

Graph Geodesics to Find Progressively Similar Skin Lesion Images

Jeremy Kawahara, Kathleen P. Moriarty, and Ghassan Hamarneh

Medical Image Analysis Lab, Simon Fraser University, Burnaby, Canada
{jkawahar, kmoriart, hamarneh}@sfu.ca

Abstract. Skin conditions represent an enormous health care burden worldwide, and as datasets of skin images grow, there is continued interest in computerized approaches to analyze skin images. In order to explore and gain insights into datasets of skin images, we propose a graph based approach to visualize a progression of similar skin images between pairs of images. In our graph, a node represents both a clinical and dermoscopic image of the same lesion, and an edge between nodes captures the visual dissimilarity between lesions, where dissimilarity is computed by comparing the image responses of a pretrained convolutional neural network. We compute the geodesic/shortest path between nodes to determine a path of progressively visually similar skin lesions. To quantitatively evaluate the quality of the returned path, we propose metrics to measure the number of transitions with respect to the lesion diagnosis, and the progression with respect to the clinical 7-point checklist. Compared to baseline experiments, our approach shows improvements to the quality of the returned paths.

Keywords: Graph geodesics, skin lesions, visualizing similar images

1 Introduction

Globally, skin conditions are the fourth most common cause of healthy years lost due to disability [6], and represent the most common reason for a patient to visit their general practitioner in studied populations [16]. Skin conditions such as malignant melanoma, a common cancer, can be fatal [14]. Many groups recognize the potential for computerized systems to analyze skin lesions and help reduce the burden on health care, and much work has gone into developing computerized systems to diagnose skin disorders [13]. Typically, such systems take as input a skin lesion image, and output either a discrete label or the probability that this lesion has a particular diagnosis. For example, Estevan et al. [5] used a human designed taxonomy to partition clinically similar images into classes that have a similar number of samples. They fine-tuned a Convolutional Neural Network pretrained over ImageNet [15] to classify skin lesions, and achieved results comparable to human dermatologists. While knowing the probability that the image contains a particular type of skin lesion is a worthwhile goal, a disadvantage to this approach is that it is a “black box”, where the user gains no insights into the automated diagnosis or of the underlying dataset of skin images.

A different approach from classification that offers some insights into the dataset or diagnosis is to adopt an image retrieval based approach. For example, Bunte et al. [3] extracted colour features from clinical skin images, learned a supervised transformation of these features, and retrieved images in a dataset based on the k nearest neighbours to these features. These returned images can be displayed to the user, giving insights into the appearance of similarly diseased images and allow the diagnosis to be inferred. Kawahara et al. [10] displayed a network graph visualization based on the nearest neighbours to a single query image, which allow users to efficiently search the space of similar lesion images.

Another approach to visualize *general images* was proposed by Hegde et al. [7], where rather than retrieving the k nearest images to a single query image, their approach uses two query images (a source and target) to retrieve a list of images that progress in visual similarity between the two images. They accomplish this by representing images as nodes in a graph, where the edges between nodes indicate their pair-wise distance, and the geodesic (shortest path) between source and target nodes represents a visually smooth progression of images. A similar approach for *general images* was recently implemented online [12], which is based on an experimental visualization tool as part of Google Arts and Culture [11]. In other works, representing images as nodes to find an optimal path between nodes has been used to guide subject-template brain registration in MR images [8].

In this work, we apply a similar method to find images of skin lesions that visually progress between a source and target lesion. This visualization approach may be useful for clinicians who wish to find reference images of hard to classify, visually challenging “borderline” cases across types of skin diseases (e.g., note the visually challenging aspects in distinguishing clark nevus from melanoma in Fig. 1 *bottom row*). Another use may be to show or predict the visual progression over time between a low-risk benign lesion to a malignant lesion (e.g., progression in Fig. 3 *bottom row*). This may give insights into the progression of the disease (e.g., Clark/Dysplastic nevi is potentially a precursor to melanoma and studies estimate that 20-30% of melanomas come from nevi [4]), or serve as a useful reference for patients to monitor and compare the progression of their own lesion. In these potential applications, the target images could be from either a set of predefined reference images, or the geodesics to each of the nearest unique diseases could be automatically shown.

To the best of our knowledge, this is the first work that has applied geodesic paths to visualize skin lesion images. In contrast to previous work [7,8,11,12], we propose to let each node in our graph represent images from two modalities (a dermoscopic and a clinical image), where the edge weights are influenced by both types of images. We apply an exponential function to the pair-wise dissimilarity measures, and show how this results in longer paths of higher quality without risking disconnected graphs. Finally, we propose measures to quantitatively evaluate the quality of our paths, which is lacking in prior work. These proposed quality measures are particularly important as without them, we would need to qualitatively inspect each path.

2 Methods

A skin lesion can be captured by both a dermatoscope (producing a *dermoscopic* image x_d), and a photo camera (producing a *clinical* image x_c), where the dermoscopic images show a more standardized view of the lesion, and the clinical images are non-standardized and often show additional contextual information (e.g., the body part the lesion is on) not available in the dermoscopic images. Given a dataset of skin lesions, the i -th skin lesion is represented by a dermoscopic and clinical pair of images $(x_d^{(i)}, x_c^{(i)})$. We create a graph where each pair of images $(x_d^{(i)}, x_c^{(i)})$ are represented by a single node $v^{(i)}$, and an edge $e^{(ij)}$ encodes the dissimilarity between nodes i and j . Our goal is to find a set of nodes $(v_0^{(s)}, v_1^{(i)}, \dots, v_{R-1}^{(j)}, v_R^{(t)})$ of an unknown length R such that the initial node $v_0^{(s)}$ is a given source node (the superscript identifies the lesion, and the subscript indicates the position in the returned path), the R -th node is a given target node $v_R^{(t)}$, and the intermediate nodes $(v_1^{(i)}, \dots, v_{R-1}^{(j)})$ represent lesions that visually progress between the source and target nodes. We find these intermediate nodes using Dijkstra’s algorithm, which computes the geodesic between the source and target and returns a path of nodes representing a progression of visually similar lesions.

The key components that we now examine in detail are how to: extract image features that capture the salient properties of skin images, compute local dissimilarity between pairs of skin lesion images, weigh and connect the node edges using multi-modal images, and quantitatively evaluate the quality of the returned paths.

Skin Images as Deep Pretrained Neural Nets Responses. The responses of skin images with deep convolutional neural networks pretrained over ImageNet [15] have shown to be effective feature vectors for skin lesion classification despite the differences in appearance between skin lesions and natural images [9]. We use a similar approach to compute feature vectors as in [9], and for a particular image, extract the responses from the first fully connected layer of VGG16 [17], and average the responses over a set of predefined image augmentations,

$$\Phi(x)_m = \frac{1}{|II|} \sum_{\pi \in II} \phi(\pi(x - \mu))_m \quad (1)$$

where π is a function to augment an image (e.g., left-right flip); II is the set of $|II|$ number of image augmentations; $\phi(\cdot)_m$ extracts the m -th response of the first fully connected layer of VGG16; and, μ represent the mean pixel over the training data from ImageNet, which is subtracted from the skin lesion image x . The resulting feature vector $\Phi(x)$ represents a single lesion image by averaging the augmented responses over a single image, without increasing the dimensions of the feature vector.

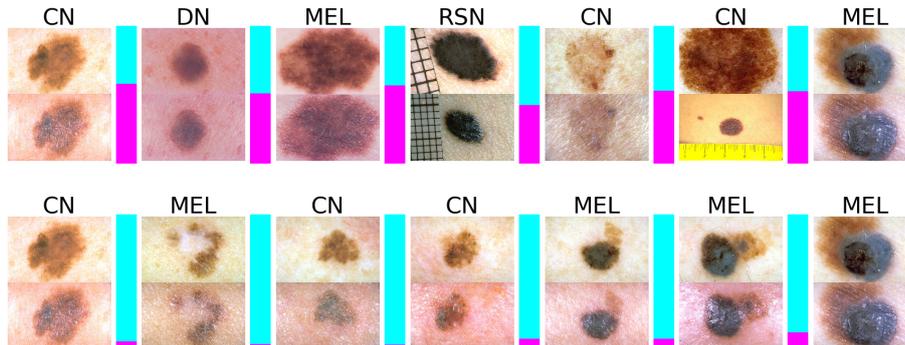


Fig. 1: An example random path (*top*) and geodesic returned from the proposed method (*bottom*), where the *leftmost* and *rightmost* image represent the source and target nodes, respectively. The dermoscopic image above the clinical image in each row. The *magenta bar* indicates the dissimilarity between adjacent images, where a higher bar indicates that they are more dissimilar.

Local Image Dissimilarity. Given two feature vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^M$ (which represent the responses of two skin images), we compute the dissimilarity between them as the cosine distance raised to the p -th power,

$$\mathcal{D}(\mathbf{u}, \mathbf{v}) = \left(1 - \frac{\sum_i^M u_i v_i}{\sqrt{\sum_i^M u_i^2} \sqrt{\sum_i^M v_i^2}} \right)^p \quad (2)$$

where setting $p \neq 1$ non-linearly changes the dissimilarity between vectors. By using a high p (e.g., $p = 4$), we assign very low values to edges connecting similar images, thus encouraging geodesics to pass through nearby nodes of similar images, avoiding very short paths even in the case of complete graphs, i.e., fully connected graphs (further discussed in the Results section). Other distance measures are possible (e.g., L_1, L_2), and we found them to give empirically similar results. Fig. 1 shows the dissimilarity between pairs of images (dissimilarity is displayed in magenta using $p = 1$ for clarity).

Multi-modal Edge Weights. We define the edge weight $e^{(ij)}$ between nodes i and j as a weighted sum based on both the dermoscopic and clinical images,

$$e^{(ij)} = \alpha \mathcal{D}(\Phi(x_d^{(i)}), \Phi(x_d^{(j)})) + (1 - \alpha) \mathcal{D}(\Phi(x_c^{(i)}), \Phi(x_c^{(j)})) \quad (3)$$

where $\mathcal{D}(\cdot)$ is a function that computes the dissimilarity (Eq. 2) between the feature vectors $\Phi(\cdot)$ computed in Eq. 1; and α weighs the influence of the dermoscopic and clinical images ($0 \leq \alpha \leq 1$). Increasing α causes an edge to be more influenced by the dermoscopic image than the clinical image, which may be desired as dermoscopic images contain more salient lesion properties.

Node Connectivity. To form the graph, we must decide on the connectivity of nodes. This can be done by connecting the k nearest neighbours (where nearest is defined via Eq. 2) to each node with an edge. However, choosing k is challenging as a large k (e.g., a complete graph) increases computational complexity and can lead to very short paths being returned when a direct edge exists between any pair of source and target nodes. Too small a k can lead to disconnected graphs, where no path exists between the source and target nodes. In the Results, we experiment with different values of k and show that by setting a high value of p in Eq. 3, the returned paths remain longer even in the case of complete graphs.

Surrogate Measures of Path Quality. While we provide qualitative results through visualizing the returned paths (Fig. 3), we also propose the following measures to quantitatively evaluate the quality of the returned paths. We define a *quality path* as a smooth visual progression of images. However, this definition is hard to precisely define and directly measure. Thus we propose a surrogate measure that uses the diagnoses of the lesions, as skin lesion datasets are often accompanied with a corresponding clinical diagnosis $y^{(i)}$ (e.g., melanoma, nevus), indicating the disease type of the i -th lesion $x^{(i)}$, where $y^{(i)}$ is an attribute of node $v^{(i)}$. Our assumption is that lesions with the same diagnosis will likely be visually similar, and that a high quality path will have a smooth progression with respect to the lesion diagnosis. In order to give a high cost to paths that frequently change neighbouring labels, we define the *transition cost* as,

$$\text{trans}(v_0^{(s)}, v_1^{(i)}, \dots, v_{R-1}^{(j)}, v_R^{(t)}) = \frac{1}{R-1} \sum_{r=1}^R \left(1 - \delta(y_r^{(a)} - y_{r-1}^{(b)})\right) \quad (4)$$

where R is the number of nodes in the returned path; and $y_r^{(i)}$ indicates the skin lesion diagnosis for the r -th returned path node corresponding to node $v_r^{(i)}$ (e.g., $y_0^{(s)}$ and $y_R^{(t)}$ correspond to the labels of the source and target nodes $v_0^{(s)}, v_R^{(t)}$ respectively). The Dirac delta function $\delta(\cdot)$ returns 1 if the two labels have the same class, and 0 otherwise.

Our second surrogate quality measure quantifies the progression of the 7-point score between the source and target nodes. The 7-point score is a clinical measure of melanoma based on the visual presence of seven criteria (e.g., irregular streaks) within a lesion. The weighted sum of these seven criteria form the 7-point score $\hat{y} \in \mathbb{Z} [1]$, where $\hat{y}^{(i)}$ is an attribute of node $v^{(i)}$. We assume that a quality path will have 7-point scores that smoothly progress from a low to high score, as higher scores indicate the presence of lesion more indicative of melanoma (and vice versa). We define the *progression cost* as,

$$\text{progress}(v_0^{(s)}, \dots, v_R^{(t)}) = \frac{1}{R} \sum_{r=1}^R \left(\max \left[\left(\text{sgn}(\hat{y}_0^{(s)} - \hat{y}_R^{(t)}) (\hat{y}_r^{(i)} - \hat{y}_{r-1}^{(j)}) \right), 0 \right] \right) \quad (5)$$

where $\text{sgn}(z)$ returns the sign of the difference between the source and target node scores,

$$\text{sgn}(z) = \begin{cases} 1, & \text{if } z = 0 \\ \frac{z}{|z|}, & \text{otherwise.} \end{cases} \quad (6)$$

This measure returns a cost of 0 if the 7-point score consistently decreases, increases, or remains constant along the path between the source and target nodes, and penalizes by the magnitude of the change otherwise. This approach, however, will always compute a 0 cost if the path only consists of the source and target nodes. As this is a degenerate case, we ignore the progression costs for paths of length two when computing results, and note that this measure is biased to return lower costs for shorter paths, and is thus most informative when comparing paths with the same number of nodes.

3 Results

Data. We test our proposed approach and surrogate measures using the Interactive Atlas of Dermoscopy [2] skin dataset. This dataset contains 1011 cases of skin lesions, where all but four cases are captured by both a clinical x_c and dermoscopic x_d image (in the four cases missing x_c , we set $x_c = x_d$). Each case has a class label y that represents a known lesion diagnosis, and a 7-point score \hat{y} . The diagnosis y can take on one of the 15 class labels: basal cell carcinoma (BCC), blue nevus (BN), clark nevus (CN), combined nevus (CBN), congenital nevus (CGN), dermal nevus (DN), dermatofibroma (DF), lentigo (LT), melanoma (MEL), melanosis (MLS), miscellaneous (MISC), recurrent nevus (RN), reed or spitz nevus (RSN), seborrheic keratosis (SK), and vascular lesion (VL). The 7-point score $\hat{y} \in \mathbb{Z}$ ranges between 0 and 7 (in this dataset), where a higher score indicates the lesion has visual properties more indicative of melanoma. The lesion diagnosis and the 7-point score are only used to quantify the quality of the returned paths, and are not used to form the graph. We randomly select a set of 1000 pairs of source and target nodes which are used across all experiments.

Recovering Synthetic Paths. We start by testing if our proposed approach can recover the path of images created by a progressive synthetic transformation. To do this, we crop the image by removing 15% of the pixels at the borders of the images, and repeat this five times. This progressively enlarges the lesion over a series of five images. We added these five synthetic images to our dataset, select the original image as the source and the final synthetic image as the target ($p = 4$ and $k = 30$). We find our approach not only recovers all synthetic images, but it recovers the correct sequence of synthetic images, i.e. in the order they were synthesized (Fig. 2), indicating that this approach and the feature vectors are sensitive to scale despite the CNN being trained on images at multiple scales.

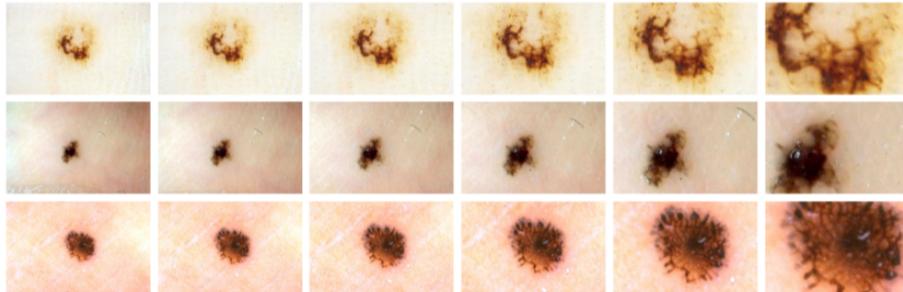


Fig. 2: Synthetic examples: Here the *leftmost* images represent the source nodes, which belong to the original (non-enlarged) dermoscopic images in the dataset. The *rightmost* images represent target nodes, which were the last of the progressively enlarged images. The returned geodesic path is represented by the images in between. Note that the returned geodesic included all five synthetic images, in proper order of increasing enlargement.

Retrieving Paths from a Complete and Non-Complete Graphs. For our first experiment, in Table 1 **row 1.1** (*complete graph with $p = 1$*) we report results using a complete graph (i.e., $k = 1011$) using only the dermoscopic images (i.e., $\alpha = 1$ in Eq. 1) and setting $p = 1$ in Eq. 3. We observe that when using a complete graph, the returned paths often consist of only the source and target nodes as their shared edge yields the shortest path. This experiment highlights the need to either prune the edges in the graph or modify the edge weights. Following the approach of [12], we form a new graph where each node is connected to its $k = 30$ neighbours. **Row 1.2** (*non-complete graph with $p = 1$*) shows that restricting the node connectivity increases the number of nodes in the returned path and improves the transition cost (note that the progression cost performs worse as it is biased towards paths with fewer nodes, and is thus most informative when comparing paths with a similar number of nodes).

Paths with Exponential Edge Weights. While decreasing node connectivity (i.e., lowering k) results in longer paths, care must be taken when choosing k , as reducing k increases the risk of forming disconnected graphs where no path exists between a source and target node. Thus instead of pruning edges, our next experiment (**row 1.3** *complete graph with $p = 4$*) shows how applying an exponential function (i.e., $p = 4$ in Eq. 3) to the dissimilarity function results in longer paths of higher quality even in a complete graph. By removing the need to prune graphs (i.e., choose k), we guarantee a path to exist, while still preventing short paths. If we are not concerned with disconnected graphs, we can combine edge pruning using k neighbours with the increased p , to match the computational efficiency of a pruned graph without penalty to quality (**row 1.4** *non-complete graph with $p = 4$*). For the remaining experiments, we use $p = 4$ and non-complete graphs with $k = 30$, as our graphs remained connected.

Table 1: Quantitative results of the returned paths using the proposed surrogate quality measures. The *Img.* column indicates if the input was a dermoscopic image x_d , clinical image x_c , or included both. k represents the number of nearest neighbours used to form edges that connect nodes. *Aug.* indicates if the image was augmented or not when forming the image feature vector. *Trans.*, *Progress.*, indicates the average and standard deviation transition and progression cost as defined in the text. *Num. Path* shows the average and standard deviation number of nodes in the computed path.

| Exp. | Img. | Aug. | p | k | Ordered | Trans. | Progress. | Num. Path |
|----------|------------|------|-----|------|----------|-----------------|-----------------|------------------|
| 1.1 | x_d | ✗ | 1 | 1011 | min-path | 0.76 ± 0.42 | 0.10 ± 0.19 | 2.02 ± 0.13 |
| 1.2 [12] | x_d | ✗ | 1 | 30 | min-path | 0.64 ± 0.34 | 0.23 ± 0.26 | 3.59 ± 0.85 |
| 1.3 | x_d | ✗ | 4 | 1011 | min-path | 0.56 ± 0.26 | 0.37 ± 0.20 | 8.11 ± 2.87 |
| 1.4 | x_d | ✗ | 4 | 30 | min-path | 0.56 ± 0.26 | 0.37 ± 0.20 | 8.12 ± 2.87 |
| 1.5 | x_d | ✓ | 4 | 30 | min-path | 0.55 ± 0.25 | 0.35 ± 0.17 | 9.16 ± 3.62 |
| 1.6 | - | - | - | - | random | 0.76 ± 0.19 | 0.54 ± 0.24 | 9.16 ± 3.62 |
| 1.7 | x_d | ✓ | - | - | linear | 0.58 ± 0.25 | 0.44 ± 0.21 | 9.16 ± 3.62 |
| 1.8 | x_c | ✗ | 4 | 30 | min-path | 0.65 ± 0.18 | 0.46 ± 0.20 | 10.64 ± 5.08 |
| 1.9 | x_d, x_c | ✗ | 4 | 30 | min-path | 0.45 ± 0.24 | 0.34 ± 0.19 | 7.90 ± 3.27 |
| 1.10 | x_d, x_c | ✓ | 4 | 30 | min-path | 0.45 ± 0.23 | 0.34 ± 0.17 | 8.86 ± 3.73 |

Comparing Random and Linearly Interpolated Path. In row 1.5 (*augmented images*) we augment the feature vector with left-right image flips (Eq. 1), which results in longer geodesics paths and minor improvements to the path quality. We form a path with an equal number of nodes as those returned in the geodesic path in the previous experiment (from row 1.5) by randomly sampling nodes (without replacement). As the labels in our dataset are highly imbalanced, these random paths give us a baseline quality score (**row 1.6** *random paths*). We also compare our method by ignoring the graph, and instead using linearly interpolated feature vectors between the source and target feature vectors. These interpolated feature vectors are uniformly separated to match the number of returned nodes in row 1.5. The nearest unique neighbour to this interpolated feature vector is used to form the path. **Row 1.7** (*linear paths*) shows that this approach yields paths of worse quality when compared to using graph geodesics. We highlight that the graph geodesic approach has the additional advantage of automatically determining the number of nodes in the path, whereas the linearly interpolated approach requires this to be specified (we set it equal to the length of the geodesic path).

Using Clinical Image Features. In row 1.8 (*clinical images*) we use only the clinical image (i.e., $\alpha = 0$ in Eq. 3) and notice a marked decrease in the quality of the paths when compared to dermoscopic images. This is expected since dermoscopic images are more standardized and focused on the lesion, while clinical images have a non-standard field of view and can capture background artifacts.

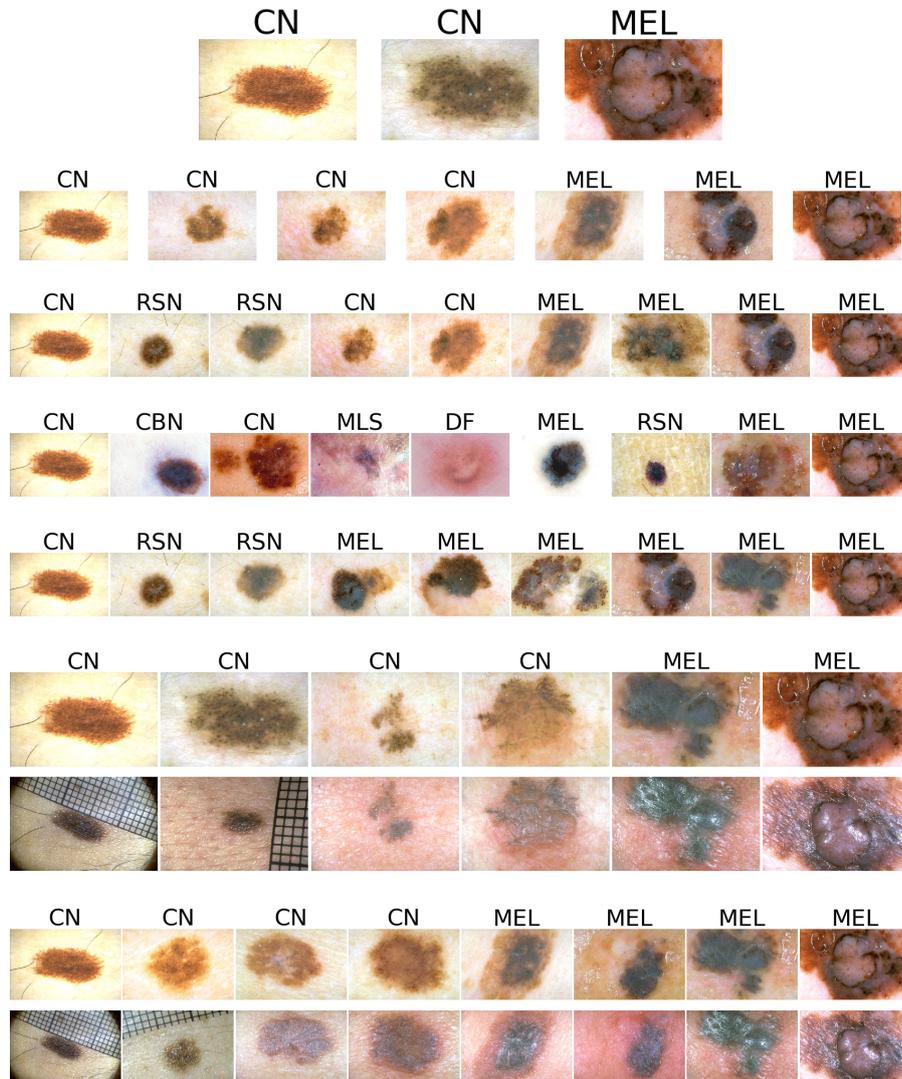


Fig. 3: Visualizing Paths. The *leftmost* and *rightmost* dermoscopic images are the given source (clark nevus) and target (melanoma) node, where the images in each row in between them correspond to the computed geodesic/minimal path. Each row, starting from the *top* to *bottom* row, correspond to the following experiments in Table 1: 1.2 (*non-complete graph with $p = 1$*), 1.4 (*non-complete graph with $p = 4$*), 1.5 (*augmented images*), 1.6 (*random paths*), 1.7 (*linear paths*), 1.9 (*dermoscopic and clinical images*), and 1.10 (*full approach*). The geodesic of Experiments 1.9 and 1.10 incorporates clinical images, shown directly below the dermoscopic images.

Combining Dermoscopic and Clinical Image Features. In **row 1.9** (*dermoscopic and clinical images*) we include both the clinical and dermoscopic images, weighting the dermoscopic images higher (i.e., $\alpha = 0.8$ in Eq. 3) as the dermoscopic images better capture the salient lesion features and avoid irrelevant background artifacts. The returned paths now respect both imaging modalities, yielding improvements to the quality of the paths, most noticeable with transition costs. Finally, in **row 1.10** (*full approach*) we show the full proposed approach, which uses augmented images from both modalities with the dissimilarity measure raised to the power of $p = 4$ on a non-complete graph. While the path quality measures remain similar to the previous experiment, the total path length increases.

4 Conclusions

We proposed a method to visualize a smooth progression of similar skin lesion images between two skin lesions. Our graph geodesic based approach applies an exponential dissimilarity function and considers information from multiple modalities (clinical and dermoscopic images) to form the graph edges, leading to longer paths of higher quality. We proposed surrogate measures of path quality based on the diagnostic labels of the skin lesions to quantitatively assess the resulting paths. Future work would explore how to improve the feature vectors that represent the skin images (e.g., fine-tuning the CNN over a skin dataset), and examine how to make the progression quality measure less sensitive to the length of the path.

Acknowledgments. Thanks to the Natural Sciences and Engineering Research Council (NSERC) of Canada for funding and to the NVIDIA Corporation for the donation of a Titan X GPU used in this research. Thanks to Sara Daneshvar for preparing the data used in this work.

References

1. Argenziano, G., Fabbrocini, G., Carli, P., Vincenzo, D.G., Sammarco, E., Delfino, M.: Arch.Dermatol. (12), 1563–1570
2. Argenziano, G., Soyer, H.P., Giorgio, V.D., Piccolo, D., Carli, P., Delfino, M., Ferrari, A., Hofmann-Wellenhof, R., Massi, D., Mazzocchetti, G., Scal-venzi, M., Wolf, I.H.: Interactive atlas of dermoscopy: a tutorial (Book and CD-ROM) (2000)
3. Bunte, K., Biehl, M., Jonkman, M.F., Petkov, N.: Learning effective color features for content based image retrieval in dermatology. Pattern Recognition 44(9), 1892–1902 (2011)
4. Duffy, K., Grossman, D.: The dysplastic nevus: From historical perspective to management in the modern era: Part I. Historical, histologic, and clinical aspects. Journal of the American Academy of Dermatology 67(1), 1–27 (2012)
5. Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S.: Dermatologist-level classification of skin cancer with deep neural networks. Nature 542(7639), 115–118 (2017)

6. Hay, R.J., Johns, N.E., Williams, H.C., Bolliger, I.W., Dellavalle, R.P., Margolis, D.J., Marks, R., Naldi, L., Weinstock, M.a., Wulf, S.K., Michaud, C., J L Murray, C., Naghavi, M.: The Global Burden of Skin Disease in 2010: An Analysis of the Prevalence and Impact of Skin Conditions. *The Journal of Investigative Dermatology* 134, 1527–1534 (2014)
7. Hegde, C., Sankaranarayanan, A.C., Baraniuk, R.G.: Learning Manifolds in the Wild. *Journal of Machine Learning Research* 5037 (2012)
8. Jia, H., Wu, G., Wang, Q., Wang, Y., Kim, M., Shen, D.: Directed Graph Based Image Registration. *MLMI MICCAI Workshop* 7009, 175–183 (2011)
9. Kawahara, J., BenTaieb, A., Hamarneh, G.: Deep features to classify skin lesions. In: *IEEE ISBI*. pp. 1397–1400 (2016)
10. Kawahara, J., Hamarneh, G.: Image Content-Based Navigation of Skin Conditions. In: *World Congress of Dermatology* (2015)
11. Klingemann, M., Doury, S.: X Degrees of Separation (2016), <https://artsexperiments.withgoogle.com/xdegrees/>
12. Kogan, G.: Shortest path between images (2017), <https://github.com/ml4a/ml4a-guides/blob/master/notebooks/image-path.ipynb>
13. Korotkov, K., Garcia, R.: Computerized analysis of pigmented skin lesions: A review. *Artificial Intelligence in Medicine* 56(2), 69–90 (2012)
14. Markovic, S., Erickson, L.A., Rao, R., Creagan, E.T., et al.: Malignant Melanoma in the 21st Century, Part 1: Epidemiology, Risk Factors, Screening, Prevention, and Diagnosis. *Mayo Clinic Proceedings* 82(3), 364–380 (2007)
15. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L.: ImageNet large scale visual recognition challenge. *International Journal of Computer Vision (IJCV)* 115(3), 211–252 (2015)
16. Schofield, J.K., Fleming, D., Grindlay, D., Williams, H.: Skin conditions are the commonest new reason people present to general practitioners in England and Wales. *British Journal of Dermatology* 165, 1044–1050 (2011)
17. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. *Intl. Conf. on Learning Representations (ICLR)* (2015)