

Chapter 1

Introduction

1.1 Background and Problem Statement

In recent years, the way people obtain dermatological advice has changed significantly, mainly due to the COVID-19 pandemic. Teledermatology, a branch of telemedicine, has become increasingly popular as a method to diagnose and manage skin conditions remotely. Telemedicine involves the use of telecommunications technology to provide healthcare services from a distance, allowing patients to consult with healthcare providers without needing to be physically present.

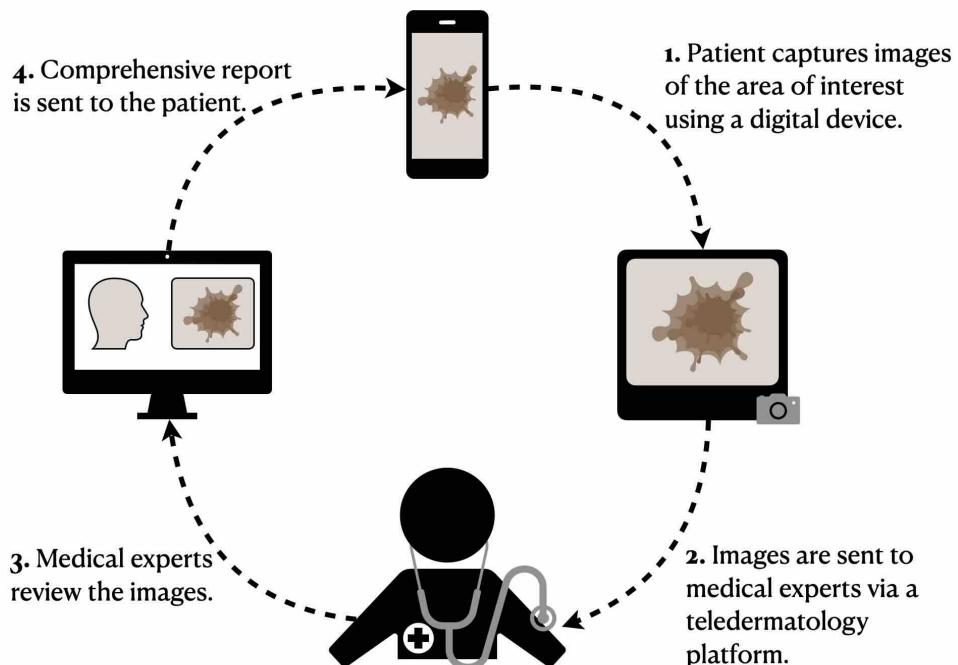


Figure 1.1: This diagram illustrates the streamlined process of a teledermatology consultation, starting from the patient capturing images of their skin condition to the receipt of a detailed medical report. Each step highlights the essential role of image quality in ensuring accurate diagnosis and effective patient care. (Created by the Author)

In teledermatology, patients use mobile applications to take pictures of their skin conditions with everyday devices like smartphones and tablets. These images are then sent to dermatologists for analysis, eliminating the need for face-to-face appointments. Figure 1.1 illustrates the streamlined process of a teledermatology consultation, starting from the patient capturing images of their skin condition to the receipt of a detailed medical report. Each step highlights the essential role of image quality in ensuring accurate diagnosis and effective patient care.

However, the success of teledermatology depends heavily on the quality of the images patients capture. Despite the convenience of modern technology, many images sent by patients do not meet the required standards. Issues such as poor lighting, blurred images, and inadequate representation of skin conditions can greatly hinder a dermatologist's ability to make accurate diagnoses. These challenges with image quality reduce the effectiveness of teledermatology.

This common problem highlights the critical need to improve the quality of images taken through mobile applications. This thesis aims to address this problem by developing and implementing automated image quality assessment techniques to enhance the reliability and effectiveness of teledermatology.

1.2 Objectives of the Thesis

The primary goal of this thesis is to develop and evaluate automated methods for assessing image quality within the context of teledermatology. The objectives are varied, starting with a comprehensive literature review of image quality assessment methods from the general imaging domain to determine their suitability for teledermatology applications. This thesis also aims to select appropriate quality metrics, apply these methods to relevant dermatological datasets, and create a reproducible repository for future research.

The specific objectives of this thesis are detailed as follows:

- An extensive review of the literature on image quality assessment (IQA) methods, focusing on their application in teledermatology.
- Identifying and selecting image quality metrics that are most suitable for assessing the quality of dermatological images.
- Evaluate the performance of selected image quality metrics on dermatological datasets to determine their effectiveness in assessing image quality.
- Develop a reproducible repository of image quality assessment tools and methodologies for teledermatology applications.

Achieving these objectives is expected to significantly enhance the efficiency and accuracy of teledermatology services by establishing a standardized approach to image quality assessment. This improvement is anticipated to streamline workflow, save time, and reduce frustration in teledermatology, providing effective tools and methodologies for evaluating the quality of patient images remotely. Ultimately, the advancements from this research will contribute to better diagnostic precision and overall patient care in remote dermatological consultations.

Reformulate this paragraph

1.3 Organisation of this Thesis

This thesis is structured into six chapters to provide a clear and systematic exploration of image quality assessment in teledermatology. Chapter 2 covers the literature review, discussing

previous and related works on image quality assessment (IQA) and teledermatology. Chapter 3 details the methodologies, including those used in the literature review and those specific to IQA and teledermatology. In Chapter 4, the experiments conducted are described, offering insights into the metrics used. Chapter 5 presents the results of these investigations. Finally, Chapter 6 concludes the thesis, summarizing the findings and suggesting directions for future research.

All figures and tables in this thesis are created by the author unless otherwise referenced. If any code is referenced, the path or module is provided in the footnotes.

Chapter 2

Literature Review

2.1 Image Evaluation

There are three ways to evaluate an image: assessing its quality, aesthetics, or fidelity. Each method focuses on different aspects of image evaluation and has unique applications.

Image Quality Assessment (IQA) measures the degradation of an image. This involves comparing an original, undistorted image with a processed version that has undergone changes such as compression, noise addition, or artifact introduction. The goal is to quantify how much the image quality has declined due to these changes.

Image Aesthetics Assessment focuses on the visual appeal of an image. It evaluates how pleasing an image is to the human eye, considering factors like composition, color, and overall aesthetic impact. While related to IQA, since both involve human judgment, this area is not the focus of this thesis because it deals more with subjective perceptions of beauty rather than measurable quality degradations.

Image Fidelity Assessment deals with how accurately an image represents the original scene or view. This is especially relevant in applications involving multiple views or stereo cameras, assessing the correctness of image reconstruction. However, this thesis will also not cover image fidelity assessment, as it pertains more to the accuracy of recreating an image rather than evaluating its quality after processing.

The primary focus of this thesis is on IQA, specifically looking at various types of image degradation. The following subsections will discuss common distortions, datasets that contain these distortions, and the state-of-the-art (SOTA) methods in IQA. But first, it is important to distinguish between subjective and objective quality assessment.

2.1.1 Subjective Quality Assessment

Subjective quality assessment involves human observers evaluating the quality of images based on their visual perception. This method is essential for understanding how humans perceive image quality in real-world situations, especially when technical measurements might not fully capture what people actually see and experience. There are two primary methods used in subjective quality assessment:

- Absolute Categorical Rating: In this approach, human observers are presented with a unlabeled image and asked to rate its quality based on predefined categories. Each observer evaluates the image independently, without comparing it to any reference image. This method allows evaluators to provide a direct judgment on the image's quality based on their subjective experience.
- Paired Comparison: In this method, human observers are presented with two images: a unlabeled image and a reference image. Observers then assess the quality of the test image by comparing it directly to the reference image, assigning a score based on the perceived differences in quality.

Subjective quality assessment is highly valued for its ability to accurately reflect human perception of image quality. However, this method is also resource-intensive, requiring significant time and effort from human evaluators. Additionally, subjective assessments can be influenced by variability and biases introduced by individual scorers. For example, differences in monitor color calibration, the scorer's domain knowledge, and personal preferences can affect the consistency and reliability of the evaluations. Despite these challenges, subjective quality assessment remains a critical component of comprehensive image quality evaluation, particularly in applications where the human response to an image is the ultimate measure of its quality.

2.1.2 Objective Quality Assessment

Objective quality assessment relies on mathematical algorithms rather than human judgment to evaluate image quality. This approach uses our understanding of human vision system attributes to develop mathematical equations that measure quality, even though not all methods rely on these attributes. Essentially, it involves comparing data points or features extracted from images to determine quality. This assessment is mainly categorized into three methods based on the reference data used: Full-Reference IQA (FR-IQA), Reduced-Reference IQA (RR-IQA), and No-Reference IQA (NR-IQA).



Figure 2.1: General framework of FR-IQA algorithms. Features are extracted from both images, and then the feature distance is calculated.

Full-Reference IQA (FR-IQA) involves a comprehensive comparison between a distorted image and a reference image (see Figure 2.1). Features are extracted from both images, and their differences are quantitatively analyzed to compute a quality score. While FR-IQA offers detailed assessments, it requires a reference image for every distorted image evaluated, which can limit its practicality.

Reduced-Reference IQA (RR-IQA) operates similarly to FR-IQA but does not need the complete reference image. Instead, it uses a reduced set of features extracted from both the distorted and reference images (see Figure 2.2). This method balances the exhaustive comparison of FR-IQA and the independence of NR-IQA, reducing computational demands while still providing meaningful quality assessments based on partial reference data.

Both FR-IQA and RR-IQA utilize two methods to analyze quality:

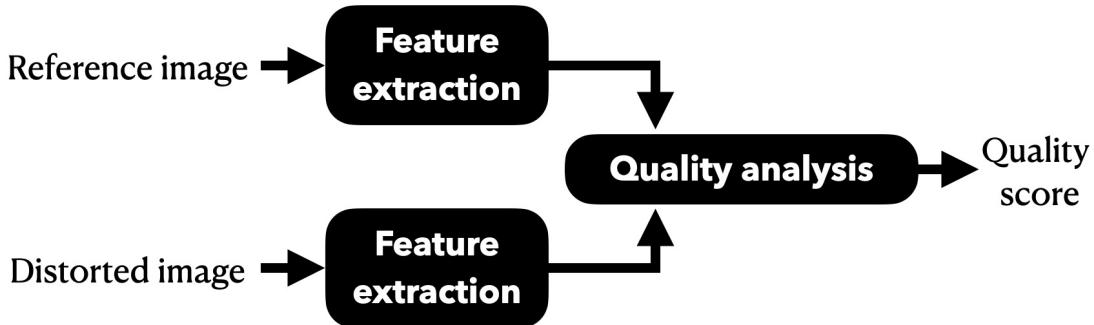


Figure 2.2: General framework of RR-IQA algorithms. Features of the reference and distorted images are extracted and used collectively to compute the quality.

- Spatial-Based Analysis: This method compares images pixel by pixel or region by region, offering straightforward interpretation and efficient computation. However, it may not fully align with how humans process images, lacking robustness in some scenarios.
- Transform-Based Analysis: This approach transforms images into a different domain (such as the frequency domain) that more closely mimics the human visual system. While robust, it is complex and computationally intensive.



Figure 2.3: General framework of no-reference image quality assessment algorithms.

No-Reference IQA (NR-IQA) does not rely on any reference image. Instead, it analyzes the distorted image alone by extracting features indicative of quality (see Figure 2.3). This method is particularly useful when no reference images are available, such as in many practical applications of teledermatology. NR-IQA can be tailored to address specific types of distortions or designed for general-purpose quality assessment, providing versatility across various domains.

For this thesis, the focus will be on No-Reference IQA because it is especially relevant for evaluating teledermatology images where reference images are usually not available. Since IQA measures distortions and NR-IQA can handle various types, it is important to identify the most common distortions encountered. The next subsection will discuss these distortions in detail.

2.1.3 Common Distortions in Image Quality Assessment

Image Quality Assessment (IQA) must address various distortions that can significantly affect the perceived quality of images. Understanding these common distortions is crucial for developing effective IQA algorithms, particularly in contexts like teledermatology, where accurate image assessment is critical. Below are the common distortions typically considered in IQA, with a reference image shown first for better comparison:

The common distortions are:

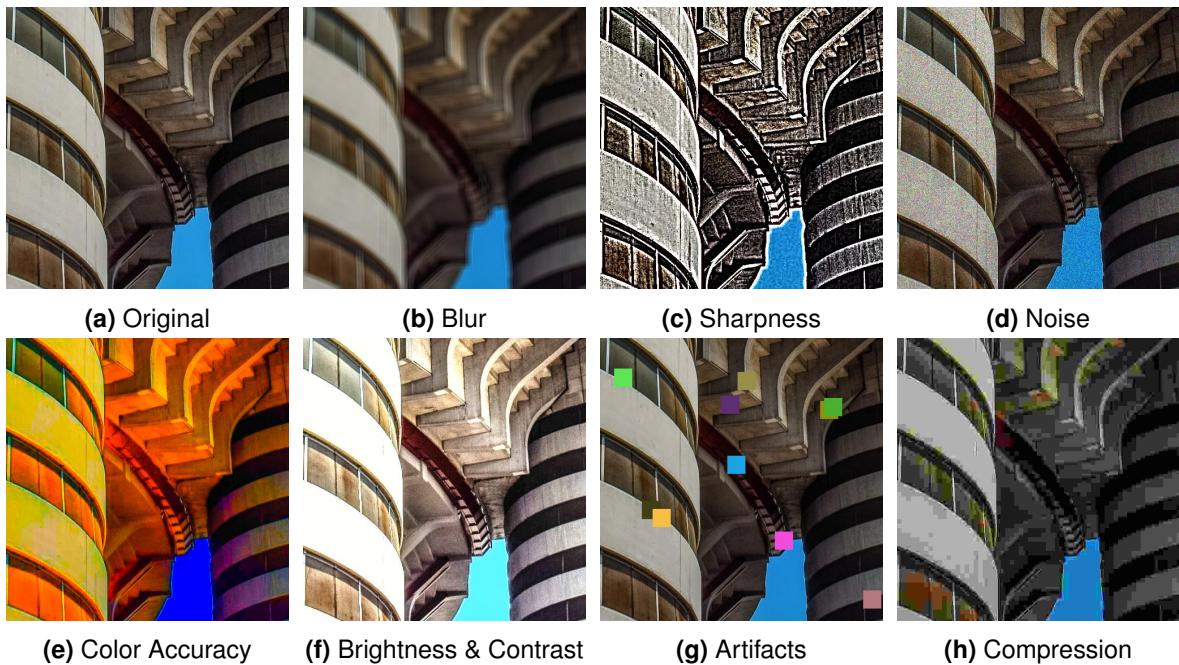


Figure 2.4: Examples of Common Distortions in Images. (adapted from (Agnolucci et al., 2023))

1. **Blur:** Blurred images lack sharpness and clarity, often resulting from motion during capturing, incorrect focus settings, or imperfections in the camera lens. See Figure 2.4b for an example of a blurred image.
2. **Sharpness:** Sharpness refers to how well-defined the edges and fine details in an image appear. High sharpness indicates clear, crisp images, while low sharpness makes an image look soft and unclear. See Figure 2.4c for an example of a sharpened image.
3. **Noise:** Noise appears as random variations in brightness or color and is often due to the limitations of the camera's sensor, particularly under low light conditions or at high ISO settings. See Figure 2.4d for an example of a noisy image.
4. **Color Accuracy:** Color accuracy refers to how faithfully colors are reproduced in an image. Distortions in color accuracy can lead to inaccurate or unrealistic color representation. See Figure 2.4e for an example of a color-distorted image.
5. **Brightness & Contrast:** Brightness is the overall light level of an image, while contrast refers to the range between its darkest and lightest areas. Proper balance of both is crucial for maintaining image visibility and detail. Excessive or insufficient brightness and contrast can make an image unusable for detailed analysis. See Figure 2.4f for an example of an image with altered brightness.
6. **Artifacts:** Artifacts are unwanted visual anomalies introduced during image acquisition or processing, such as halos, or jagged edges. See Figure 2.4g for an example of an image with artifacts.
7. **Compression:** When images are compressed to reduce file size, this often results in lost detail and visible quality degradation. See Figure 2.4h for an example of a compressed image.

Each type of distortion affects the visual quality and perceived accuracy of images, influencing the effectiveness of IQA methodologies in assessing image quality. Understanding these distortions

is essential for developing robust quality assessment algorithms and improving image clarity in various applications, including teledermatology.

2.1.4 Benchmark Datasets for IQA

Benchmark datasets play a vital role in advancing Image Quality Assessment (IQA). They provide standardized and diverse image sets with known distortions and corresponding quality annotations, helping researchers evaluate and improve IQA algorithms. These annotations, often in the form of Mean Opinion Score (MOS) and Differential Mean Opinion Score (DMOS), serve as benchmarks for algorithm performance.

Mean Opinion Score (MOS) is calculated by averaging ratings from human observers who judge the quality of images on a predefined scale. This score reflects the overall perceptual quality as seen by typical viewers and is widely used to compare the performance of different IQA methods against human visual judgment.

Differential Mean Opinion Score (DMOS), on the other hand, is derived from MOS and measures the perceived difference in quality between a reference image and a distorted version. This score is particularly useful for understanding the impact of specific distortions on image quality.

These datasets enable researchers to thoroughly test the robustness, accuracy, and generalization capabilities of different IQA methods. They also help in developing new algorithms by providing reliable quality scores, which are essential for ensuring reproducible.

An overview of IQA databases is provided in Table 2.1, and more detailed descriptions can be found in Appendix B.

2.1.5 State-of-the-Art in Image Quality Assessment

The current state-of-the-art in Image Quality Assessment (IQA) is ARNIQA (Agnolucci et al., 2023), with version 2 released on November 4, 2023. ARNIQA (leArning distoRtion maNifold for Image Quality Assessment) represents a major advancement in No-Reference Image Quality Assessment (NR-IQA). This technology aims to measure image quality based on human perception, even without a reference image. This capability is crucial in fields like teledermatology, where the quality of images directly impacts diagnostic accuracy.

Overview of ARNIQA: ARNIQA is developed using a self-supervised learning approach. It learns a comprehensive model of all possible image distortions, focusing on the types and quality of distortions rather than the content of the images themselves. This makes it highly adaptable across various domains where image content can differ significantly.

Key Features of ARNIQA:

1. **Image Degradation Model:** ARNIQA can synthetically degrade images through up to 1.9 billion distinct degradation patterns. This model can apply up to seven different types of distortions in one sequence, covering a wide range of real-world scenarios. Training with such diverse distortions ensures that ARNIQA can accurately assess image quality across various conditions and avoid the need for large labeled datasets.
2. **SimCLR Framework:** At the core of ARNIQA is the SimCLR (Simple Framework for Contrastive Learning) framework. This framework enables the model to learn meaningful representations of image quality by comparing different versions of the same image and focusing on their similarities and differences. SimCLR constructs positive pairs by applying

Table 2.1: An overview of IQA databases

Category	Database	Year	#Ref.	#Dist.	#Dist. Type	#Dist. Level	Resolution Type	Ground-truth
General	LIVE	2004	30	779	JPEG, JP2K, WN, GB, FF	5 or 4	768 × 512	DMOS
	TID2008	2008	25	1700	17 ^a	4	512 × 384	MOS
	TID2013	2013	25	3000	24 ^b	5	512 × 384	MOS
	CSIQ	2009	30	866	JPEG, JP2K, WN, GB, APGN, GCD	5 or 4	512 × 512	DMOS
	A57	2007	3	54	DWT, AGWN, JPEG, JP2K, JP2K-DCQ, GB	3	512 × 512	MOS
	WED	2017	4744	94880	JPEG, JP2K, GB, WN	5	-	-
	KADID-10k	2019	81	10125	25 ^c	5	512 × 384	DMOS
Multiple Dist.	KADIS-700k	2020	140000	700000	25 ^d	5	512 × 384	DMOS
	LIVEMD	2012	15	405	GB followed by JPEG, GB followed by WN	-	1280 × 720	DMOS
	MDID2013	2013	12	324	corrupted successively by GB, WN, and JPEG	-	768 × 512 or 1280 × 720	DMOS
	MDID2016	2016	20	1600	GB or CC first, JPEG or JP2K second and WN last	-	512 × 384	MOS
Screen content	SIQAD	2014	20	980	WN, GB, CC, JPEG, JP2K, MB, LSBC	7	700 × 700	DMOS
	SCIQ	2017	40	1800	WN, GB, MB, CC, JPEG, JP2K, CSC, CQD	5	1280 × 720	MOS
	CCT	2017	72	1320	HEVC and HEVC-SCC coding	11	1280 × 720 to 1920 × 1080	MOS
	HSNID	2019	20	600	WN, GB, MB, CC, JPEG, JP2K	5	-	MOS
Authentic Dist.	LIVE Wild	2016	0	1162	-	-	500 × 500	MOS
	CID2013	2015	0	480	-	-	1600 × 1200	MOS

Note: #Ref.: Total number of pristine images. #Dist.: Total number of distorted images. AGWN: Additive Gaussian white noise. WN: White noise.

APGN: Additive pink Gaussian noise. CC: Contrast change. CSC: Color saturation change. CQD: Color quantization with dithering.

DWT: Quantization of the LH subbands of a 5-level DWT. FF: Simulated fast fading Rayleigh channel. GB: Gaussian blur. MB: Motion blur.

GCD: Global contrast decrements. HEVC-SCC: Screen content coding extension of high efficiency video coding. JPEG: JPEG compression.

JP2K: JPEG2000 compression. JP2K-DCQ: JPEG-2000 compression with DCQ. LSBC: Layer segmentation based compression.

^aSee detailed types on database page: <https://www.ponomarenko.info/tid2008.htm>

^bSee detailed types on database page: <https://www.ponomarenko.info/tid2013.htm>

^cSee detailed types on database page: <https://database.mmsp-kn.de/kadid-10k-database.html>

^dSee detailed types on database page: <https://database.mmsp-kn.de/kadid-10k-database.html>

the same distortion settings to two different images, ensuring that the model concentrates on the distortions rather than the content. To further enhance the model's discriminative capability, SimCLR introduces subtle variations by downsampling images before cropping and applying distortions, creating hard negative examples. These examples help the model differentiate between similar-looking images with different types of degradation. By using this approach, the SimCLR framework ensures that ARNIQA effectively learns to recognize and assess various distortions, enhancing its ability to provide accurate image quality

assessments (see Figure 2.5).

3. Linear Regressor: A Linear Regressor maps the features learned by SimCLR to quality score ranging from 0 to 1. This score reflects the relative quality of the image based on the distortions present.

Training Strategy of ARNIQA: ARNIQA's training strategy involves two main phases:

1. Encoder Pre-training: The model is first trained on a large set of unlabeled images that are synthetically degraded. This helps the encoder learn features related to different types and levels of image degradation.
2. Regressor Training: In the second phase, a specific regressor is trained using the Mean Opinion Scores (MOS) of images. This step translates the learned features into actual quality scores.

Advantages of ARNIQA: ARNIQA achieves high performance with only up to 0.5% of the data required by other methods, thanks to its focus on distortion patterns rather than image content. It provides reliable and consistent quality assessments across a wide range of distortions and severities, demonstrating its robustness. Additionally, ARNIQA is particularly suitable for tele-dermatology as it can handle varying image quality resulting from different lighting conditions, camera quality, and patient handling.

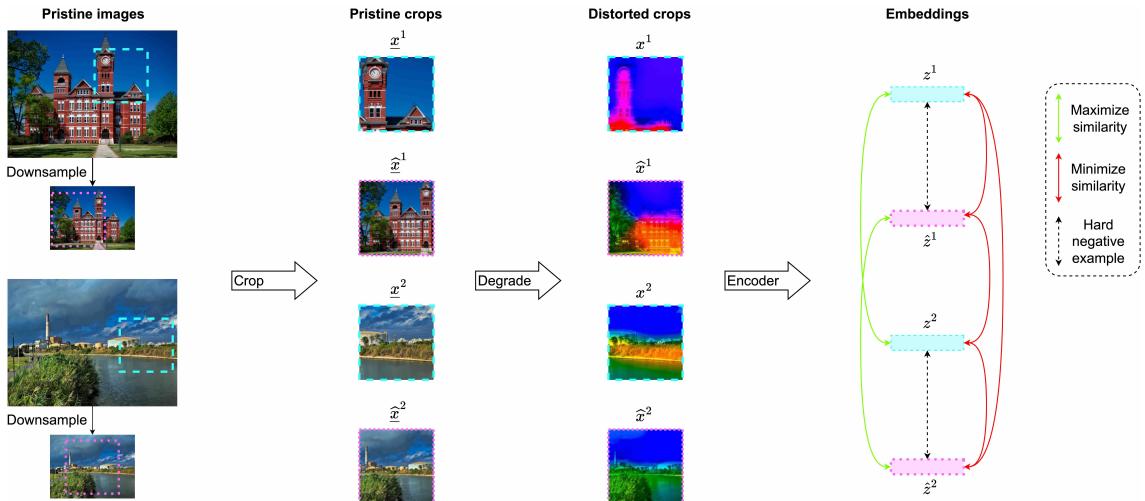


Figure 2.5: Overview of the training strategy for ARNIQA. Two pristine images are cropped and equally degraded. The model maximizes the similarity of their embeddings while minimizing the similarity to embeddings from degraded crops of half-scale versions of the original images. This process creates hard negative examples by introducing downsample distortion, demonstrating how original and half-scale degraded crops differ despite identical degradation. (Agnolucci et al., 2023).

2.1.6 Challenges and Opportunities in Image Quality Assessment

In the realm of Image Quality Assessment (IQA), practitioners face the challenge of accounting for the variability of conditions under which images are captured. This is especially true in tele-dermatology, where variables such as lighting and camera quality can significantly affect image

consistency. Moreover, the subjectivity inherent in human visual perception adds complexity to developing algorithms that accurately reflect human assessments of image quality.

Another significant challenge is the diversity of image content, which makes it difficult to apply uniform quality criteria across different types of images. Additionally, real-world images often present multiple interacting distortions, unlike the isolated distortions typically studied in laboratory settings, complicating the quality assessment process. Scalability also presents a hurdle, as the increasing volume of image data demands efficient processing for quality evaluation.

On the other side, the landscape of IQA also presents several opportunities. Advances in machine learning, particularly with self-supervised approaches like ARNIQA, open up new possibilities for training models that require less annotated data and can generalize across various conditions. Such technological progress bodes well for teledermatology, where enhanced IQA could lead to more accurate diagnoses.

The drive towards standardized image capturing and processing protocols represents another opportunity to improve image quality consistency. Additionally, interdisciplinary research combining insights from computer science, imaging, and medical fields is essential to tailor IQA methods for specific medical applications. Finally, leveraging big data analytics can provide a comprehensive understanding of common quality issues, informing the development of more refined IQA tools.

2.2 Teledermatology

Within the dynamic spectrum of telemedicine, teledermatology emerges as a distinct application that utilizes information and communication technologies to facilitate dermatological consultations remotely. This modality of healthcare has seen widespread adoption, especially in resource-rich regions like Europe and North America, where the availability of advanced technologies has allowed for the provision of high-quality images essential for accurate diagnoses.

The following section provides an overview of teledermatology, a specialized field of dermatology that utilizes telecommunications technology to provide remote diagnosis and consultation for skin conditions. This section discusses the importance of image quality in teledermatology, quality criteria for teledermatology images, as well as challenges and opportunities associated with the practice.

2.2.1 Introduction to Teledermatology

Teledermatology can be effectively categorized into two primary approaches: real-time (RT) and store-and-forward (S&F). RT teledermatology facilitates live interactions between patients and physicians through video calls, while S&F involves capturing and sending images for later review by a dermatologist. The S&F method has gained prominence due to its convenience and adaptability to varying schedules.

A typical teledermatology workflow begins with the patient capturing an image of their skin condition using a digital device. This image is then transmitted through a teledermatology platform to a medical expert who reviews the image's quality and details. The dermatologist then provides a report or prescription back to the patient, completing the consultation cycle. This workflow underscores the critical nature of image quality in teledermatology, as diagnostic accuracy is heavily reliant on the clarity and fidelity of the transmitted images.

2.2.2 Importance of Image Quality in Teledermatology

In teledermatology, the caliber of transmitted images is critically pivotal. Clear and detailed images are the foundation upon which dermatologists rely for diagnosing and managing skin conditions from afar. When the images are of high quality, they capture essential details such as texture and color nuances that can be key to distinguishing between benign and more severe dermatological issues.

The resolution, focus, and accurate color representation in these images can markedly streamline the teledermatological process. They minimize the necessity for additional consultations due to poor image clarity, thereby improving the overall efficiency of the healthcare system and reducing patient wait times.

As the primary conduit for remote dermatological assessment, the images not only facilitate immediate patient care but also feed into the broader ecosystem of teledermatology that includes emerging technologies such as artificial intelligence. High-fidelity images are integral to training sophisticated AI algorithms, which promise to further enhance diagnostic precision and expedite the triage process.

In summary, the emphasis on image quality in teledermatology is not merely a current requirement but a crucial investment in the future of dermatological care, ensuring continued improvements in patient outcomes and the evolution of healthcare delivery methods.

2.2.3 Quality Criteria for Teledermatology Images

In teledermatology, the efficacy of remote diagnoses heavily relies on the quality of the images. Just as Image Quality Assessment (IQA) must confront various distortions affecting image perception, teledermatology faces its own set of quality criteria that are essential for effective practice. Proper lighting, background uniformity, appropriate field of view, accurate orientation, precise focus and depth of field, high resolution, and correct color calibration are paramount in ensuring that the dermatological images transmitted for evaluation are of the highest possible quality.

1. **Lighting:** Adequate illumination is critical. It should be even and diffuse, avoiding harsh shadows or overexposure that could obscure skin lesions or lead to misinterpretation of the skin's condition. **Remark** Position the light source evenly to avoid shadows and overexposure. Natural light or diffused artificial light can help illuminate the skin lesion uniformly and reduce glare.
2. **Background:** A neutral and uncluttered background helps to focus attention on the dermatological issue without distraction, ensuring that the skin lesion is the most prominent feature in the image. **Remark** Use a plain, non-reflective background to minimize distractions and ensure the focus remains on the skin lesion. A neutral-colored backdrop, such as white or gray, is ideal for providing contrast with the lesion.
3. **Field of View:** The image should be framed to include a clear view of the lesion as well as sufficient surrounding area to provide context, which can be crucial for accurate diagnosis. **Remark** Center the skin lesion or area of interest within the frame to ensure complete coverage and avoid cutting off important details. Maintain a consistent distance between the camera and the skin to prevent distortion.
4. **Orientation:** Proper orientation of the image is vital. It should align with standard anatomical positions, allowing the dermatologist to easily interpret the image in relation to the

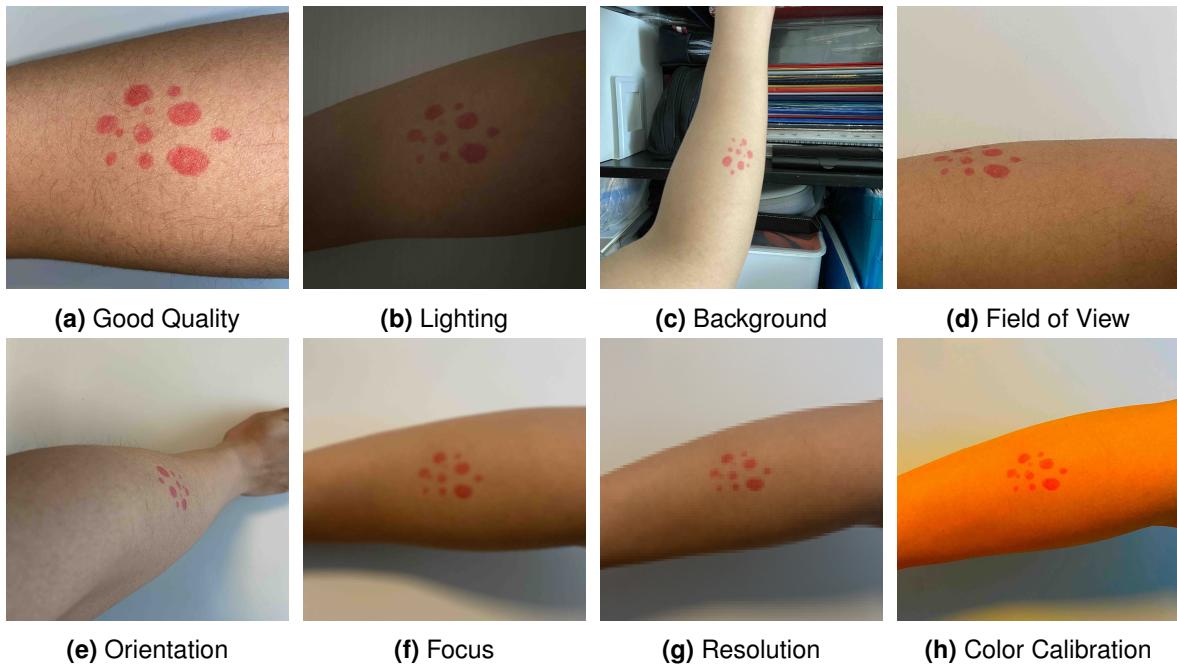


Figure 2.6: Example images illustrating common distortions used in Teledermatology Image Quality Assessment. (Created by the Author)

patient's body. **Remark** Orient the camera perpendicular to the skin surface to capture images in the correct orientation. Align the camera with the skin lesion to maintain consistency and facilitate accurate comparison between images.

5. **Focus & Depth of Field:** Sharp focus on the lesion is a necessity, with a depth of field that keeps the entire area of interest in clear detail, as blurring can mask important characteristics of skin conditions. **Remark** Ensure the camera is in focus and adjust the aperture to achieve sufficient depth of field. Focus on the skin lesion to capture sharp, detailed images without blurriness or loss of clarity.
6. **Resolution:** The image must be high resolution to reveal fine details of the skin. A higher pixel count can facilitate a more thorough examination and better clinical decision-making. **Remark** Use a camera with high-resolution capabilities to capture fine details and nuances of the skin lesion. Adjust the camera settings to the highest resolution possible to ensure clarity and precision in the image.
7. **Color Calibration:** Accurate color reproduction is necessary for the assessment of skin lesions. Any color distortion can lead to misdiagnosis, especially in conditions where hue is a diagnostic clue. **Remark** Calibrate the camera settings to accurately reproduce colors and skin tones. Avoid harsh lighting or color casts that may distort the color representation of the skin lesion. Use a color reference chart or white balance settings to ensure color accuracy.

Each of these quality criteria contributes to the diagnostic accuracy in teledermatology by ensuring that the images convey the true nature of the skin condition. Just as in traditional dermatology, where the dermatologist's visual assessment is a key diagnostic tool, in teledermatology, the image serves as the eyes of the dermatologist. Therefore, optimizing these criteria is crucial to the successful application of teledermatology services.

2.2.4 Teledermatology Datasets

The following datasets are valuable resources for teledermatology research and applications, especially notable for their accessibility and diversity. These datasets are not strictly confined to clinical settings or dermoscopic images and include patient-taken images, making them highly relevant for practical teledermatology purposes where clinical settings may vary:

- **ACNE04:** This dataset focuses on acne severity and lesion counting, containing 1,457 images with detailed annotations for training and testing purposes (Wu et al., 2019).
- **DDI:** Provides 656 high-quality images curated by dermatologists for detailed skin tone evaluation and diagnostic accuracy (Daneshjou et al., 2022).
- **Derm7pt:** Utilizes 1,011 lesion cases to train a neural network for classifying skin lesions and melanoma using the 7-point checklist (Kawahara et al., 2019).
- **Fitzpatrick17k:** Includes 16,577 images annotated for Fitzpatrick skin type across 114 different skin conditions (Groh et al., 2021).
- **Monkeypox Dataset 2022:** Contains approximately 1,905 images focused on monkeypox, useful for developing diagnostic tools (Ahsan et al., 2022).
- **PAD-UFES-20:** Comprises 2,298 clinical images from smartphones, enriched with clinical metadata for comprehensive research (Pacheco & Krohling, 2020).
- **SCIN:** Emerged from a crowdsourcing initiative, this dataset contains 10,408 images capturing a broad spectrum of dermatological conditions (Ward et al., 2024).

These datasets collectively contribute to the advancement of teledermatology by providing varied, real-world data crucial for developing effective diagnostic tools and algorithms.

2.2.5 Approaches to Image Quality Assessment in Teledermatology

To understand the various approaches to Image Quality Assessment (IQA) within teledermatology, it is instructive to review related works that have contributed to the field. These studies not only offer insights into the methodologies employed for assessing image quality but also highlight the specific image distortions that are often targeted. Furthermore, by examining the architectures of the algorithms and the criteria used to classify image quality, we can discern the strengths and weaknesses of each approach.

TruelImage: A Machine Learning Algorithm to Improve the Quality of Telehealth Photos

The TruelImage algorithm prioritizes real-time, interactive feedback for patients taking dermatological images via their smartphones. Its three-stage process includes semantic segmentation to identify skin regions, feature generation focusing on blur, lighting, and zoom, and logistic regression classifiers that predict image quality and specific reasons for poor quality (Vodrahalli et al., 2020).

The prototype exhibits promise, demonstrating the capability to reject approximately half of subpar quality images while retaining around 80% of good quality images. This suggests its potential utility in a clinical setting, where it could save time for both clinicians and patients by pre-screening image quality and offering specific feedback to improve poor submissions (Vodrahalli et al., 2020).

Yet, TruelImage's current limitation lies in its modest dataset and the training data's lack of diversity regarding skin types, which could lead to biased quality assessments. Furthermore, while the

algorithm excels at detecting blurriness, it is less effective at assessing lighting conditions and zoom, which are critical factors in teledermatology (Vodrahalli et al., 2020).

ImageQX: Explainable Image Quality Assessments in Teledermatological Photography

ImageQX offers an automated, deployable method for IQA in teledermatology, addressing common image distortions: such as poor framing, bad lighting, blur, low resolution, and distance issues. With an architecture that employs the EfficientNet-B0 as a lightweight feature extractor. The network then employs linear layers, batch normalization, and dropout layers to predict poor image quality reasons, integrating them to forecast overall image quality (Jalaboi et al., 2023).

The approach is data-driven, validated on a substantial dataset annotated by dermatologists, which underscores its potential for real-world applicability. With a macro F1-score of 0.73 for image quality assessment, ImageQX demonstrates expertise comparable to dermatologists and a capability for explainability via attention maps that can guide patients to retake pictures more effectively (Jalaboi et al., 2023).

However, while ImageQX achieves high specificity, suggesting minimal disruption to patient experience by incorrectly rejecting high-quality images, its predictive performance for certain poor quality explanations like 'bad framing' remains low, with an F1-score of 0.37 (Jalaboi et al., 2023). This indicates room for improvement in addressing certain types of distortions, though the network's size (15MB) makes it an attractive option for mobile deployment.

ImageQX stands out with its deep learning approach that uses a convolutional neural network (CNN) trained on a dataset of images labeled for quality by board-certified dermatologists. The network architecture is based on the lightweight EfficientNet-B0, which facilitates deployment on mobile devices. This model's training utilized a large dataset of 36,509 images, where the photographs were annotated for common quality issues, such as framing, lighting, blur, resolution, and distance from the subject.

The strength of ImageQX lies in its explainability and its ability to provide reasons for poor image quality, aligning closely with the common distortions identified by dermatologists. It achieves a high macro F1-score, suggesting that its performance in image quality assessment is on par with expert human raters. However, some limitations include the difficulty of explaining certain quality issues like 'blurry' images, which had lower inter-rater agreement scores. The model's reliance on dermatologist-annotated images also points to the importance of the quality and diversity of the training dataset for its generalizability and accuracy.

In conclusion, both ImageQX and Truelimage present pioneering steps towards automated IQA in teledermatology, each with strengths such as deployability and real-time feedback capabilities. Their current challenges include the need for more diverse training data, improved detection of certain distortions, and further validation in real-world settings. Addressing these challenges will be crucial for their successful integration into teledermatological practices, ultimately contributing to more efficient and effective remote dermatological care.

2.2.6 Challenges and Opportunities in Teledermatology

Teledermatology has fundamentally changed how dermatological care is accessed, particularly in remote or underserved areas. However, the field faces several challenges that intersect with the focus of this thesis on image quality assessment.

One major challenge is the variability in image quality, which stems from patients using a wide range of devices in uncontrolled environments to capture images. This variability can severely impact the consistency and reliability of diagnoses made remotely. Additionally, technological barriers such as disparities in access to high-quality digital devices and varying levels of digital literacy among patients can further affect the quality of submitted images. Moreover, ensuring the privacy and security of sensitive dermatological images transmitted over the internet remains a critical concern that needs continuous attention to safeguard patient data.

On the opportunity front, recent advancements in image processing and machine learning technologies present significant potential to automatically enhance image quality and correct common distortions. This can greatly aid in improving diagnostic accuracy and efficiency. There is also a considerable opportunity to develop and implement standardized guidelines for image capture in teledermatology. Such standardization can help minimize quality variability and streamline the diagnostic process. Furthermore, the integration of artificial intelligence to provide real-time feedback to patients on improving image quality can enhance the effectiveness of teledermatology services by ensuring that only high-quality images are evaluated by dermatologists.

By tackling these challenges and leveraging the emerging opportunities, teledermatology can continue to evolve and play a crucial role in modern healthcare, enhancing access and efficiency in dermatological care across diverse populations.

Chapter 3

Methodology

This chapter outlines the methodologies used throughout the project to achieve the research objectives and answer the thesis questions concerning image quality assessment (IQA) in teledermatology. It discusses the exploratory approach adapted during the project, characterized by cycles of decision-making and agile methodologies.

3.1 Explorative Approach

Teledermatology, particularly focused on IQA, presents a broad scope for innovation due to the variety of possible image distortions and the different ways these can be addressed. The exploratory approach adopted in this research is characterized by its high degree of innovation and flexibility, which sometimes comes at the expense of linearity. Traditional phased approaches are not always applicable, leading to the adoption of a methodology based on explorative cycles with decision-making processes and agile practices.

Initially, the problem statement for this project was only broadly defined, setting the stage for an adaptive and fluid approach as the research unfolded. As the project progressed, it increasingly focused on creative problem-solving, facilitated through multiple learning cycles that allowed for iterative refinement of ideas and methods. The research was structured into two distinct phases to manage this process effectively. In the early stages, known as the Diverging Phase, the scope of the research question was deliberately expanded. This expansion enabled the continuous generation of new ideas, each informed by the insights gathered during the ongoing investigation. Later, in the Converging Phase of the project, the emphasis shifted towards synthesizing these ideas into a cohesive set of findings and conclusions, aiming to consolidate the diverse insights into a unified understanding of the initial problem statement (Hoffmann et al., 2016). A schematic of this exploratory model is depicted in Figure 3.1.

3.2 Project Control

Despite the exploratory approach's lack of traditional milestones, it is crucial to have a rough timeline to successfully navigate the research tasks. Before commencing, a workflow was established, detailed in the planning document attached to this thesis. There are three essential milestones identified in the first half of the project, each critical to its success:

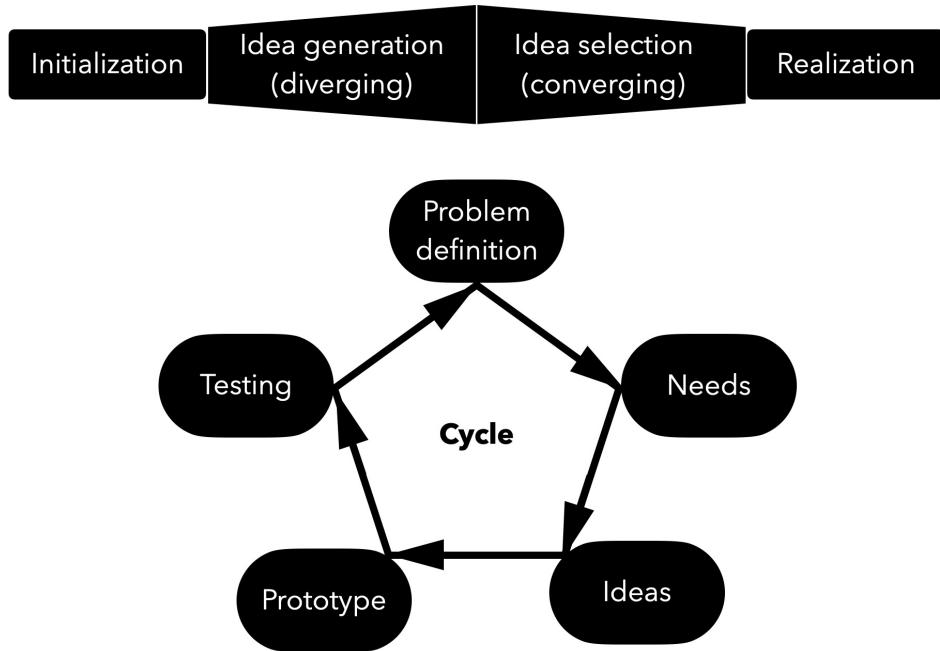


Figure 3.1: Illustration of the explorative approach, including the stages of initialization, idea generation, idea selection, and implementation. The lower part of the figure shows the decision cycles (Hoffmann et al., 2016).

Understanding Teledermatology: Gaining a deep understanding of the domain to ensure all subsequent actions are relevant and informed.

State of the Art in IQA: Identifying the latest developments in IQA to ensure the methods used are cutting-edge.

Availability of Teledermatology Data: Ensuring access to appropriate datasets for conducting meaningful IQA.

These milestones are pivotal as each subsequent phase relies on the successful completion of the previous. Failure to achieve these could severely impact the project, potentially necessitating a fundamental reassessment of the objectives according to Section 1.2.

3.3 Research Steps

Due to the explorative nature of this study, a linear progression through predefined research steps was not feasible. However, key steps were categorized and executed as follows:

3.3.1 Literature Review

The literature review is foundational to the research, setting the stage for framing the research questions and determining the appropriate methodologies. This initial phase involved a thorough examination of existing research on IQA, focusing specifically on methodologies previously employed and their effectiveness within teledermatology. By analyzing these methodologies, the review highlighted both their strengths and limitations, providing a nuanced understanding of the current landscape and the gaps in knowledge and application.

The choice of ARNIQA as the primary methodological tool for this thesis was influenced signif-

icantly by the insights gained from the literature. ARNIQA's advanced capabilities in handling a range of image distortions—which are critical for teledermatology—made it an ideal choice. Traditional IQA methods often fall short in managing the complex and varied image qualities encountered in teledermatology; however, ARNIQA, with its robust image degradation model and its use of SimCLR for learning from unlabeled images through contrastive learning, offers a substantial improvement. This approach is particularly advantageous in scenarios where high-quality labeled datasets are scarce or incomplete.

The literature review not only reinforced the methodology's alignment with the project's aims but also highlighted ARNIQA's potential to enhance current IQA practices in teledermatology. This phase was crucial for ensuring that the subsequent steps were built on a solid theoretical foundation, guided by informed insights into the state of the art, and tailored to meet the specific demands of assessing image quality in a teledermatological context. The outcomes of this foundational step provided the necessary groundwork for the detailed investigations and methodological applications that followed.

3.3.2 Data Collection and Preparation

In the pursuit of a suitable dataset for evaluating image quality within teledermatology, a significant challenge encountered was the absence of teledermatology datasets that included Mean Opinion Score (MOS) or Differential Mean Opinion Score (DMOS) similar to those commonly available in traditional IQA datasets mentioned in Subsection 2.1.4. This scarcity is largely due to the clinical nature of most teledermatology images, which are often captured using dermoscopes in controlled settings, thereby not representing the typical use case in remote dermatological assessments.

To address this gap, the SCIN dataset was chosen for its relevance and uniqueness. Unlike many dermatological datasets that predominantly focus on skin cancer diagnostics through the classification of malignant and benign tumors, the SCIN dataset encompasses a broader spectrum of common dermatological conditions. These conditions primarily include allergic, inflammatory, and infectious diseases, which are frequently encountered in everyday clinical practice but underrepresented in existing datasets. What makes the SCIN dataset particularly valuable for this research is that it captures images of early-stage concerns—over half of the images were taken less than a week from the onset of symptoms, with 30% captured less than a day after. These are conditions patients are likely to consult via teledermatology platforms before they could be seen in a traditional healthcare setting.

Given the project's focus on image quality assessment rather than diagnostic accuracy, I adapted the SCIN dataset for my specific research needs. This adaptation involved creating two distinct sets of images:

Good Quality Set: This set was compiled based on my assessment of what constitutes 'good quality' in dermatological images. Although I am not a dermatologist, the selection was guided by general quality criteria relevant to both clinical and non-clinical settings.

Test Set with Labeled Images: To quantitatively assess image quality, I personally labeled a test set, applying scores from 0 to 1 for each of the seven quality criteria Subsection 2.2.3 identified as critical for teledermatology. In this scoring system, a score of 1 indicates extreme distortion relevant to the specific criterion, and a score of 0 signifies no noticeable distortion.

This approach to dataset preparation not only tailored the data to the specific needs of this research but also established a framework for systematically assessing image quality in teleder-

matology. This preparation is crucial for the next phases of the project, which involve training and validating the image quality assessment model to ensure it can reliably perform in real-world teledermatology applications.

3.3.3 Training and Validation

The model training process begins with the application of a series of predefined distortions to a set of high-quality images. These distortions are carefully selected to simulate real-world imperfections commonly encountered in teledermatology. Each image is then labeled according to the severity and type of distortion applied, creating a dataset that not only includes the distorted images but also features precise annotations regarding their quality.

Feature extraction is a critical next step where the SimCLR model from ARNIQA is utilized to derive meaningful representations from the distorted images. These features are expected to capture the underlying patterns of distortions that affect image quality.

Overview of Training and Validation Processes

Model Selection and Training: A regression model will be employed to correlate the extracted features with the labeled image quality scores. The choice of model—be it a Random Forest or a Linear Regressor—will be based on a balance of accuracy, interpretability, and computational efficiency. This model will learn to predict the quality of an image, effectively turning the features extracted by SimCLR into actionable insights.

Validation Strategy: Validation plays a crucial role in ensuring that the model not only performs well on the training data but is also effective and reliable when deployed in real-world teledermatology settings. The model will be validated using a separate set of images that were not included in the training phase. Standard statistical metrics such as accuracy, precision, and recall, among others, will be used to evaluate the model's performance, ensuring its efficacy and robustness.

For a comprehensive understanding of the specific methodologies employed — including the intricacies of the distortion pipeline, the operational details of SimCLR, and the exact validation protocols — please refer to the Implementation Chapter 4. This section will delve into the technical specifics, providing a detailed description of each step involved in the process, from the initial data handling to the final stages of model evaluation.

3.3.4 Testing and Experiments

The testing and experimental phase of the research is crucial for validating the efficacy and accuracy of the image quality assessment model developed in the previous stages. The testing phase employs a set of images that I personally labeled based on seven critical quality criteria for teledermatology, with scores ranging from 0 (no distortion) to 1 (extreme distortion). This phase evaluates the model's ability to assess image quality accurately, using metrics such as Mean Squared Error (MSE) and R-Squared. These metrics will help determine how well the model's predictions align with the actual labels, providing a baseline for assessing the model's performance and guiding future improvements. Detailed procedures and metrics will be elaborated further in the Implementation Chapter 4.

3.3.5 Discussion and Further Development

The final phase of the project involves analyzing and discussing the results obtained from the model testing. This discussion will assess how effectively the model meets the research objectives and will highlight areas where the model excelled or fell short. Insights gained from this

analysis will inform potential areas for further development, such as refining the model's ability to handle specific distortions or improving its generalization across diverse teledermatology images. This ongoing cycle of evaluation and enhancement is crucial for advancing the field of image quality assessment in teledermatology, ensuring that the methodologies continue to evolve in line with technological advancements and clinical needs.

Note:

The approach to training and validation in this thesis involves a novel method of applying and assessing distortions based on the defined quality criteria critical to teledermatology. The process begins by taking good quality images and subjecting them to a systematic distortion pipeline, which is designed to simulate real-world imperfections that could affect diagnostic accuracy in teledermatology.

Distortion Implementation:

Each of the seven quality criteria identified has a corresponding distortion type, such as motion blur for focus. These distortions are quantified across five levels, from 0 (no distortion) to 8 (severe distortion). The application of these distortions is controlled by a probabilistic model based on a Gaussian distribution, with a mean of 0 and a standard deviation of 2.5. This statistical approach ensures a natural variability in the application of distortions, closely mimicking the random nature of image quality issues in practical settings.

The specific levels of distortion for each image are selected using normalized probabilities, ensuring that each level of severity has a realistic chance of being applied. The distortions are then applied to the images, and each image is labeled with values scaled between 0 to 1 corresponding to the severity of the applied distortions, with these labels indicating the level of quality degradation for each criterion.

Feature Extraction and Regression Model Training:

Once the images have undergone distortion, they are processed through the SimCLR framework from ARNIQA to extract features. SimCLR is particularly suited for this task as it is designed to learn useful representations from unlabeled data in an unsupervised manner, making it ideal for scenarios where explicit labels might not be available or fully reliable.

The extracted features, along with the generated labels, are then used to train a regression model. The choice of the regression model, whether it be a Random Forest or a simple Linear Regressor, will depend on the performance criteria such as accuracy and computational efficiency. The regression model will learn to predict the quality level of an image based on the features extracted by SimCLR, which are indicative of the various distortions applied.

Validation:

The validation phase involves assessing the trained model's accuracy and its ability to generalize across different sets of images. This is crucial for ensuring that the model performs well not only on the training data but also on unseen images, which would be representative of a real-world application in teledermatology. The model's performance is evaluated using standard metrics such as Mean Squared Error (MSE) or R-squared, which provide insights into how well the model predicts image quality.

This training and validation process is critical as it directly influences the reliability and effectiveness of the IQA in a teledermatology context. It ensures that the system can accurately identify and quantify the severity of various distortions, thus supporting dermatologists in making informed decisions based on the quality of the images they review.

Considerations:

It's essential to ensure that the distribution and application of distortions during the training phase closely mimic realistic scenarios to prevent the model from overfitting to unrealistic patterns. Additionally, considering ethical aspects and the potential impact on patient outcomes, it is crucial to maintain high accuracy and reliability in the model's predictions to support effective

and safe teledermatology practices.

Literature Review on IQA and Teledermatology

Purpose: Start by discussing the significance of the literature review in framing your research questions and methodology.

Content: Outline how the literature influenced the selection of your methods and tools, particularly focusing on past approaches to IQA and their applicability or limitations in teledermatology.

Rationale: Explain the choice of ARNIQA based on gaps or strengths identified in the literature, establishing why it's well-suited for addressing current challenges in teledermatology image quality assessment.

Image Quality Assessment Methodology

Introduction to ARNIQA: Detail why ARNIQA was chosen for IQA in teledermatology, emphasizing its strengths such as the sophisticated image degradation model and its ability to train with fewer labeled examples.

Explaining SimCLR: Provide an in-depth explanation of SimCLR, discussing how it works (contrastive learning mechanism), why it is effective (ability to learn useful representations from unlabeled data), and its particular advantages for your research (e.g., robustness to various distortions).

Utility and Implementation: Describe how you implemented ARNIQA and SimCLR, including any modifications or optimizations made for teledermatology. Mention the availability of code and weights, which ensures reproducibility and facilitates future research.

Teledermatology Image Quality Assessment

Dataset Description: Introduce the SCIN dataset, explaining why it's suitable for your study, its composition, and any preprocessing steps involved.

Distortion Model Creation: Discuss the design of your custom distortion model, detailing the types and layers of distortions you included. Justify why these particular distortions are relevant to teledermatology.

Test Set and Labeling: Explain how you created and labeled your test set, including the criteria used for labeling and the process of validation.

Architecture Overview: Provide a comprehensive overview of the entire system architecture, showing how each component (data input, processing, analysis, and output) integrates to form a cohesive workflow.

Summary of Methodological Approach

Synthesis: Briefly summarize how each part of your methodology contributes to addressing the research questions or hypotheses stated in earlier chapters.

Justification: Reinforce the rationale behind your methodological choices, linking back to the literature review and the specific challenges identified in teledermatology IQA.

Chapter 4

Implementation

This chapter delves into the detailed implementation of the methodologies outlined in Chapter 3, focusing on the specific processes, experiments, and analyses conducted. It includes the practical steps taken to prepare data, apply distortions, extract features, and train the regression model to assess image quality in teledermatology.

4.1 Image Selection and Labeling Process

This section describes the initial stages of the implementation, focusing on the selection and preparation of the image datasets used in the study.

4.1.1 Image Filtering and Selection

The first step in preparing the images involves carefully choosing high-quality pictures from the SCIN dataset, which mainly includes pictures of early-stage dermatological conditions. This selection is done manually as it ensures that each image is clear and useful for clinical use. The primary focus during selection is for images that are well-framed and free of any distortions that might affect their usefulness in diagnosis.

Each selected image is checked to make sure it doesn't have any blurring that might hide important details of the skin condition, as clear images are crucial for accurate diagnosis. Additionally, it's important that the images are properly lit and show true contrast, this means they shouldn't be too bright or too dark. Proper lighting and contrast help in accurately showing the skin's condition. Lastly, the images must represent realistic skin tones and colors. Accurate color representation is critical because wrong colors can lead to incorrect diagnoses. This careful selection process of images ensures that the baseline images used for further distortion and analysis are of good quality, providing a solid foundation for the subsequent experimental stages.

4.1.2 Labeling of the Test Set

The labeling process involves manually scoring approximately 50 high-quality images and 200 images with various distortions. These images are evaluated based on seven key quality criteria crucial for teledermatology: lighting, focus, orientation, color calibration, background, resolution, and field of view. Each image is scored on a scale from 0 to 1 for each criterion, where 0 indicates no distortion and 1 indicates extreme distortion.

This manual labeling is facilitated through a custom Python script, which displays each image and prompts the user to enter scores for each distortion criterion. The scores are collected in a structured format and stored in a JSON file for subsequent analysis. This structured and meticulous approach ensures that each image is evaluated consistently and comprehensively.

Visualization of Label Distribution

To understand the distribution of labels and how frequently distortions occur across different criteria, histograms are generated. These histograms are particularly useful for visualizing the prevalence and severity of distortions in the dataset. Two histograms are plotted for each criterion:

The first histogram shows the distribution of images scored as '0' for a specific criterion, representing images where no distortion is observed.

The second histogram displays the distribution for images where a distortion is present (scores >0), showing the varying levels of distortion severity.

These histograms provide valuable insights into the commonality and impact of each type of distortion, aiding in the analysis of how distortions affect overall image quality. They also highlight the criteria that may require more focused attention during the model training and validation process.

4.2 Distortion Pipeline

The distortion pipeline is central to simulating realistic image quality issues. Each quality criterion has multiple types of distortions, each having five ranges of intensity, increasing in severity. All distortion types begin at zero, indicating no distortion applied, and progress to higher values that represent increasing levels of the specified distortion. Visual representations of the types of degradations at different ranges for each quality criterion is provided in the Appendix A.

This section describes each distortion type and its impact on image quality, highlighting the adaptability of the pipeline in generating training images that reflect a range of realistic conditions.

4.2.1 Distortion Types

Here, each distortion type is briefly described, highlighting how they simulate different aspects of image degradation:

1. Lighting:

- *Brighten*: applies a sequence of color space transformations, curve adjustments, and blending operations to enhance the brightness of an input image, resulting in an output image with increased visual intensity;
- *Darken*: similar to brighten operation, but it leads to a decreased visual intensity;

2. Focus:

- *Gaussian blur*: filters every pixel of the image with a simple Gaussian kernel.
- *Lens blur*: filters every pixel of the image with a circular kernel;
- *Motion blur*: filters every pixel of the image with a linear motion blur kernel to simulate the effect of a moving camera or a moving object in the scene. Consequently, the image appears blurred in the direction of the motion;

Histogram:
 "Distribution of images scored as 0 for the focus criterion, indicating no distortion." or "Distribution of scores for the focus criterion where distortion is present, illustrating the range and frequency of severity."

3. Orientation:

- *Top perspective*: temp;
- *Bottom perspective*: temp;
- *Left perspective*: temp;
- *Right perspective*: temp;

4. Color calibration:

- *Color saturation 1*: converts the image to the HSV-color space and then multiplies the saturation channel by a factor;
- *Color saturation 2*: converts the image to the LAB-color space, then multiply each color channel by a factor;

5. Background:

- *temp*: temp;

6. Resolution:

- *temp*: temp;

7. Field of view:

- *temp*: temp;

4.3 Distortion Implementation Process

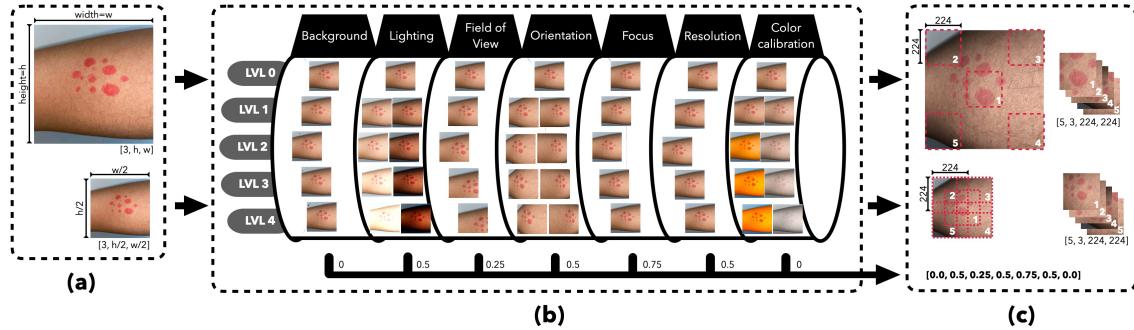


Figure 4.1: Distortion pipeline for generating training images with varying levels of distortion.

The good quality images selected from the SCIN (Ward et al., 2024) dataset are processed, where each image is first resized to a target resolution of 512 pixels while maintaining the aspect ratio. This resizing is crucial to standardize the input size and improve the consistency of feature extraction across different images. For each image, a downsampled version is created. The downsampled image serves as a means to generate hard negative examples, enhancing the learning process.

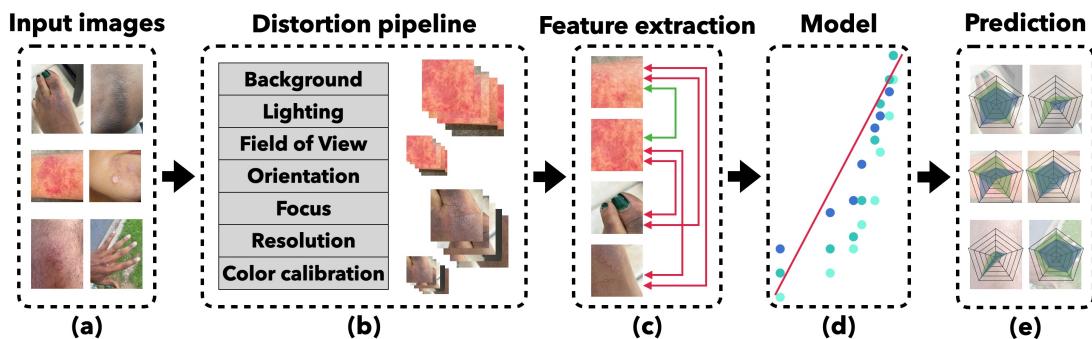
The images then undergo a series of distortions through a defined pipeline, as mentioned in Section 4.2, which applies a sequence of distortion functions. These functions are chosen based on a Gaussian distribution with a mean of 0 and a standard deviation of 2.5, ensuring a probabilistic variation and to add a realistic randomness in the severity of applied distortions. The distortion values are mapped to a uniform scale from 0 to 1, representing the intensity of the distortion from none to extreme. This mapping is crucial for training the model to recognize and quantify these distortions accurately.

Table 4.1: Mapping of numerical values to levels of distortion.

0	0.25	0.5	0.75	1
No distortion applied	Slight distortion	Moderate distortion	Significant distortion	Extreme distortion

After the distortions are applied, both the original and downsampled images are processed to extract smaller crops from significant areas such as the center and the four corners. This step is vital as it ensures that the model learns from various perspectives within the same image, enhancing the robustness of the learned features. The crops are then stacked to form a batch of images ready for feature extraction. But first, the images are normalized using standard values (mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225]) to match the input requirements of the neural network used in SimCLR. This normalization helps in stabilizing the training process by scaling the input features to a common range.

The prepared images are then passed through the SimCLR network to extract features. The network maximizes the similarity of embeddings from images degraded in a similar manner while maximizing the dissimilarity with embeddings from different images or from downsampled versions, which serve as hard negative examples.

**Figure 4.2:** Architecture.

4.4 Feature Extraction with SimCLR

Once the crops from the distorted and downsampled images have been generated, they are processed through the SimCLR framework implemented within the ARNIQA system to extract meaningful features. SimCLR, a pivotal component of ARNIQA, operates based on a self-supervised contrastive learning approach. This framework is adept at maximizing the similarity of features between images that have been distorted in a similar manner, regardless of their initial content.

The implementation of the SimCLR framework from ARNIQA forms a pivotal part of the feature extraction process for this project. SimCLR stands out due to its robust self-supervised learning capabilities, particularly suited for handling the complex distortions applied to the images in this study.

Overview of SimCLR in ARNIQA:

SimCLR operates by maximizing the similarity between embeddings of differently cropped images from the same source but subjected to identical distortions, while also minimizing their similarity to embeddings from differently distorted crops. This methodology effectively trains the model to recognize and quantify similar degradation patterns across diverse image content, leveraging a contrastive loss function that is critical for learning detailed distortion characteristics.

Using the mean and std of Imagenet is a common practice. They are calculated based on millions of images. If you want to train from scratch on your own dataset, you can calculate the new mean and std. Otherwise, using the Imagenet pretrianed model with its own mean and std is recommended.

Training Strategy with SimCLR:

The training strategy involves processing pairs of images through a series of steps:

Image Pair Selection: Two pristine images are selected randomly from the dataset. **Distortion Application:** Each image is cropped and subjected to the same distortion, based on predefined distortion compositions. **Embedding Comparison:** The embeddings of the distorted image crops are compared to ensure that similar distortions produce similar embeddings. This process is visualized in Figure 3 from the ARNIQA paper, which illustrates how the embeddings of similarly degraded image crops are aligned, while those of differently degraded crops serve as hard negative examples. This setup is crucial for training the model to discriminate subtle differences in degradation, enhancing its sensitivity to variations in image quality.

Hard Negative Examples:

To enhance the model's ability to discern subtle degradation differences, images are also down-scaled before cropping and distorting, creating hard negative examples. These examples are crucial for training because they share similar content but differ in distortion due to the downsampling process, challenging the model to fine-tune its discrimination capabilities.

Reproducibility and Practical Implementation:

The availability of code and weights for the SimCLR implementation in ARNIQA ensures that this methodology is not only reproducible but also accessible for future research. This open access to resources supports ongoing improvements and adaptations in the field, fostering further developments in image quality assessment for teledermatology.

Benefits of SimCLR in Image Quality Assessment:**Utilizing SimCLR provides significant advantages:**

Unlabeled Data Usage: It can effectively learn from unlabeled data, which is particularly beneficial in scenarios where high-quality labeled datasets are scarce. **Robust Feature Extraction:** By learning to identify and differentiate between various image distortions, SimCLR helps in building a robust model capable of accurate image quality assessment. **Enhanced Learning from Distortions:** The framework's focus on maximizing similarity for similarly distorted images and maximizing dissimilarity for differently distorted images (even when content is similar) enables the model to learn a comprehensive distortion manifold. These capabilities make SimCLR an ideal choice for this project, addressing the challenges of assessing image quality in teledermatology where diverse and subtle image distortions can significantly impact diagnostic accuracy. The implementation details and insights gained from using SimCLR are foundational for the subsequent stages of model training and validation discussed in the following sections.

4.4.1 Framework Details

4.5 Regression Model Training and Validation

text

4.5.1 Model Selection and Training

4.5.2 Performance Metrics

4.6 Model Testing

The final model is tested against the labeled test set to evaluate its performance in real-world scenarios. Plots illustrating the model's performance across various quality criteria will be shown, highlighting areas where the model performs well or where there is significant variance, indicating uncertainty in quality assessment.

4.6.1 Testing with Labeled Test Set

Chapter 5

Results and Analysis

Realisierung

Dies ist das Hauptkapitel Ihrer Arbeit! Hier wird die Umsetzung der eigenen Ideen und Konzepte (Kapitel 3) anhand der gewählten Methoden (Kapitel 4) beschrieben, inkl. der dabei aufgetretenen Schwierigkeiten und Einschränkungen.

Evaluation und Validation

Auswertung und Interpretation der Ergebnisse. Nachweis, dass die Ziele erreicht wurden, oder warum welche nicht erreicht wurden.

Chapter 6

Discussion and Conclusion

text

Bibliography

- Agnolucci, L., Galteri, L., Bertini, M., & Del Bimbo, A. (2023, November 4). *ARNIQA: Learning Distortion Manifold for Image Quality Assessment*. arXiv: 2310.14918 [cs]. Retrieved April 23, 2024, from <http://arxiv.org/abs/2310.14918>
- Ahsan, M. M., Uddin, M. R., & Luna, S. A. (2022, June 3). *Monkeypox Image Data collection*. arXiv: 2206.01774 [cs, eess]. Retrieved April 28, 2024, from <http://arxiv.org/abs/2206.01774>
- Chandler, D. M. (2010). Most apparent distortion: Full-reference image quality assessment and the role of strategy. *Journal of Electronic Imaging*, 19(1), 011006. <https://doi.org/10.1117/1.3267105>
- Chandler, D., & Hemami, S. (2007). VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images. *IEEE Transactions on Image Processing*, 16(9), 2284–2298. <https://doi.org/10.1109/TIP.2007.901820>
- Daneshjou, R., Vodrahalli, K., Novoa, R. A., Jenkins, M., Liang, W., Rotemberg, V., Ko, J., Swetter, S. M., Bailey, E. E., Gevaert, O., Mukherjee, P., Phung, M., Yekrang, K., Fong, B., Sahasrabudhe, R., Allerup, J. A. C., Okata-Karigane, U., Zou, J., & Chiou, A. S. (2022). Disparities in dermatology AI performance on a diverse, curated clinical image set. *Science Advances*, 8(32), eabq6147. <https://doi.org/10.1126/sciadv.abq6147>
- Ghadiyaram, D., & Bovik, A. C. (2016). Massive Online Crowdsourced Study of Subjective and Objective Picture Quality. *IEEE Transactions on Image Processing*, 25(1), 372–387. <https://doi.org/10.1109/TIP.2015.2500021>
- Groh, M., Harris, C., Soenksen, L., Lau, F., Han, R., Kim, A., Koochek, A., & Badri, O. (2021, April 20). *Evaluating Deep Neural Networks Trained on Clinical Images in Dermatology with the Fitzpatrick 17k Dataset*. arXiv: 2104.09957 [cs]. Retrieved April 28, 2024, from <http://arxiv.org/abs/2104.09957>
- Gu, K., Xu, X., Qiao, J., Jiang, Q., Lin, W., & Thalmann, D. (2020). Learning a Unified Blind Image Quality Metric via On-Line and Off-Line Big Training Instances. *IEEE Transactions on Big Data*, 6(4), 780–791. <https://doi.org/10.1109/TBDA.2019.2895605>
- Gu, K., Zhai, G., Yang, X., & Zhang, W. (2014). Hybrid No-Reference Quality Metric for Singly and Multiply Distorted Images. *IEEE Transactions on Broadcasting*, 60(3), 555–567. <https://doi.org/10.1109/TBC.2014.2344471>
- Hoffmann, C. P., Lennerts, S., Schmitz, C., Stölzle, W., & Uebelnickel, F. (Eds.). (2016). *Business Innovation: Das St. Galler Modell*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-07167-7>
- Jalaboi, R., Winther, O., & Galimzianova, A. (2023, January 23). *Explainable Image Quality Assessments in Teledermatological Photography*. arXiv: 2209.04699 [cs]. Retrieved April 23, 2024, from <http://arxiv.org/abs/2209.04699>
- Jayaraman, D., Mittal, A., Moorthy, A. K., & Bovik, A. C. (2012). Objective quality assessment of multiply distorted images. *2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, 1693–1697. <https://doi.org/10.1109/ACSSC.2012.6489321>

- Kawahara, J., Daneshvar, S., Argenziano, G., & Hamarneh, G. (2019). Seven-Point Checklist and Skin Lesion Classification Using Multitask Multimodal Neural Nets. *IEEE Journal of Biomedical and Health Informatics*, 23(2), 538–546. <https://doi.org/10.1109/JBHI.2018.2824327>
- Ma, K., Duanmu, Z., Wu, Q., Wang, Z., Yong, H., Li, H., & Zhang, L. (2017). Waterloo Exploration Database: New Challenges for Image Quality Assessment Models. *IEEE Transactions on Image Processing*, 26(2), 1004–1016. <https://doi.org/10.1109/TIP.2016.2631888>
- Min, X., Ma, K., Gu, K., Zhai, G., Wang, Z., & Lin, W. (2017). Unified Blind Quality Assessment of Compressed Natural, Graphic, and Screen Content Images. *IEEE Transactions on Image Processing*, 26(11), 5462–5474. <https://doi.org/10.1109/TIP.2017.2735192>
- Ni, Z., Ma, L., Zeng, H., Chen, J., Cai, C., & Ma, K.-K. (2017). ESIM: Edge Similarity for Screen Content Image Quality Assessment. *IEEE Transactions on Image Processing*, 26(10), 4818–4831. <https://doi.org/10.1109/TIP.2017.2718185>
- Pacheco, A. G., & Krohling, R. A. (2020). The impact of patient clinical information on automated skin cancer detection. *Computers in Biology and Medicine*, 116, 103545. <https://doi.org/10.1016/j.combiomed.2019.103545>
- Ponomarenko, N., Jin, L., Ieremeiev, O., Lukin, V., Egiazarian, K., Astola, J., Vozel, B., Chehdi, K., Carli, M., Battisti, F., & Jay Kuo, C.-C. (2015). Image database TID2013: Peculiarities, results and perspectives. *Signal Processing: Image Communication*, 30, 57–77. <https://doi.org/10.1016/j.image.2014.10.009>
- Ponomarenko, N., Lukin, V., Zelensky, A., Egiazarian, K., Astola, J., Carli, M., & Battisti, F. (2009). TID2008 – A Database for Evaluation of Full- Reference Visual Quality Assessment Metrics.
- Sheikh, H., Sabir, M., & Bovik, A. (2006). A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms. *IEEE Transactions on Image Processing*, 15(11), 3440–3451. <https://doi.org/10.1109/TIP.2006.881959>
- Sun, W., Zhou, F., & Liao, Q. (2017). MDID: A multiply distorted image database for image quality assessment. *Pattern Recognition*, 61, 153–168. <https://doi.org/10.1016/j.patcog.2016.07.033>
- Virtanen, T., Nuutinen, M., Vaahteranoksa, M., Oittinen, P., & Hakkinen, J. (2015). CID2013: A Database for Evaluating No-Reference Image Quality Assessment Algorithms. *IEEE Transactions on Image Processing*, 24(1), 390–402. <https://doi.org/10.1109/TIP.2014.2378061>
- Vodrahalli, K., Daneshjou, R., Novoa, R. A., Chiou, A., Ko, J. M., & Zou, J. (2020, October 1). *TruelImage: A Machine Learning Algorithm to Improve the Quality of Telehealth Photos*. arXiv: 2010.02086 [cs, eess]. Retrieved April 23, 2024, from <http://arxiv.org/abs/2010.02086>
- Ward, A., Li, J., Wang, J., Lakshminarasimhan, S., Carrick, A., Campana, B., Hartford, J., S, P. K., Tiyasirichokchai, T., Virmani, S., Wong, R., Matias, Y., Corrado, G. S., Webster, D. R., Siegel, D., Lin, S., Ko, J., Karthikesalingam, A., Semturs, C., & Rao, P. (2024, February 28). *Crowdsourcing Dermatology Images with Google Search Ads: Creating a Real-World Skin Condition Dataset*. arXiv: 2402.18545 [cs]. Retrieved April 28, 2024, from <http://arxiv.org/abs/2402.18545>
- Wu, X., Wen, N., Liang, J., Lai, Y.-K., She, D., Cheng, M.-M., & Yang, J. (2019). Joint Acne Image Grading and Counting via Label Distribution Learning. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 10641–10650. <https://doi.org/10.1109/ICCV.2019.01074>
- Yang, H., Yuming Fang, Lin, W., & Wang, Z. (2014). Subjective quality assessment of Screen Content Images. *2014 Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*, 257–262. <https://doi.org/10.1109/QoMEX.2014.6982328>

Appendix A

Supplementary Material

The following pages contain the supplementary material for this thesis. This section includes documents specific to project planning and management. The documents are attached in this order:

- Project Assignment
- Risk Management
- Project Planning

Documents and code relevant to the thesis can be downloaded from the following link:

<https://github.com/Schoggi-Mimi/bachelor-thesis>.

Aufgabenstellung

Modul:	Dept I BAA FS24
Titel:	Automated Image Quality Assessment in Teledermatology
Ausgangslage und Problemstellung:	ABIZ has been researching artificial intelligence applications in dermatology for the past decade with the objective to develop decision support systems to effectively support clinical practice. In collaboration with the University Hospital of Basel and the Swiss company Derma2go, we are tackling the issue of automatically assessing the quality of patient images for diagnosis, since this factor heavily impacts the effectiveness of teledermatology workflows.
Ziel der Arbeit und erwartete Resultate:	The objective of this work is to conduct an extensive review of state-of-the-art quality assessment methods in the general image domain and evaluate how they can be applied to teledermatology. The project deliverables include: <ul style="list-style-type: none"> - A comprehensive review of state-of-the-art image quality assessment methods. - A review of image quality criteria for teledermatology diagnosis. - An evaluation of selected quality assessment methods on public dermatology datasets. - A well-written repository enabling to reproduce reported results and assess the quality of new patient images.
Gewünschte Methoden, Vorgehen:	The project will start with a literature review of existing quality assessment methods and patient image quality criteria in dermatology. Together with the supervisor, adapted methods will be selected, which the student will then evaluate on public dermatology datasets. The student will present his work to the supervisor on bi-weekly meetings. One day before the meeting, the student will share a 1-page document describing in bullet points: <ul style="list-style-type: none"> - What work was performed during the last reporting period. - What work is planned for the next period. - Project status, comparison with planning, reasons for deviations if applicable. - Top three risks incl. planned measures. For the meeting, the student will prepare slides to present these information in more details.
Kreativität, Methoden, Innovation:	This thesis will encourage innovative approaches, including but not limited to proposing new metrics and relevant changes to adapt methods to the teledermatology context. The student will have the opportunity to fine-tune deep learning models on public dermatology datasets and work closely with both clinicians and researchers from ABIZ and the partner institutions.
Sonstige Bemerkungen:	Candidates should have a strong background in computer science. Prior experience with medical imaging or teledermatology is beneficial but not mandatory. The project will require a creative approach to problem-solving and an eagerness to work in interdisciplinary teams.

Projektteam

Student:in 1:	Choekyel Nyungmartsang
Betreuer:in:	Dr. Ludovic Amruthalingam

Auftraggeber

Firma:	Algorithmic Business Research Lab
Ansprechperson:	Dr. Ludovic Amruthalingam
Funktion:	
Strasse:	
PLZ/Ort:	
Telefon:	+41 41 349 30 74
E-Mail:	ludovic.amruthalingam@hslu.ch
Website:	

Version 13.06.2023 / bcl

APPENDIX A. SUPPLEMENTARY MATERIAL

include
risk man-
age-
ment and
project
planning
pdfs

Appendix B

Dataset

Detailed information on image quality assessment (IQA) databases:

- **LIVE** (Laboratory for Image & Video Engineering) dataset (Sheikh et al., 2006) includes 29 reference images and 779 manually distorted images corrupted by 5 types of distortions: JPEG compression (JPEG), JPEG2000 compression (JP2K), white noise (WN), Gaussian blur (GB), and simulated fast fading Rayleigh channel (FF). Each distortion type contains 5 or 4 distortion levels. Most images are 768×512 pixels in size. Each distorted image in this dataset is associated with a Differential Mean Opinion Score (DMOS), scaled from 0 to 100, where 0 indicates no perceivable distortion.
- **TID2008** (Tampere image database 2008) dataset (Ponomarenko et al., 2009) includes 25 reference images and 1700 distorted images corrupted by 17 types of distortions, with 4 levels for each distortion type. All images have a fixed resolution of 512×384 . This dataset provides MOS values and their standard deviations, with MOS ranging from 0 to 9, where 9 signifies a distortion-free image.
- **TID2013** (Tampere image database 2013) dataset (Ponomarenko et al., 2015) is extended from TID2008 (Ponomarenko et al., 2009) by increasing the number of distortion levels to 5, and the number of distortion types to 24. Therefore, 3000 distorted images are generated from 25 pristine images. The subjective testing and data processing steps are similar to that of TID2008. DMOS values for this dataset were derived from over half a million ratings given by nearly a thousand observers, with values ranging from 0 to 9, where higher values denote poorer image quality.
- **CSIQ** (Categorical subjective image quality (CSIQ) database) (D. M. Chandler, 2010) contains 30 pristine images and 866 distorted images corrupted by JPEG, JP2K, WN, GB, additive pink Gaussian noise, and global contrast decrements, with 5 or 4 levels for each distortion type. The resolution is 512×512 . Each image in CSIQ is associated with DMOS values obtained from subjective ratings by 25 testers, with DMOS values scaled from 0 to 1, where higher values indicate worse quality.
- **A57** (D. Chandler & Hemami, 2007) includes 3 pristine images and 54 distorted images corrupted by 6 types of distortions, with 3 levels for each distortion type. All images are in gray scale. The resolution is 512×512 .
- **WED** (Waterloo exploration database) (Ma et al., 2017) includes 4744 pristine natural images and 94880 distorted images corrupted by JPEG, JP2K, GB, and WN, with 5 levels

for each distortion type. The images have various resolutions. No human opinion score is provided, but the authors introduce several alternative test criteria to evaluate the IQA models.

Multiple Distortions IQA Databases

- **LIVEMD** (LIVE multiply distorted) (Jayaraman et al., 2012) database consists of 15 reference images and 405 multiply distorted images. The database includes one/double-fold artifacts. Each multiply distorted image is corrupted under two multiple distortion scenarios: Gaussian blur followed by JPEG and Gaussian blur followed by white noise. All images have a resolution of 1280×720 . DMOS values for each distorted image range from 0 to 100.
- **Multiply distorted image database 2013 (MDID2013)** (Gu et al., 2014): MDID2013 has a total of 12 pristine images and 324 distorted images. Each pristine image is corrupted successively by Gaussian blur, white noise, and JPEG. The images have resolutions of 768×512 or 1280×720 .
- **Multiply distorted image database 2016 (MDID2016)** (Sun et al., 2017): MDID2016 consists of 20 reference images and 1600 distorted images. Five distortion types are introduced, i.e., white noise, Gaussian blur, JPEG, JPEG2000, and contrast change (CC). The order of distortions is as follows: Gaussian blur or CC first, JPEG or JPEG2000 second, and white noise last. All distorted images are with random types and levels of distortions. The image resolution is 512×384 .

Screen Content IQA Databases

- **Screen Image Quality Assessment Database (SIQAD)** (Yang et al., 2014): SIQAD includes 20 pristine and 980 distorted screen content images (SCIs). Distortion types include white noise (WN), Gaussian blur (GB), color cast (CC), JPEG, JPEG2000 (JP2K), motion blur (MB), and layer segmentation-based compression, with 7 levels for each type. The images have various resolutions near 700×700 .
- **Screen Content Image Quality (SCIQ) Database** (Ni et al., 2017): SCIQ consists of 40 pristine and 1800 distorted SCIs corrupted by 9 types of distortions, including WN, GB, MB, CC, JPEG, JP2K, color saturation change (CSC), color quantization with dithering (CQD), and the screen content coding extension of High Efficiency Video Coding (HEVC-SCC). Five distortion levels are considered. The resolution is fixed at 1280×720 .
- **Cross-Content-Type (CCT) Database** (Min et al., 2017): CCT is constructed to conduct cross-content-type IQA research. CCT consists of 72 pristine and 1320 distorted natural scene images (NSIs), computer graphic images (CGIs), and SCIs. Two distortion types are considered, i.e., HEVC and HEVC-SCC coding, with 11 distortion levels for each type. The image resolution is either 1920×1080 or 1280×720 .
- **Hybrid Screen Content and Natural Scene Image Database (HSNID)** (Gu et al., 2020): HSNID has 10 pristine NSIs and 10 pristine SCIs, and 600 distorted NSIs and SCIs corrupted by WN, GB, MB, CC, JPEG, and JP2K, with 5 distortion levels for each type.

Authentic Distortions IQA Databases

- **LIVE in the wild image quality challenge database** (Ghadiyaram & Bovik, 2016) includes 1162 authentically distorted images captured using a variety of mobile devices. Complex real distortions, which are not well-modeled by the synthetic distortions are included. All images are cropped to the resolution of 500×500 . A novel crowdsourcing system was employed to gather over 350,000 opinion scores from 8100 observers, ensuring the objectivity of the MOS values obtained.

- **Camera image database (CID2013)** (Virtanen et al., 2015): CID2013 is designed to test no-reference IQA algorithms. It includes 480 real images captured from 8 typical scenes using 79 consumer cameras and mobile phones. The images are rated from 5 aspects: the overall quality, sharpness, graininess, lightness, and color saturation scales. The images are scaled to a size of 1600×1200 .

Appendix C

Degradation Types

As mentioned in Subsection 2.2.3, the dataset used in this thesis is augmented with synthetic degradations. The following figures Figure C.1, Figure C.2, Figure C.3, Figure C.4, Figure C.5, Figure C.6, Figure C.7 show the different levels of intensity for the degradations of each distortion group.

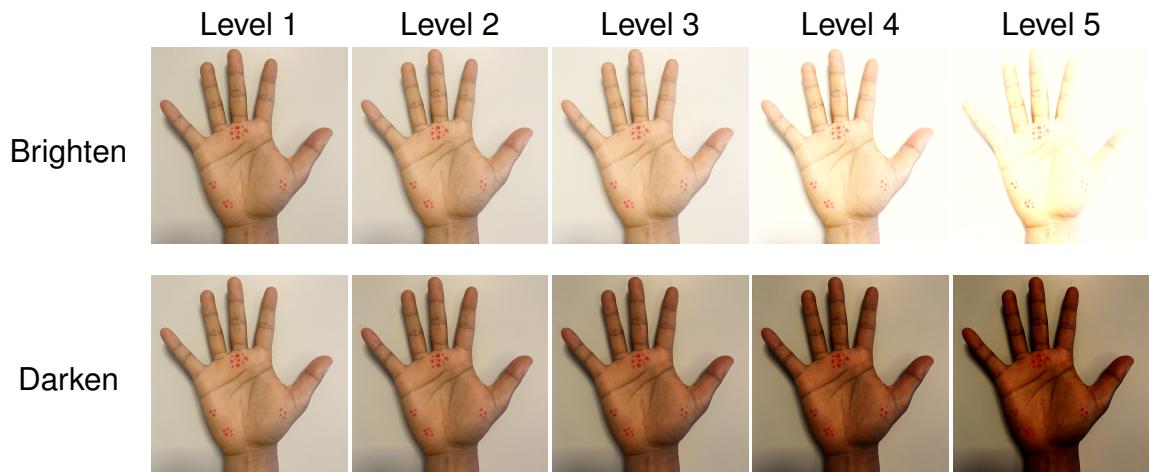


Figure C.1: Visualization of the degradation types belonging to the *Brightness change* group for increasing levels of intensity.

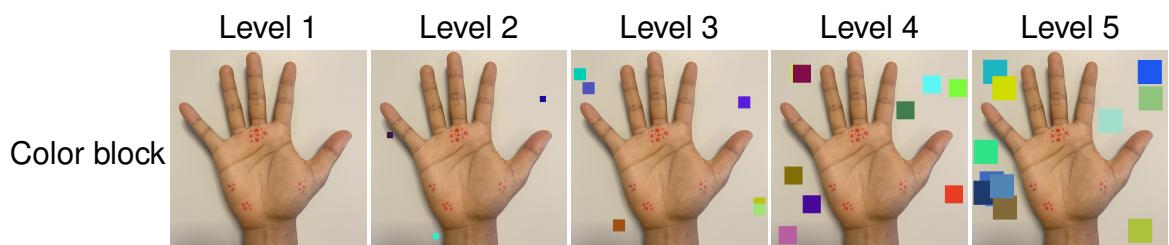


Figure C.2: Visualization of the degradation types belonging to the *Background color* group for increasing levels of intensity.

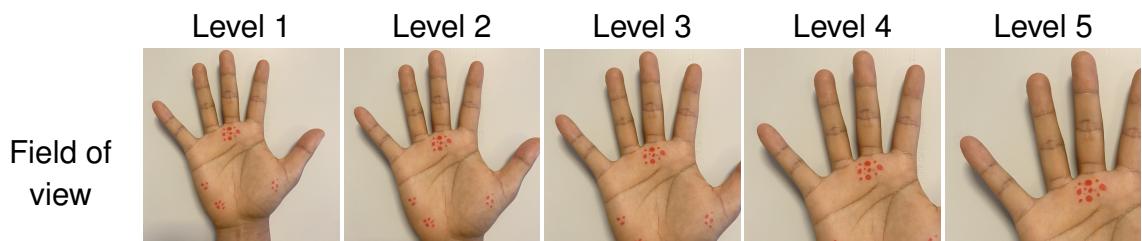


Figure C.3: Visualization of the degradation types belonging to the *Field of View* group for increasing levels of intensity.

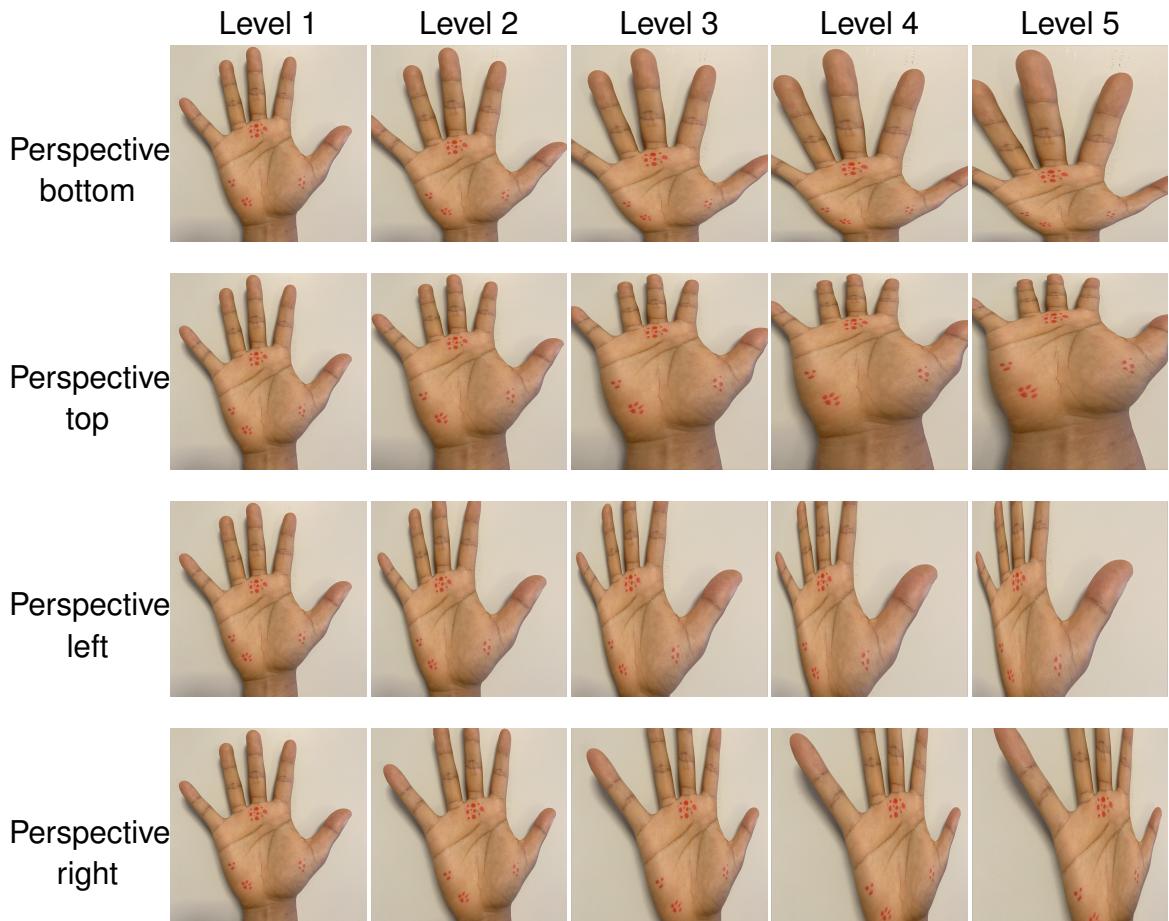


Figure C.4: Visualization of the degradation types belonging to the *Image orientation* group for increasing levels of intensity.

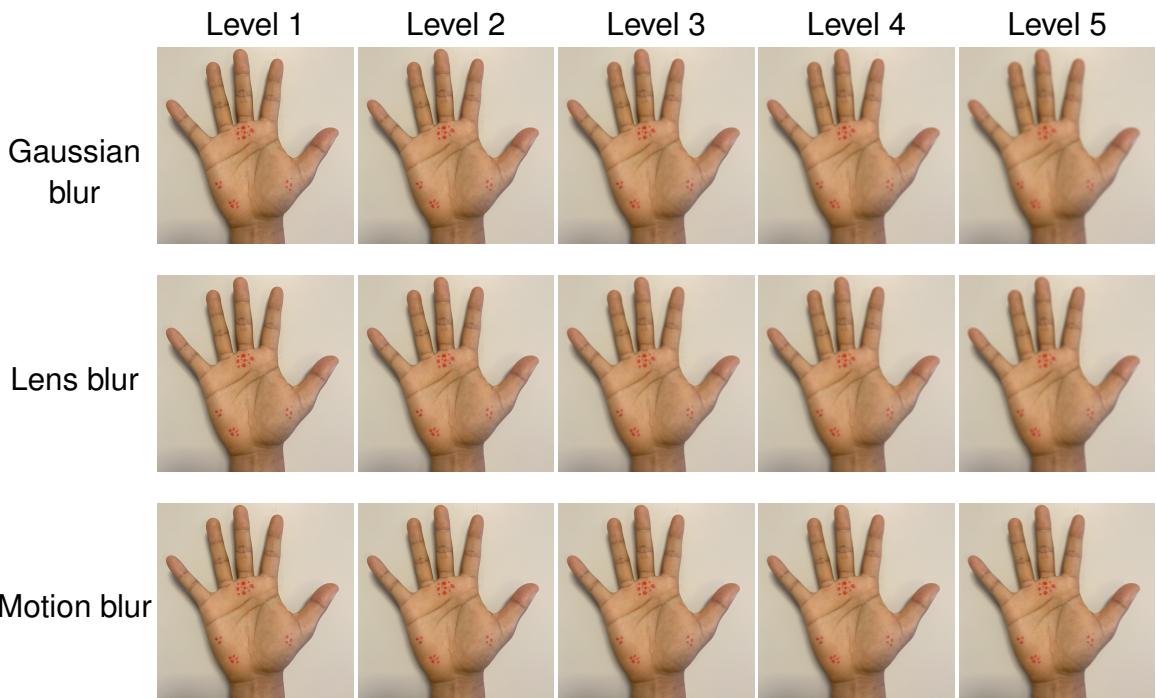


Figure C.5: Visualization of the degradation types belonging to the *Focus* group for increasing levels of intensity.

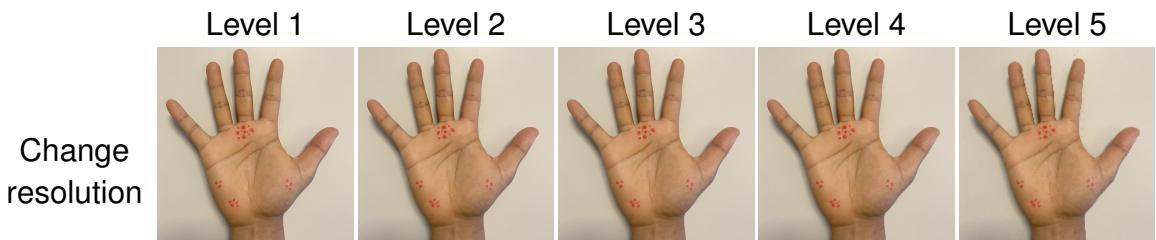


Figure C.6: Visualization of the degradation types belonging to the *Resolution* group for increasing levels of intensity.

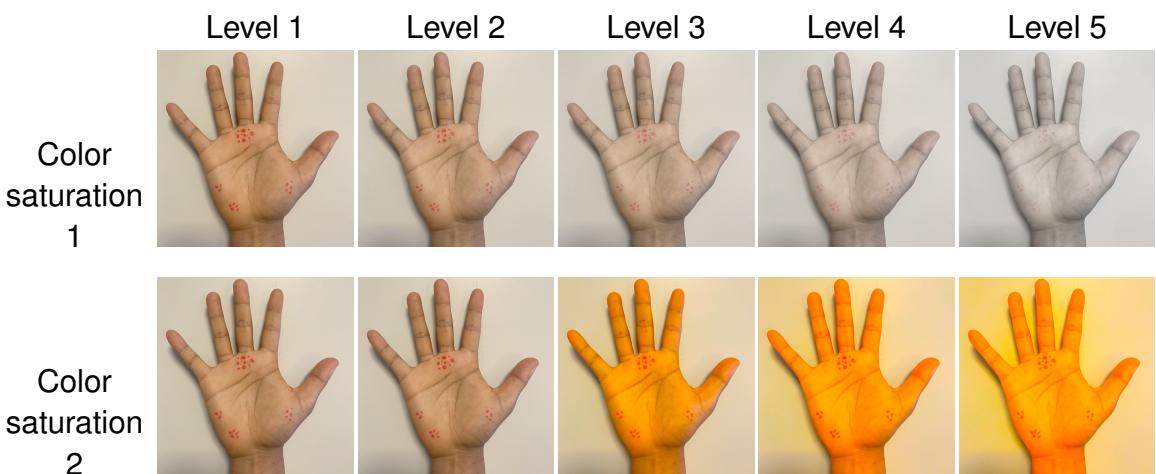


Figure C.7: Visualization of the degradation types belonging to the *Color calibration* group for increasing levels of intensity.

Appendix D

Code

Anhang, Abkürzungs-, Abbildungs-, Tabellen-, Formel-Verzeichnis, Literaturverzeichnis nicht vergessen!

Anhänge

Projektspezifisch können weitere Dokumentationsteile angefügt werden wie:

Aufgabenstellung, Projektmanagement-Plan/Bericht, Testplan/Testbericht, Bedienungsanleitungen, Details zu Umfragen, detaillierte Anforderungslisten, Referenzen auf projektspezifische Daten in externen Entwicklungs- und Datenverwaltungstools etc.

Listing D.1: Caption on PDF

```
import numpy as np
```