OASIS: Generating Synthetic Skin Artifacts

Elena Sizikova*, Niloufar Saharkhiz*, Jana G. Delfino, Aldo Badano
Office of Science and Engineering Laboratories
Center for Devices and Radiological Health
U.S. Food and Drug Administration
Silver Spring, MD 20993 USA
∗ - These authors contributed equally to this work.
elena.sizikova@fda.hhs.gov

Abstract

Artificial intelligence (AI) techniques have gained significant interest in recent years for automating many dermoscopic image analysis tasks. The presence of artifacts in patient datasets, either inherent to skin images (e.g., hair or blood vessels) or those introduced during image acquisition (e.g., calibration charts, rulers, dark frames) has challenged the application of these techniques since artifacts can lead to decreased performance. Furthermore, data-driven methods used to synthesize data to address patient data limitations are themselves affected by these artifacts. We present OASIS, a knowledge-based, tunable simulator for generating synthetic skin images with artifacts in the object domain prior to imaging. We describe how to simulate commonly occurring clinical artifacts, including calibration charts, rulers, dark frames, hair, and blood vessels. We then demonstrate that OASIS can be used to generate controllable, artifact-present images with pixel-level mask annotations and artifact-free counterparts, and show how the resulting images can be used for augmenting limited patient datasets in downstream tasks.

1. Introduction

In recent years, dermatologic artificial intelligence (AI) applications for the automatic analysis of skin images have emerged [37]. Skin lesion segmentation, for instance, is used to classify lesions according to their shape, e.g., using the Asymmetry, Border, Color, Diameter (ABCD) rule [21], or as a pre-processing step for automatic lesion classification or feature extraction. In this work, we present an approach to generate synthetic skin images with common artifacts (calibration charts, rules, dark frames, hair, and blood vessels) that negatively affect automatic skin lesion segmentation. We analyze how the resulting synthetic examples can be used together or in lieu of patient image datasets, which are scarce and may lack appropriate annotations.

Skin cancer is the fifth most common cancer in the United States (US) [42] and seventeenth world-wide (excluding non-melanoma) [14]. The survival rate for individuals with melanoma depends on the stage of the disease at the time of diagnosis, with the 5-year survival rate reported to be approximately 30% at late stages [22]. In particular, early-stage diagnosis of melanoma relies on the analysis of morphological characteristics of the lesion's shape and the associated segmentation of the lesion boundaries, which can be subjective and impacted by the clinician's expertise. One common limiting factor for automatic lesion segmentation is the unavailability of publicly available large-scale skin image datasets for model training that encompass representative populations and contain enough examples with fine-grained annotations. Existing public datasets also frequently contain images that are not directly intended for AI segmentation analysis, but rather as secondary tools. As a result, associated markings or "artificial" artifacts, such as rulers, calibration charts or dark frames are often present. Hair and blood vessels are other frequent features or "natural" artifacts on those images. Several studies have shown the complication of lesion segmentation due to the presence of these artifacts [16, 34, 36]. Shifts in correlation between invariant and spurious features in the test dataset bias the models toward detecting hair and rulers instead of melanomas [12]. Winkler et al. showed that surgical markings or scale bars at the periphery of the dermoscopic images increase the false-positive rate for melanoma detection and are significantly associated with changes to malignancy scores provided by the deep learning network [49]. Artifacts also complicate the synthesis of dermoscopic images using data-driven generative editing methods, such as diffusion models (DDPM) and generative adversarial networks (GANs) because these models do not understand artifact concepts, and training generative models on limited

artifact-present patient images risks replicating dataset biases or imaging distortions. On the other hand, simulation models offer a controlled way of generating paired examples (artifact vs. clean image) for training generative editing models previously unknown concepts about skin artifacts.

In this work, we introduce OASIS, a synthetic skin simulator with the ability to simulate five types of commonly occurring artifacts in skin imaging: calibration chart, ruler, dark frame, hair, and blood vessels (see Fig. 1 and Fig. 2 for examples). This simulator is open-source and extends the S-SYNTH pipeline [29] to include skin artifacts, creating paired datasets of images, segmentation masks (of lesions and artifacts) for downstream training of generative image editing models. This last application, unique to our work, is important to address domain gaps and broaden applicable use cases.

Simulating artifacts enables controlled generation images with realistic artifact variations. We found that our proposed simulator is able to simulate common skin artifacts to create training datasets needed to adapt generative models (e.g., InstructPix2Pix [15]) to skin image editing tasks. Using 48,642 synthetic images, we analyzed the effect of training data augmentation in comparison to training only with patient datasets for the task of skin lesion segmentation. We demonstrate that using synthetic skin images with artifacts not only improves the overall segmentation performance on patient datasets, but also enhances the segmentation of skin lesions in subgroups with darker skin tones. Our contributions are:

- We introduce OASIS[1], an open-source, object-domain synthetic skin artifact simulator capable of simulating five types of common skin artifacts: calibration chart, ruler, dark frame, hair, and blood vessels.
- We conduct an in-depth analysis of the proposed artifact simulator with a publicly available skin imaging dataset, ISIC18 [19], and demonstrate that the addition of synthetic data with artifacts in training may improve performance in patient data, particularly in examples that contain artifacts and in patients with dark skin.
- We compare OASIS images and simulated artifacts with other artifact addition methods and show that OASIS examples more realistically capture the artifact insertion process and may be used to compare artifact inpainting methods.
- We show that OASIS images and masks are a rich source of data for training image editing pipelines for adding and removing artifacts. Specifically, we show that datasets from OASIS can be used to finetune diffusion pipelines to teach them previously unknown concepts about dermatologic imaging.

[1]Code and supporting data are available at: https://github.com/DIDSR/ssynth-release/tree/OASIS.

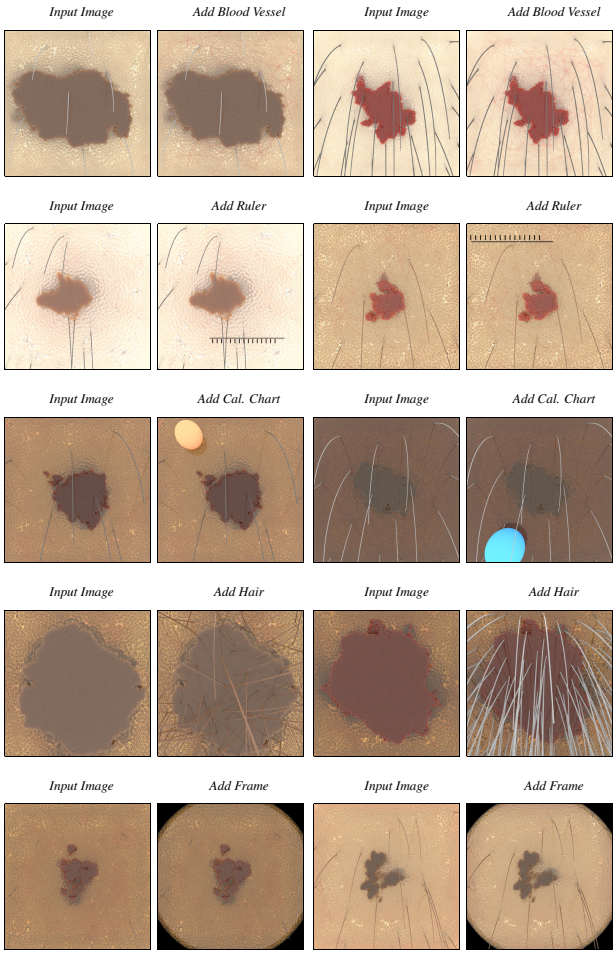


Figure 1. Examples of artifacts that can be generated using the OASIS pipeline, which allows paired generation of images with updates properties. This allows for generating training datasets for finetuning generative models to desired artifact properties independent of limited patient datasets with artifacts.

2. Related Work

Artifacts (e.g., dark corners, hair) are commonly found in skin images and often occlude lesion boundary and texture details. There has been active research on hair removal algorithms using traditional methods such as linear interpolation, inpainting by nonlinear partial differential equation (PDE), exemplar-based repairing, and fast marching [3]. Nevertheless, these techniques encounter limitations associated with thin and overlapping hairs and hairs of similar color than that of underlying skin [32]. A number of hand-crafted [27, 32] and data-driven techniques for hair removal in dermoscopy images have been proposed [8, 11, 30, 33, 47].

One limitation of data-driven hair removal in dermatology images is the absence of an appropriate training dataset:

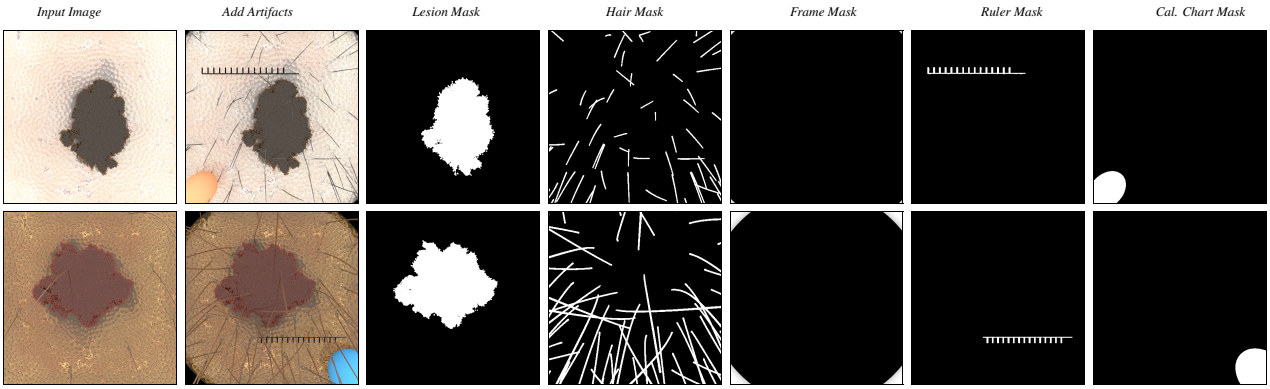


Figure 2. OASIS artifacts may be combined. For lesions, hair, frame, ruler and calibration charts, we generate pixel-level segmentation masks by manipulating rendering materials.

creating hair-free images manually, as shown in [11], is impractical for large datasets. If paired images (with and without hair) are needed, hair simulation [47] can add artificial hair to patient dermatology images, but such techniques may suffer from lack of realism since the added hair objects do not interact with the same light sources as the original skin area. Furthermore, the inherent skin tone imbalances and other biases of the original datasets would be propagated into the adjusted dataset, compromising its application to real-world scenarios [8]. Finally, poor quality of the reconstructed images is another challenge that hinders the use of these techniques, as noted in [30].

The dark corner artifact (DCA), another common artifact characterized by the presence of a black frame, is caused by the circular shape of the dermoscopic lens when pressed against the skin or by a non-contact dermoscope for a lesion located on non-flat skin [39]. There are variations across DCA presence, intensity and shape and some DCA may have the same level of luminescence intensity as nevi and may interfere with lesion border detection [43, 45]. However, as the dark corner artifact shape and intensity depends on the magnification of the camera, existing algorithms such as [4, 38, 45, 46] may not work well for all inputs.

The removal of artifacts including hair, ruler and dark corners may also help improve AI performance. For example, Pewton and Yap [38] removed the dark corner artifact by extracting the largest contours, reducing the dark corners' intensity, followed by inpainting the dark corner region, and found that removing the artifact shifted the network activations to the skin lesions. In an effort to provide ground truth annotated data for these algorithms or benchmark datasets to compare their performance, some studies annotated patient images with artifact masks [23, 33], however, this involved a laborious labeling process. Another line of work focused on simulating artifacts and adding them to the artifact-free skin images. For instance, artificial hair was generated using modelling [35, 41] or with image-to-image translation techniques [5]. In Skin_Hair dataset [24], 252 patient images were superimposed with artificially added hairs to create hair-free, hair-present and hair mask triplets. The authors of [24] also highlighted the need for similar efforts for other types of artifacts. Finally, DERMAHAIR [9] is a recent dataset consisting of 700 patient images with artificially superimposed hair used to train and evaluate a data-data driven hair removal model. [28] provided a tool to add three types of artifacts (bubbles, rulers and ink markings) to skin images, where artifacts are extracted from dermoscopic images manually using Adobe Photoshop [28]. In [18], the authors used color labeled generative style transfer network (CL-GSTN) to generate synthetic dermoscopic images with color patch artifacts. They reported that adding synthetic images to a subset of patient images improved the Dice score by up to 1%. However, their approach performed worse than when the full set of patient images was used for training of the segmentation model. A related research direction has looked at debiasing and addressing the limitations of skin imaging datasets, which are often small, biased, or do not contain the annotations (e.g., segmentation masks, patient or acquisition device characteristic labels) needed to carry out tasks of interest. Bissoto et al. [12] discussed the challenges of artifact and other debiasing methods for skin lesion classification and showed that more work is needed in this field. In Table 1, we provide an overview of existing datasets for skin artifact analysis. To our knowledge, OASIS is the first fully synthetic and largest source of examples with five types of artifacts available.

Another potential way to augment limited patient data is to generate examples using data-driven models. In MaskMedInpaint [25], authors showed that models trained using patient images augmented with synthetic images created using a text-to-image diffusion model (a generative

model trained to in-paint background regions) have improved generalization. Recent image editing (also known as guided image synthesis) approaches rely on data-driven models that receive an input image and a set of instructions and predict the edited image (e.g., InstructPix2Pix [15]). Developed initially for non-medical applications, these approaches may need additional in-domain data (which is already scarce) to generate realistic skin image artifacts. In our experiments, we show that OASIS images, generated using a knowledge-based (KB) method, can be used to train this class of models. For a comparison of KB and data-driven methods, we refer the reader to [44].

Table 1. Existing datasets for dermoscopic image artifact analysis.

Year	Reference	Type of Data	Artifact	#Examples
2022	Pewton [38]	Image-level labels for patient images labelling artifacts (Borders, Hair, Measurement Device, Air Pocket(s) and Clinical Markings); annotations of size of the border/DCA artifact	DCA, hair, ruler, air pocket, clinical marking	9,810
2024	Pewton [39]	Patient skin lesion images with and without dark corner artifacts; synthetically superimposed artifact labels; image level artifact labels for patient images with and without melanoma	dark corner	10,250
2020	Talavera-Martinez [47]	Synthetically added hair to patient datasets for hair inpainting evaluation	hair	618
2021	Digital Hair Dataset [33]	Paired patient images with hair segmentation masks	hair	306
2023	Hossain [23]	Patient images with hair segmentation masks	hair	500
2022	Skin_Hair Dataset [24]	Images and hair masks from artificially added hair to patient images	hair	252
2024	Synthetic Hair Dataset [26]	Images and hair masks from artificially added hair to patient images	hair	1,064
2024	DermaHair [9]	Images and hair masks from artificially added hair to patient images	hair	700
2026	OASIS Dataset (ours)	Simulated images and masks	hair, calibration chart, ruler, frame, blood vessel	48,642

3. Materials and Methods

Image Simulator The OASIS simulator for creating synthetic images containing artifacts is based on S-SYNTH, an open-source tool for simulating dermoscopic images [29]. We simulated a skin tissue model comprised of five distinct components, including the epidermis, dermis, hypodermis, blood network, and hair in Houdini [1]. We used 27 adjustable parameters (see Suppl. Mat.) that represent the characteristics of each skin component to create different skin models. We simulated a 3D growth lesion model based on [40] with stochastic shape and size growth and inserted it into the skin tissue model. We generated 20 different growing lesion models and stored each at 10 different timepoints to allow for diverse lesion sizes and shapes.

Once the skin geometry and lesion models were ready, we imported the models into Mitsuba3 [2] to assign optical materials to each component, configure lighting conditions, and generate synthetic renderings. The imported models were configured to simulate a synthetic skin model with dimensions of 20x20 mm. The optical properties of the skin components, including the melanosome fraction, blood fraction, and those for hair, blood vessels, and lesions were selected from pre-defined ranges similar to the values of real skin tissue as described in [29]. The values for the melanosome fraction, the main factor affecting skin tone, were selected from a wide range of values to represent different skin tones. We used diffuse lighting conditions and a perspective sensor placed 13-15 mm perpendicularly above the epidermis to generate the images. We rendered the synthetic images at a sample per pixel rate of 128 and a resolution of 1024x1024 pixels (Fig. 2). The paired pixel-level segmentation masks were generated by assigning different optical materials to render the epidermis layer as a black surface and the lesion/artifact as a white object (equivalent to pixel values of 0 and 1, respectively).

Artifact Generation We generated five different artifacts specific to dermoscopic images that challenge skin lesion segmentation tasks. Three of the artifacts (calibration chart, ruler, and dark frame) were created by adding a component to the object space using Mitsuba before rendering the images. The other two artifacts (hair and blood vessels) were created by changing the mutable parameters of the skin models in Houdini, and then transferring the new models to Mitsuba to complete the remaining steps (see Suppl. Mat.). Calibration charts (blue or yellow spheres) and rulers (sets of rectangular markers) were randomly positioned near image edges using size and location constraints to avoid obscuring lesions, and were only added when lesions were smaller than a pre-specified size. Dark frames were simulated by placing a black torus below the sensor with variable radius to control coverage, also limited to scenes with smaller lesions. Hair artifacts were created by increasing hair density and scatter parameters. Blood vessel artifacts were enhanced by expanding the vascular network, reducing the dermis skin layer thickness, and limited to the images to lighter skin tones to improve vessel visibility. The design choices for each artifact were based on patient data (and can be configured).

Dataset Simulation In total, we created a synthetic dataset of 48,642 image-mask pairs, consisting of images with artifacts and images without artifacts. For the hair and blood vessel images, we generated 10,000 images each. For the ruler, calibration chart, and frame images, artifacts were rendered only when the lesion was small (this is a design choice based on observations in patient images and can be modified), so 3,988 images were created for each of these artifacts. Finally, we generated 6,678 images with combined (randomly selected artifacts). To generate each image, we randomly selected parameters from pre-defined limits for the models and their optical properties, as well as properties specific to each artifact. These parameters included the variations for each skin model component, lesion

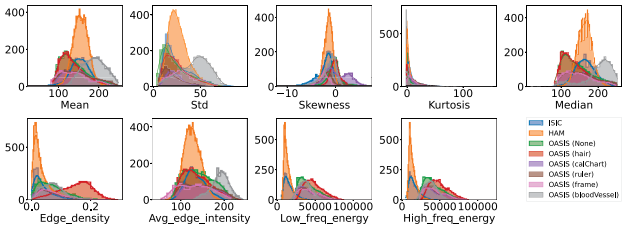


Figure 3. Comparison of features of the OASIS datasets to the ISIC and HAM patient datasets using the SyIDE-Tool synthetic data scorecard [20, 50].

model, timepoint of lesion growth, lesion material, blood fraction, melanosome fraction, hair material, and the distance between the camera sensor and the skin model. The distance was varied to increase lesion size diversity, as well as to vary the coverage of the dark frame for the images with that specific artifact. After a random selection of parameters, an artifact was added to an image to create each image pair. In addition, for every artifact inserted into the image, we generated an artifact pixel-level mask by manipulating the albedo parameters and the visibility of all materials. Using a GPU, each image-mask pair took about 2 min to simulate (at higher quality, which can be reduced as needed) and examples can be generated in parallel on either CPU or GPU.

In Fig. 3, we plot the image feature distributions of OASIS and patient images using SyIDE-Tool synthetic data scorecard [20, 50]. We find that OASIS images (all subsets) generally cover the same range of features as the patient examples, however, they differ in terms of the center of the distribution. In addition, the blood vessel subgroup tends to skew to the right in terms of intensity mean, median, and standard deviation indicating a distribution shift.

The skin tissue, lesion artifact and acquisition models are a simplification and may results in unrealistic images. However, as discussed in [7], the realism of synthetic medical data should "be assessed based on purpose". For example, in this work, we selected parameters to improve segmentation of lesions in artifact-present skin images, focused on darker skin tones, and to demonstrate a proof-of-concept for diffusion finetuning. An alternative set of parameters, focused on addressing lesion, lighting or other considerations (e.g., resolution), may be needed for different tasks.

3.1. Experimental Evaluation

Patient Data We compared the performance of OASIS images to images from the publicly available dataset of ISIC18 [19] images. The same pre-processing and data splits described in [6, 19] were applied, where the image set was divided into (1815, 259, 520) for the (training, validation, test), respectively. We used the individual typology angle (ITA) metric [31] to estimate skin tone from the non-lesion area. We also relied on artifact annotations for ISIC from Bissoto [12]. For the segmentation experiment, we varied the proportion of available artifact-present patient images in the training data, and replaced varying proportions of the training set (composed of patient images) with synthetic images containing a specific artifact. We then evaluated performance on a held-out test set. Replacement allows us to evaluate the contribution of OASIS examples in a data-scarce scenario, while still having access to the full patient set in order to establish a baseline performance.

Diffusion Model Training For image editing, we finetuned an InstructPix2Pix [15] model with paired artifact-present and artifact-absent OASIS data, and a text prompt describing the task (e.g., "add" or "remove") and the properties of the artifact (e.g., "yellow calibration chart"). We used images at 512x512 resolution, with random flip augmentation and batch size of 4. Each model was trained for 15,000 steps starting from the released pre-trained model ("timbrooks/instruct-pix2pix").

Segmentation We studied the impact of augmenting patient datasets with synthetic skin images (with and without artifacts) on patient skin lesion segmentation performance. We relied on DermoSegDiff [13], a boundary-aware diffusion model specifically designed for skin lesion delineation for the segmentation task. The DermoSegDiff model was trained using the "dsd_i01' configuration and default training parameters with `dim_x=16`, `dim_g=8`. Each model was trained five times for 120,000 steps for each experiment, with input images resized to 128x128. We evaluated the impact of data variation and reader variation to better understand the effect of artifact addition to the training data. For data and reader evaluation, we set up a multi-reader multi-case (MRMC) study [17], where each reader is an AI model trained on the same training dataset (with varying initialization). Here, the standard deviation captured variation across both examples and readers (see Suppl. Mat for more information).

4. Results and Discussion

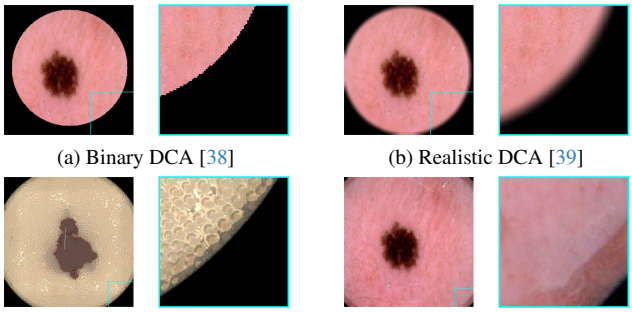
4.1. Artifact Generation Method Comparison

We report a comparison of two common (see [24, 38]) artifact inpainting methods, Telea [48] and Navier-Stokes (NS) [10], using OASIS images as the test case in Tab. 2.

The goal of this experiment is to show that OASIS can be used for controlled evaluation of artifact inpainting methods, and match results from previous work. Similar to [38], we found that the lower the area of the dark corner artifact, the higher the peak signal to noise ratio (PSNR), i.e., better performance. The NS method had smaller standard deviation than Telea method for all artifact sizes. We also

	Tiny	Small	Medium
Telea [48]	32.46 (3.48)	31.02 (3.36)	28.80 (3.21)
NS [10]	32.06 (3.32)	30.70 (3.30)	28.47 (3.15)

Table 2. Artifact inpainting evaluation using OASIS images with varying dark corner artifact area using PSNR (↑). Results are reported using mean (standard deviation) format. OASIS images can be used for parameter-controlled evaluation of artifact inpainting methods such as by artifact size (these methods are not baselines).



(a) Binary DCA [38] (b) Realistic DCA [39]

(c) OASIS (d) OASIS + InstructPix2Pix [15]

Figure 4. Qualitative comparison of artifact addition. OASIS simulated images and InstructPix2Pix [15] finetuned using OASIS offers realistic transitions between foreground and artifacts compared to existing artifact insertion approaches (a and b).

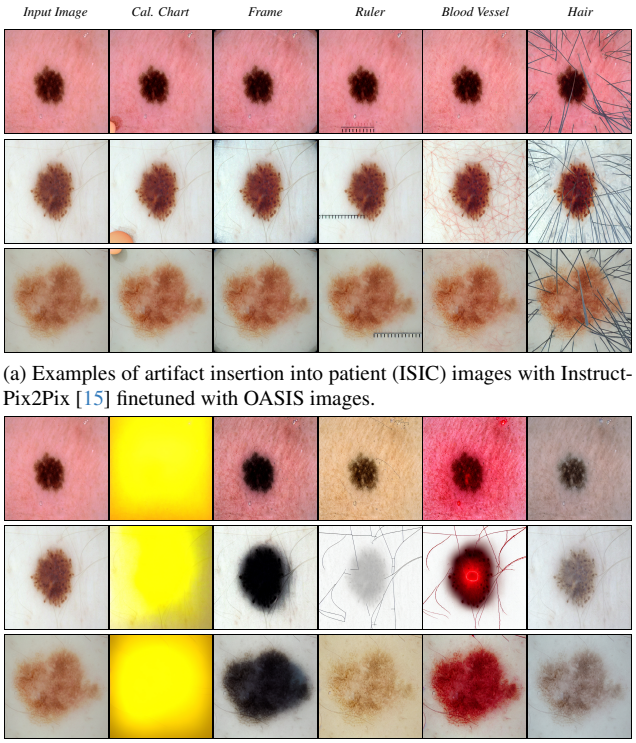
qualitatively compared our approach to existing artifact insertion methods (Binary DCA [38] and Realistic DCA [39]) in Fig. 4. These methods insert artifacts directly in the image space. We found that simulated artifacts in the object space using our method (as well as the finetuned diffusion model) generated smoother and more realistic transitions.

4.2. Generative Image Editing

Artifact Addition We provide some qualitative comparisons of how each type of artifact (calibration chart, frame, ruler, blood vessel, and hair) is inserted into patient images using a diffusion model finetuned with OASIS (see Fig. 5a). We find that artifact insertion generates realistic artifacts with minimal modifications to the patient images. In contrast, if the diffusion model has not seen skin image artifact examples before, we found it was unable to generate the artifact images (see Fig. 5b) since it does not know these domain-specific, rare concepts.

In Fig. 6a and Fig. 6b, we show that by varying the prompt, we can manipulate the characteristics of the inserted artifacts: calibration chart color or hair color.

Artifact Removal We provide sample artifact removal results in Fig. 7. We find that the OASIS-finetuned diffusion model learns well how to remove frame, black border, blood vessel and hair artifacts. The model struggles with the ruler and calibration chart artifacts, presumably because there remains a domain shift between these artifacts in patient and synthetic images.



(a) Examples of artifact insertion into patient (ISIC) images with InstructPix2Pix [15] finetuned with OASIS images.

(b) Examples of artifact insertion into patient (ISIC) images with an unfinetuned InstructPix2Pix [15].

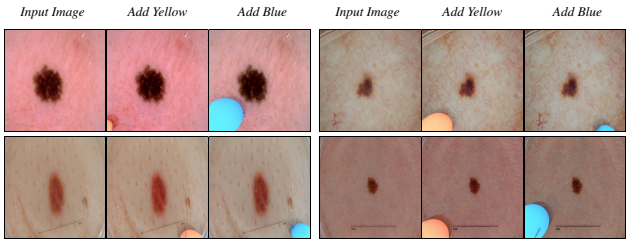
Figure 5. Comparison of artifact insertion results using finetuned vs unfinetuned diffusion models.

4.3. Effect on Lesion Segmentation

In Fig. 8, we report a MRMC analysis of lesion segmentation in the presence of artifacts. We find that OASIS images improve segmentation performance in darker skin tones (dark, tan2, and int1) across all types of artifacts considered. In many instances, the margin of error was also reduced. For the hair and ruler artifacts, we found that addition of OASIS data was more beneficial than the addition of patient data for darker skin tones, presumably, due to more reliable lesion segmentation annotations in this smaller subgroup. We did not observe a significant effect of OASIS images on lighter (lt1, lt2, and very lt) skin tones.

4.4. Limitations

Our work has several limitations. There is a distribution shift between the synthetic and patient images, which reduces the performance of models trained only on synthetic data. Although the shape features of synthetic lesions resemble those in patient skin images, the lesions themselves do not represent any specific diseases. To address this, more complex lesion models need to be simulated. We



(a) Examples of artifact property manipulation (*calibration chart*).

(b) Examples of artifact property manipulation (*hair*).

Figure 6. Artifact manipulation using finetuned diffusion.

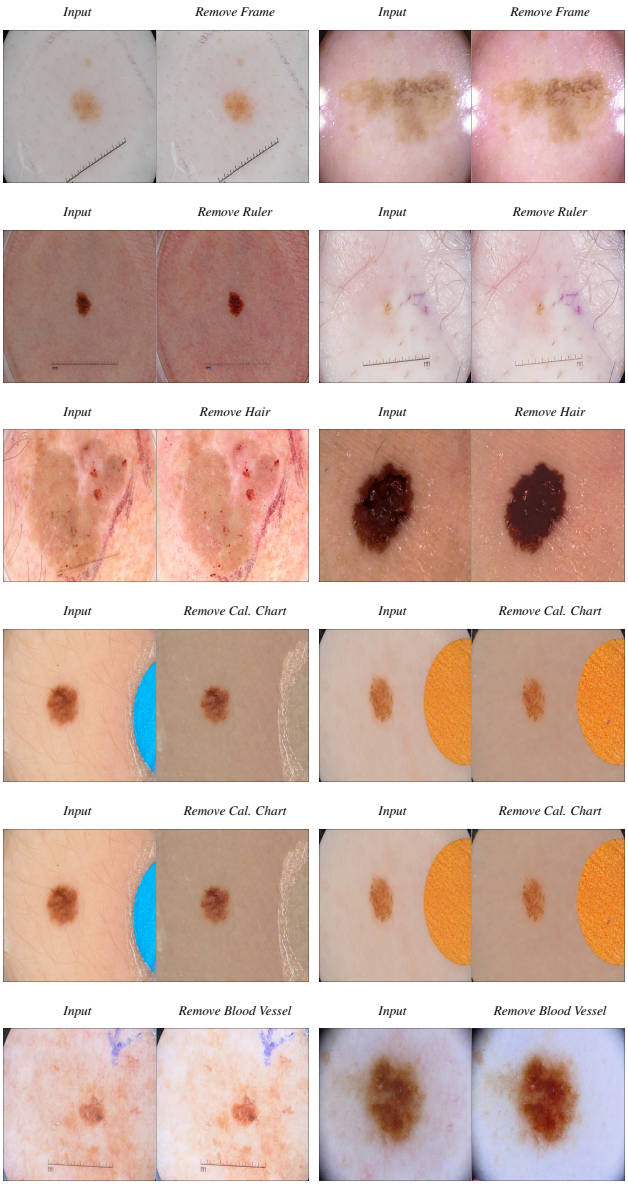


Figure 7. Examples of artifacts removed from patient images.

used a simple, constant diffuse lighting condition. However, this approach can be improved to better simulate the lighting conditions of dermoscopes, which involve specific wavelength ranges and more complex lighting properties. We chose to conduct experiments with this simplified lighting simulation to decouple lighting confounders from the analysis of the artifacts. Our work is intended as a proof-of-concept of artifact simulation within a knowledge-based pipeline and subsequent training data augmentation. Further evaluation with labeled artifact subgroups is needed.

Finally, we specifically chose not to run diagnosis/classification experiments because skin cancer diagnosis typically requires biopsy-confirmed diagnostic confirmation which is not available for synthetic images. We show that synthetic data may help improve performance in auxiliary tasks (such as segmentation).

5. Conclusion

We introduced the OASIS simulator to simulate five common artifacts (dark frames, calibration charts, ruler, hair, and blood vessels) observed in skin images that are known confounders for AI segmentation. We show that OASIS can be used to generate paired (artifact-free/with artifact) image datasets with controllable annotations for finetuning diffusion models and teaching them concepts about skin image artifacts. The resulting models can be used to add or remove artifacts. We quantitatively evaluated the resulting artifact augmented synthetic images in the task of skin image segmentation and found that synthetic data may improve performance in low data regimes common for darker skin tones. We hope that our work will inspire more research into combining knowledge-based models and generative techniques to improve task-specific synthetic data generation.

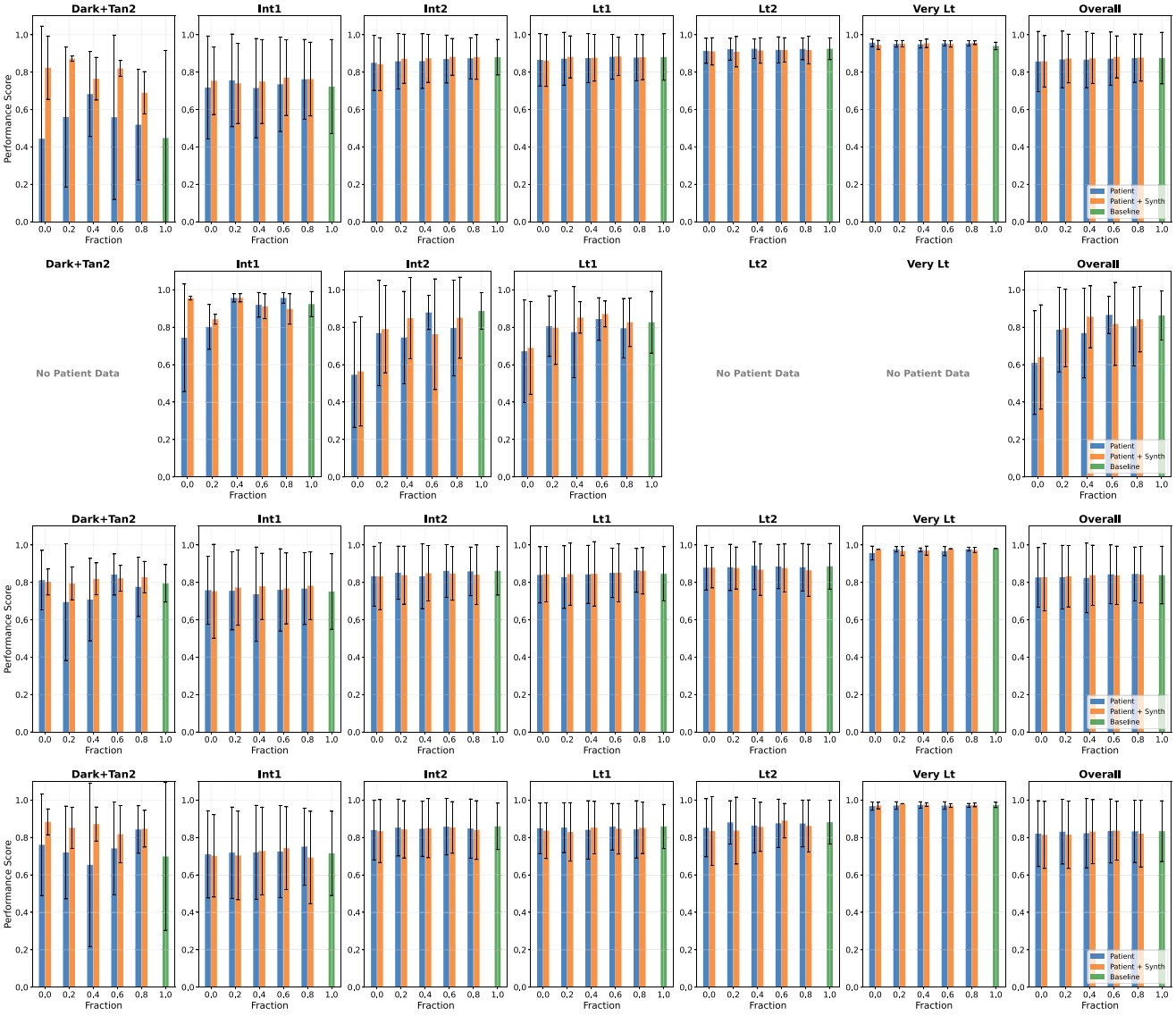


Figure 8. Multi Reader Multi Case (MRMC) evaluation of Dice segmentation performance (↑) when artifact-present training examples in the ISIC [19] are replaced with OASIS images. Results are reported with mean and standard deviation. We observe that at limited data regimes, the model trained with synthetic examples containing artifacts performs better for darker skin tones (tan2, int1). In light skin tones, all methods perform similarly. Row 1: hair, Row 2: Cal. Chart, Row 3: frame, Row 4: ruler. Blue: partial patient data (with only a fraction of artifacts present), orange: patient data supplemented with synthetic OASIS examples, and green: full patient data baseline.

6. Acknowledgments

We thank Kenny Cha, Mike Mikailov and the OpenHPC team for providing help with experiments, all colleagues in OSEL/CDRH/FDA for fruitful discussions on the project, and anonymous reviewers for providing feedback.

7. Disclaimer

This article reflects the views of the authors and does not represent the views or policy of the U.S. Food and Drug Administration, the Department of Health and Human Services, or the U.S. Government. The mention of commercial products, their sources, or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products by the Department of Health and Human Services.

References

[1] SideFX: Houdini, 2022. 4

[2] Mitsuba 3 renderer, 2022. 4

[3] Qaisar Abbas, M Emre Celebi, and Irene Fondón García. Hair removal methods: A comparative study for dermoscopy images. *Biomedical Signal Processing and Control*, 2011. 2

[4] RP Aneesh and Joseph Zacharias. Semantic segmentation in skin surface microscopic images with artifacts removal. *Computers in Biology and Medicine*, 2024. 3

[5] Mohamed Attia, Mohammed Hossny, Hailing Zhou, Saeid Nahavandi, Hamed Asadi, and Anousha Yazdabadi. Realistic hair simulator for skin lesion images: A novel benchemarking tool. *Artificial Intelligence in Medicine*, 2020. 3

[6] Reza Azad, Mohammad T Al-Antary, Moein Heidari, and Dorit Merhof. Transnorm: Transformer provides a strong spatial normalization mechanism for a deep segmentation model. *IEEE Access*, 2022. 5

[7] Aldo Badano. "how much realism is needed?"—the wrong question in silico imagers have been asking. *Medical Physics*, 44(5):1607–1609, 2017. 5

[8] Dalal Bardou, Laishui Lv, Yasmine Medjadba, Ting Zhang, Ouahiba Chouhal, Mourad Bounezra, Karima Saidi, Youcef Bezza, and Aissa Snani. GAD-VAE: generative adversarial disentanglement with variational autoencoders for hair removal in dermoscopy images. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 2024. 2, 3

[9] Bilel Benjdira, Anas M. Ali, Anis Koubaa, Adel Ammar, and Wadii Boulila. DM-AHR: A self-supervised conditional diffusion model for AI-generated hairless imaging for enhanced skin diagnosis applications. *Cancers*, 2024. 3, 4

[10] Marcelo Bertalmio, Andrea L Bertozzi, and Guillermo Sapiro. Navier-stokes, fluid dynamics, and image and video inpainting. In *CVPR*. IEEE, 2001. 5, 6

[11] Devansh Bisla, Anna Choromanska, Russell S Berman, Jennifer A Stein, and David Polsky. Towards automated melanoma detection with deep learning: data purification and augmentation. In *CVPR Workshops*, 2019. 2, 3

[12] Alceu Bissoto, Eduardo Valle, and Sandra Avila. Debiasing skin lesion datasets and models? not so fast. In *CVPR Workshops*, 2020. 1, 3, 5

[13] Afshin Bozorgpour, Yousef Sadegheih, Amirhossein Kazerouni, Reza Azad, and Dorit Merhof. Dermosegdiff: A boundary-aware segmentation diffusion model for skin lesion delineation. In *PRIME*. Springer, 2023. 5

[14] Freddie Bray, Mathieu Laversanne, Hyuna Sung, Jacques Ferlay, Rebecca L Siegel, Isabelle Soerjomataram, and Ahmedin Jemal. Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 2024. 1

[15] Tim Brooks, Aleksander Holynski, and Alexei A Efros. Instructpix2pix: Learning to follow image editing instructions. In *CVPR*, 2023. 2, 4, 5, 6

[16] Lili Cai, Keke Hou, and Su Zhou. Intelligent skin lesion segmentation using deformable attention transformer U-Net with bidirectional attention mechanism in skin cancer images. *Skin Research and Technology*, 2024. 1

[17] RST Catalog. iMRMC: Software for the Statistical Analysis of multi-reader multi-case studies, 2022. 5

[18] Yucong Chi, Lei Bi, Jinman Kim, Dagan Feng, and Ashnil Kumar. Controlled synthesis of dermoscopic images via a new color labeled generative style transfer network to enhance melanoma segmentation. In *EMBC*. IEEE, 2018. 3

[19] Noel Codella, Veronica Rotemberg, Philipp Tschandl, M Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kalloo, Konstantinos Liopyris, Michael Marchetti, et al. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC). *arXiv*, 2019. 2, 5, 8

[20] DIDSR. SyIDE-Tools: A metric library for assessing congruence and coverage of synthetic medical imaging data. https://github.com/DIDSR/SyIDE-Tools, 2025. Accessed: 2025-10-19. 5

[21] Robert J Friedman, Darrell S Rigel, and Alfred W Kopf. Early detection of malignant melanoma: the role of physician examination and self-examination of the skin. *CA: a cancer journal for clinicians*, 1985. 1

[22] Jeffrey E et al. Gershenwald. Melanoma staging: evidence-based changes in the american joint committee on cancer eighth edition cancer staging manual. *CA: a cancer journal for clinicians*, 2017. 1

[23] Sk Imran Hossain, Sudipta Singha Roy, Jocelyn De Goër De Herve, Robert E Mercer, and Engelbert Mephu Nguifo. A skin lesion hair mask dataset with fine-grained annotations. *Data in Brief*, 2023. 3, 4

[24] Joanna Jaworek-Korjakowska, Anna Wojcicka, Dariusz Kucharski, Andrzej Brodzicki, Connah Kendrick, Bill Cassidy, and Moi Hoon Yap. Skin_hair dataset: Setting the benchmark for effective hair inpainting methods for improving the image quality of dermoscopic images. In *ECCV*. Springer, 2022. 3, 4, 5

[25] Qixuan Jin, Walter Gerych, and Marzyeh Ghassemi. MaskMedPaint: Masked medical image inpainting with diffusion models for mitigation of spurious correlations. *arXiv preprint arXiv:2411.10686*, 2024. 3

[26] Lennart Jütte, Harshkumar Patel, and Bernhard Roth. Advancing dermoscopy through a synthetic hair benchmark dataset and deep learning-based hair removal. *Journal of biomedical optics*, 2024. 4

[27] Reda Kasmi, Jason Hagerty, Reagan Young, Norsang Lama, Januka Nepal, Jessica Miinch, William Stoecker, and R Joe Stanley. Sharprazor: Automatic removal of hair and ruler marks from dermoscopy images. *Skin Research and Technology*, 2023. 2

[28] Florian Katsch, Christoph Rinner, and Philipp Tschandl. Comparison of convolutional neural network architectures for robustness against common artefacts in dermatoscopic images. *Dermatology Practical & Conceptual*, 2022. 3

[29] Andrea Kim, Niloufar Saharkhiz, Elena Sizikova, Miguel Lago, Berkman Sahiner, Jana Delfino, and Aldo Badano. S-SYNTH: Knowledge-based, synthetic generation of skin images. In *MICCAI*. Springer, 2024. 2, 4

[30] Dahye Kim and Byung-Woo Hong. Unsupervised feature elimination via generative adversarial networks: application to hair removal in melanoma classification. *IEEE Access*, 2021. 2, 3

[31] Newton M Kinyanjui, Timothy Odonga, Celia Cintas, Noel CF Codella, Rameswar Panda, Prasanna Sattigeri, and Kush R Varshney. Estimating skin tone and effects on classification performance in dermatology datasets. *arXiv*, 2019. 5

[32] Ian Lee, Xian Du, and Brian Anthony. Hair segmentation using adaptive threshold from edge and branch length measures. *Computers in biology and medicine*, 2017. 2

[33] Wei Li, Alex Noel Joseph Raj, Tardi Tjahjadi, and Zhemin Zhuang. Digital hair removal by deep learning for skin lesion segmentation. *Pattern Recognition*, 2021. 2, 3, 4

[34] Zahra Mirikharaji, Kumar Abhishek, Alceu Bissoto, Catarina Barata, Sandra Avila, Eduardo Valle, M Emre Celebi, and Ghassan Hamarneh. A survey on deep learning for skin lesion segmentation. *Medical Image Analysis*, 2023. 1

[35] Hengameh Mirzaalian, Tim K Lee, and Ghassan Hamarneh. Hair enhancement in dermoscopic images using dual-channel quaternion tubularness filters and mrf-based multi-label optimization. *TIP*, 2014. 3

[36] Nabin K Mishra and M Emre Celebi. An overview of melanoma detection in dermoscopy images using image processing and machine learning. *arXiv*, 2016. 1

[37] Raj H Patel, Emilie A Foltz, Alexander Witkowski, and Joanna Ludzik. Analysis of artificial intelligence-based approaches applied to non-invasive imaging for early detection of melanoma: a asystematic review. *Cancers*, 2023. 1

[38] Samuel William Pewton and Moi Hoon Yap. Dark corner on skin lesion image dataset: Does it matter? In *CVPR*, 2022. 3, 4, 5, 6

[39] Samuel William Pewton, Bill Cassidy, Connah Kendrick, and Moi Hoon Yap. Dermoscopic dark corner artifacts removal: Friend or foe? *Computer Methods and Programs in Biomedicine*, 2024. 3, 4, 6

[40] Aunnasha Sengupta, Diksha Sharma, and Aldo Badano. Computational model of tumor growth for in silico trials. In *Medical Imaging 2021: Physics of Medical Imaging*. SPIE, 2021. 4

[41] Zhishun She, AWG Duller, Y Liu, and PJ Fish. Simulation and analysis of optical skin lesion images. *Skin Research and Technology*, 2006. 3

[42] Rebecca L Siegel, Angela N Giaquinto, and Ahmedin Jemal. Cancer statistics, 2024. *CA: a cancer journal for clinicians*, 2024. 1

[43] Katharina Sies, Julia K Winkler, Christine Fink, Felicitas Bardehle, Ferdinand Toberer, Felix KF Kommoss, Timo Buhl, Alexander Enk, Albert Rosenberger, and Holger A Haenssle. Dark corner artefact and diagnostic performance of a market-approved neural network for skin cancer classification. *JDDG*, 2021. 3

[44] Elena Sizikova, Andreu Badal, Jana G Delfino, Miguel Lago, Brandon Nelson, Niloufar Saharkhiz, Berkman Sahiner, Ghada Zamzmi, and Aldo Badano. Synthetic data in radiological imaging: current state and future outlook. *BJR—Artificial Intelligence*, 2024. 4

[45] Alina Sultana, Ioana Dumitrache, Mihai Vocurek, and Mihai Ciuc. Removal of artifacts from dermatoscopic images. In *COMM*. IEEE, 2014. 3

[46] Neda Zamani Tajeddin and Babak Mohammadzadeh Asl. A general algorithm for automatic lesion segmentation in dermoscopy images. In *ICBME*. IEEE, 2016. 3

[47] Lidia Talavera-Martinez, Pedro Bibiloni, and Manuel Gonzalez-Hidalgo. Hair segmentation and removal in dermoscopic images using deep learning. *IEEE Access*, 2020. 2, 3, 4

[48] Alexandru Telea. An image inpainting technique based on the fast marching method. *Journal of graphics tools*, 2004. 5, 6

[49] Julia K Winkler, Christine Fink, Ferdinand Toberer, Alexander Enk, Teresa Deinlein, Rainer Hofmann-Wellenhof, Luc Thomas, Aimilios Lallas, Andreas Blum, Wilhelm Stolz, et al. Association between surgical skin markings in dermoscopic images and diagnostic performance of a deep learning convolutional neural network for melanoma recognition, 2019. 1

[50] Ghada Zamzmi, Adarsh Subbaswamy, Elena Sizikova, Edward Margerrison, Jana G Delfino, and Aldo Badano. Scorecard for synthetic medical data evaluation. *Communications Engineering*, 2025. 5