X-MIRAD: EXplainable Medical Image Retrieval (Supplement)

Overview

1 Additional details on the datasets and evaluation

Models were always trained only on the training data, and tested on a separate held-out test set. The testing data was class-balanced, so that the three classes in each dataset had the same number of examples. For the COVIDx dataset, there were 100 examples of normal, pneumonia, and COVID-19 cases, for a total of 300 images. This testing data matches the publicly available test split of the COVIDx dataset. For the ISIC 2017 dataset, there were 90 examples of nevi, keratosis, and melanoma cases, for a total of 270 images. We use the 90 provided keratosis examples, and randomly selected 90 nevi and melanoma examples from the larger set of testing data.

For the image retrieval problem, each of the images in the testing data was used once as a query image, with all other images considered part of the image archive. Retrieved image results were considered correct if they had the same class as the query image. The mean average precision and mean precision metrics reported in the paper are averages over all queries in the testing data.

2 Additional t-SNE visualizations

2.1 COVID-19 chest X-ray dataset

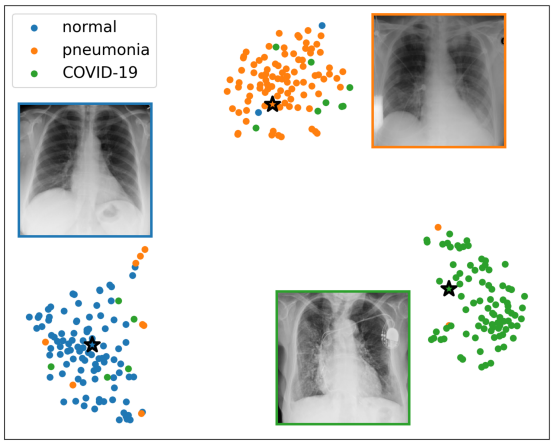

Figure S1: Two-dimensional t-SNE visualization of the learned embeddings on test data. Normal (blue), pneumonia (orange), and COVID-19 cases (green) cluster in different regions of the latent space. For each class, we show the closest image to the class centroid (marked with stars).

2.2 ISIC 2017 skin lesion dataset

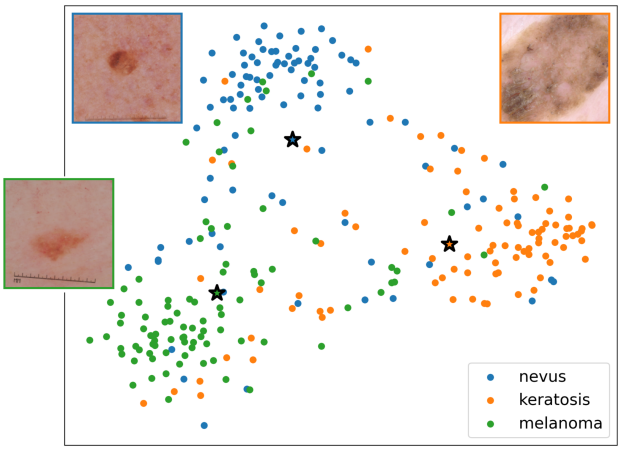

Figure S2: Two-dimensional t-SNE visualization of the learned embeddings on test data. Nevus (blue), keratosis (orange), and melanoma cases (green) cluster in different regions of the latent space. For each class, we show the closest image to the class centroid (marked with stars).

3 Additional similarity-based saliency maps

Here, we visualize additional saliency maps for each dataset. As the attention-based and activation mapping based saliency maps produced nearly equivalent results, we only visualize the activation-based saliency maps.

3.1 COVID-19 chest X-ray dataset

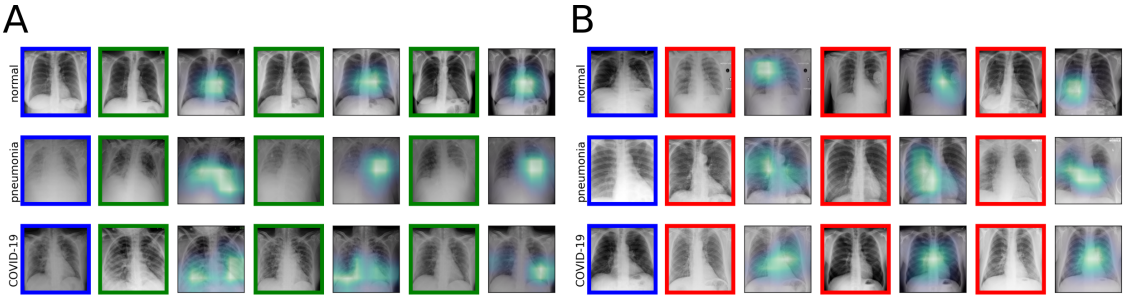

Figure S3: Example activation-based saliency maps [2, 3] on the COVID-19 dataset. (A) Queries where the top-3 retrieved images are correct. (B) Queries where the top-3 retrieved images are incorrect. Query images are marked in blue, correct results are marked in green, and incorrect results are marked in red.

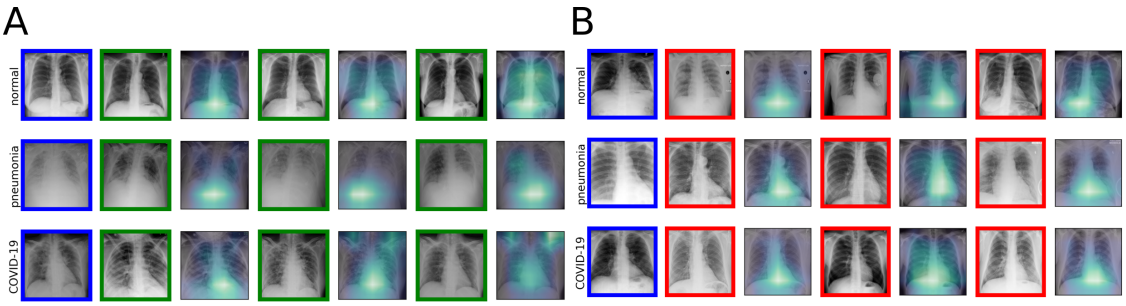

Figure S4: Activation-based saliency maps for model initialized with random weights.

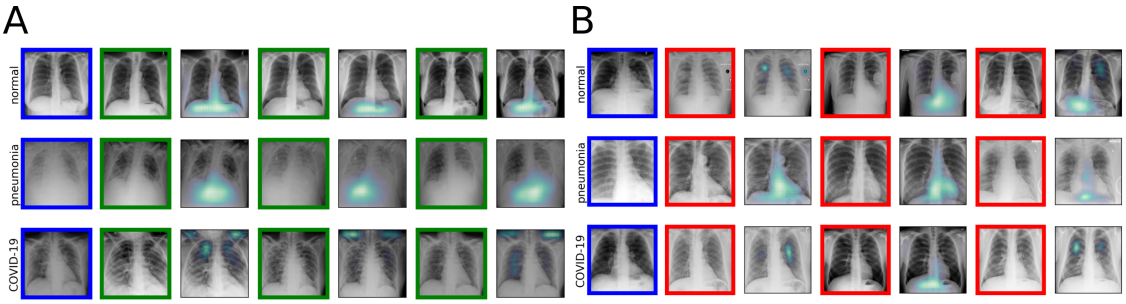

Figure S5: Occlusion-based saliency maps for model initialized with random weights.

3.2 ISIC 2017 skin lesion dataset

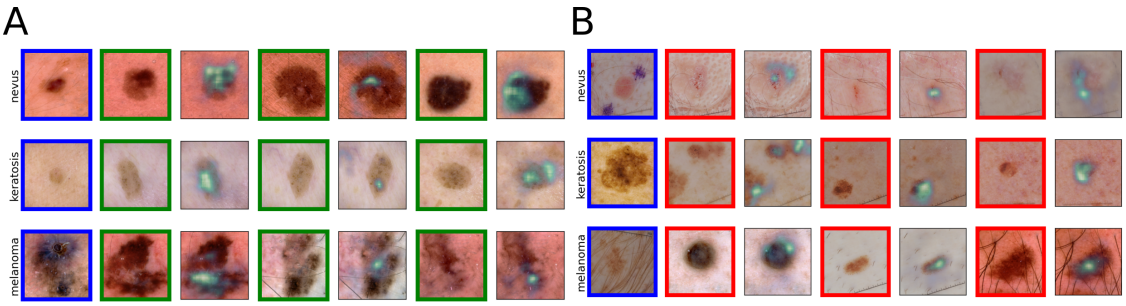

Figure S6: Example occlusion-based saliency maps [1] on the ISIC 2017 dataset. (A) Queries where the top-3 retrieved images are correct. (B) Queries where the top-3 retrieved images are incorrect. Query images are marked in blue, correct results are marked in green, and incorrect results are marked in red.

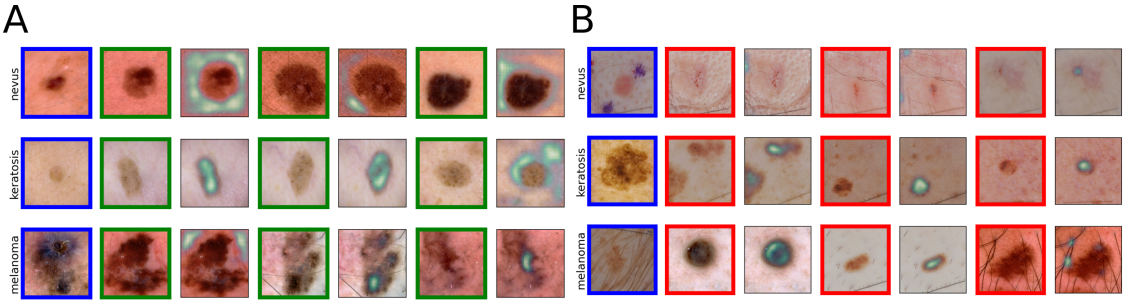

Figure S7: Occlusion-based saliency maps for model initialized with random weights.

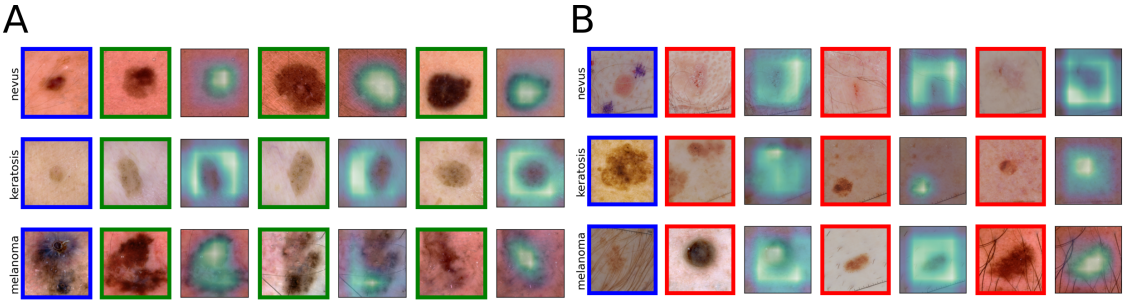

Figure S8: Activation-based saliency maps for model initialized with random weights.

References

[1] Bo Dong, Roddy Collins, and Anthony Hoogs. Explainability for content-based image retrieval. In *CVPR Workshops*, pages 95–98, 2019. 4

[2] Abby Stylianou, Richard Souvenir, and Robert Pless. Visualizing deep similarity networks. In *2019 IEEE winter conference on applications of computer vision (WACV)*, pages 2029–2037. IEEE, 2019. 3

[3] Sijie Zhu, Taojiannan Yang, and Chen Chen. Visual explanation for deep metric learning. *arXiv preprint arXiv:1909.12977*, 2019. 3