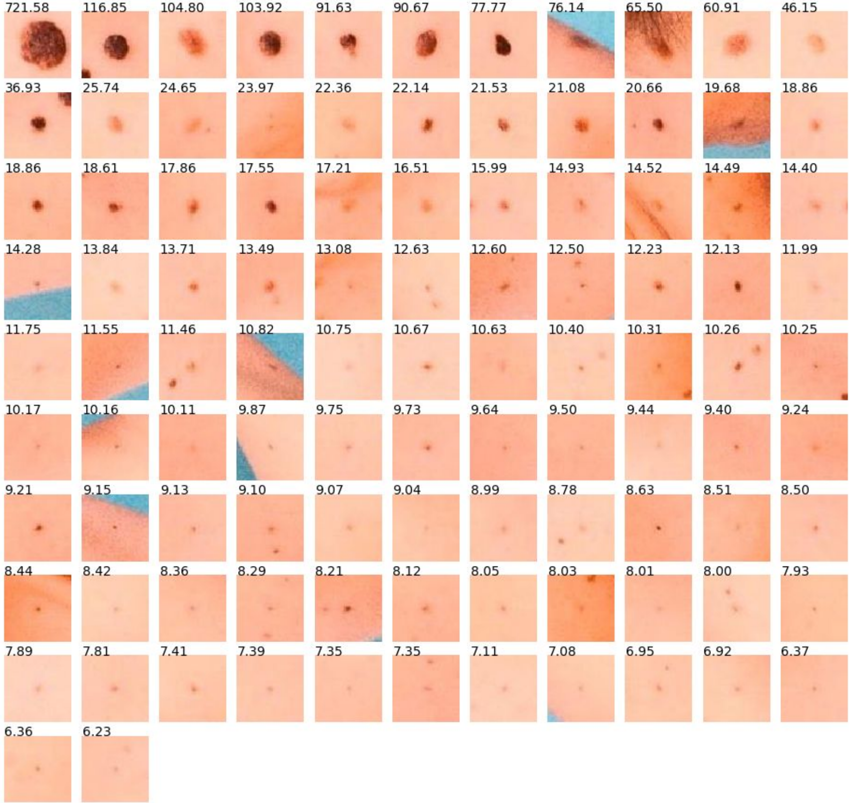
Our Approach - Adding Segmentation Module



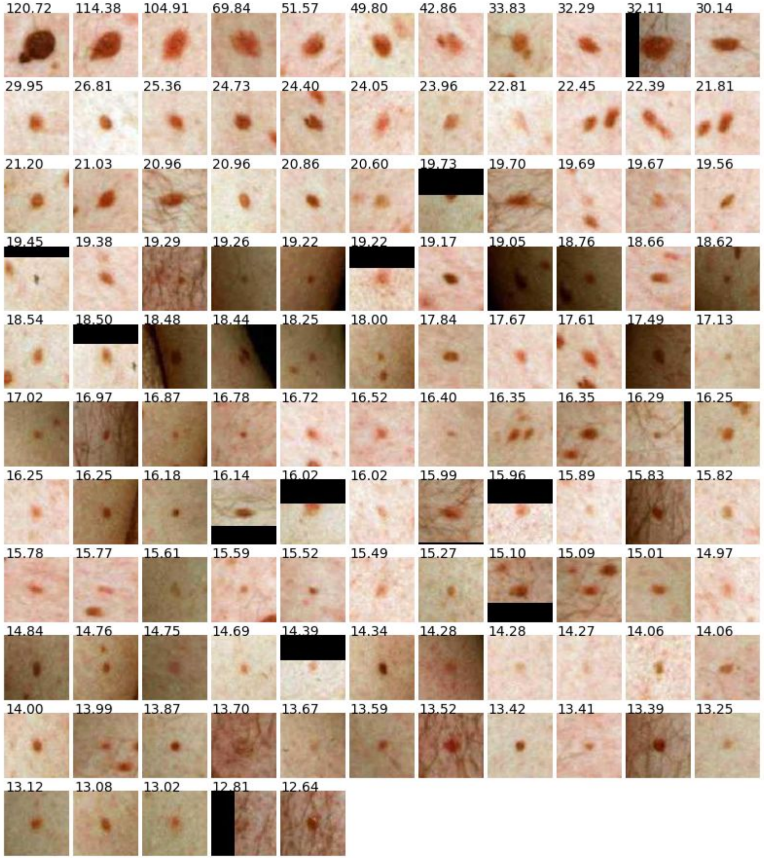
Qualitative Results for Patient #1

The numbers above each mole are showing how different it looks from the others. The upper left has the highest value and bottom right has the smallest value.



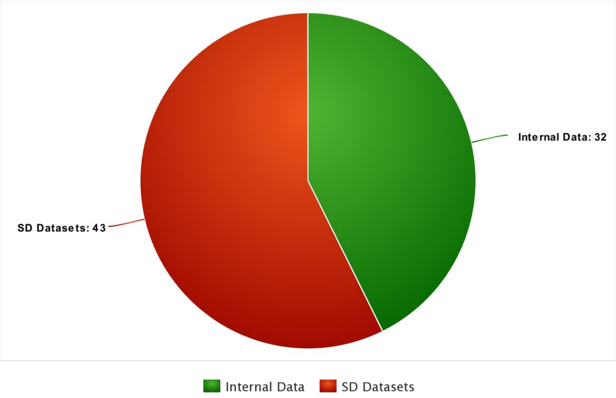


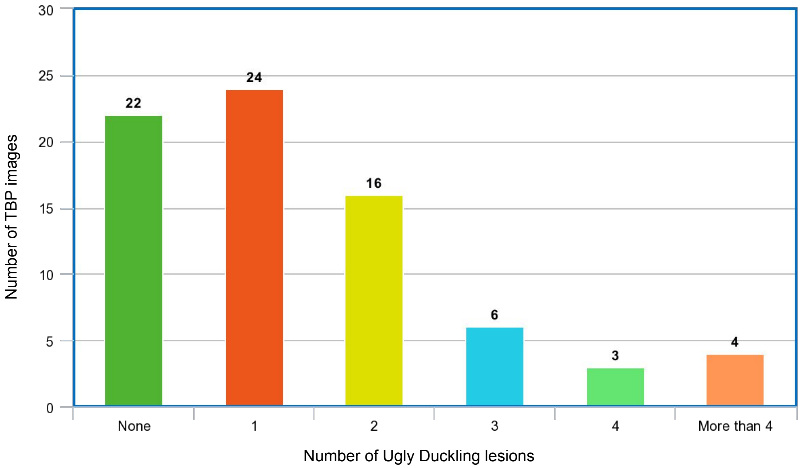
Qualitative Results for Patient #2



Evaluation Data

We evaluated our algorithm on 75 TBP images. Each lesion in a TBP image is labelled as either “Ugly Duckling” or not, by a board certified dermatologist.





Quantitative Evaluation

We report our algorithm's performance by two major types of metrics:

- Ranking based metrics
  - Mean Average Precision (MAP)
  - Mean Reciprocal Rank (MRR)
  - 3-Agreement and 7-Agreement
- Binary Classification based metrics
  - Accuracy
  - Sensitivity
  - Specificity

Evaluation - Ranking Based Metrics

For the meaningful calculation of ranking based metrics, we used a subset of TBP images which contained at least one ugly duckling.

MAP	MRR	Top-3 Agreement	Top-7 Agreement	Support Size
0.659	0.721	86.79%	94.34%	53

Evaluation - Binary Classification Based Metrics

In order to obtain binary predictions from ranking, we applied a threshold to the calculated distances. Any lesion with distance above the threshold was predicted as ugly duckling.

$$threshold = mean(distances) + \min \begin{cases} mean(distances) \\ std(distances) \end{cases}$$

Metric	Micro (Calculated over all lesions in all TBP images)	Macro (Averaged over all TBP images)
Accuracy	94.16 %	94.23 %
Sensitivity	72.07 %	71.91%
Specificity	94.70 %	94.95 %

Macro sensitivity is calculated only on the TBP images with at least one ugly duckling lesion.

Summary

In this work we showed that:

- ✓ Ugly duckling lesions can be accurately identified from TBP images with the help of self-training
- ✓ We evaluated our work against 75 images from two data sources
- ✓ We presented two types of metrics for evaluation of ugly duckling detection algorithms

Future Work

- Using the result of our work to enhance the performance of Melanoma detection from TBP images
- Investigating if features extracted by the VAE are correlated with size, shape, and color of the lesions
- Investigating the effect of replacing the features extracted by VAE with hand crafted features
- Publishing a large public dataset for better evaluation of such algorithms

References

[1] J J Grob and J J Bonerandi. The ‘ugly duckling’ sign: identification of the common characteristics of nevi in an individual as a basis for melanoma screening. Archives of dermatology, 134(1):103–4, Jan 1998.

[2] Wenqian Liu, Runze Li, Meng Zheng, Srikrishna Karanam, Ziyan Wu, Bir Bhanu, Richard J. Radke, and Octavia Camps. Towards Visually Explaining Variational Autoencoders. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 8639–8648, June 2020. ISSN: 2575-7075

[3] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. Ssd: Single shot multibox detector. Lecture Notes in Computer Science, page 21–37, 2016.

[4] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, pages 234–241, Cham, 2015. Springer International Publishing.

[5] Jufeng Yang, Xiaoping Wu, Jie Liang, Xiaoxiao Sun, Ming-Ming Cheng, Paul L. Rosin, and Liang Wang. Self-paced balance learning for clinical skin disease recognition. IEEE Transactions on Neural Networks and Learning Systems, 2019

[6] Luis R. Soenksen, Timothy Kassis, Susan T. Conover, Berta Marti-Fuster, Judith S. Birkenfeld, Jason Tucker-Schwartz, Asif Naseem, Robert R. Stavert, Caroline C. Kim, Maryanne M. Senna, Jose Aviles-Izquierdo, James J.Collins, Regina Barzilay, and Martha L. Gray. Using deep learning for dermatologist-level detection of suspicious pigmented skin lesions from wide-field images. Science Translational Medicine, 13(581):eabb3652, Feb. 2021.

Thank you for your attention!

Contact us:

Mohammadreza Mohseni: mmohseni@sfu.ca

Jordan Yap: jordanjyap@gmail.com

William Yolland: william@metaoptima.com

Arash Koochek: arash.koochek@bannerhealth.com

M Stella Atkins: stella@sfu.ca