

Contents

1 Results and Analysis	2
1.1 Range of Distortion Values	2
1.2 Model Performance	3
1.2.1 Parallel Coordinate Plot	3
1.2.2 Performance Metrics	3
1.2.3 Confusion Matrices	5
1.3 Model Predictions	6
1.3.1 Visualizations for Synthetic Distorted Images	6
1.3.2 Visualizations for Authentic Images	6
1.4 Assessing Training and Testing Images Quality	9
1.4.1 Training Images Quality	9
1.4.2 Test Images Quality	11
2 Discussion	12
2.1 Interpretation of Results	12
2.2 Answering the Research Questions	13
2.3 Comparison with Related Work	13
2.4 Reflection	13
2.5 AI Tools Used	13
3 Conclusion and Future Work	14
A Supplementary Material	III

Chapter 1

Results and Analysis

In this chapter, the performance of the trained models is analyzed and discussed. The main focus is on the final MLP regressor model, which was found to be the best performing model across multiple criteria. The selection of the MLP regressor is supported by the results shown in Figure 1.1, which presents a parallel coordinate plot comparing the best-performing models across all seven criteria and the overall SRCC.

The parallel coordinate plot shows that the MLP regressor consistently performs better than the other models across all criteria. Although all models perform similarly for different criteria, no model stands out in a specific criterion. This could be because the same features were used for every criterion, with only the target labels differing.

In addition to the parallel coordinate plot, Table 1.1 summarizes the cross-dataset evaluation results, showing the generalizability of the models. The table lists the models in the left column and their evaluation results on the SCIN and Fitzpatrick (F17K) datasets in the right columns. This table shows how well the models, trained on one dataset, perform when evaluated on another, providing insights into their robustness and adaptability. All data were synthetically distorted through the pipeline to ensure consistent evaluation conditions.

Given these findings, the MLP regressor was chosen as the final model for further testing. The following sections will detail the performance of the MLP regressor on the test images, providing a clear analysis of its strengths and weaknesses in assessing image quality in teledermatology.

1.1 Range of Distortion Values

The ranges of values for each distortion type were carefully chosen to reflect realistic scenarios for teledermatology applications. Each distortion type was visualized individually to make sure they were appropriate. Section A.2 includes images that show each criterion with different distortion types and five severity levels.

It is important to note that images should not be normalized before viewing because normalization can make the images appear overly colorful and unrealistic. However, normalization is necessary during training and testing because the feature extraction backbone from ARNIQA(Agnolucci et al., 2023) was trained on ResNet50 with ImageNet images. This step helps the model accurately extract relevant features from the images.

1.2 Model Performance

To fully understand the performance of the four different models with different architectures, a cross-dataset evaluation was done. Table 1.1 summarizes the overall Spearman’s Rank Correlation Coefficient (SRCC) on the synthetically distorted SCIN and Fitzpatrick datasets. This evaluation highlights how well the models generalize across different datasets. Note that the Fitzpatrick dataset is referred to as F17K for simplicity.

Table 1.1: Spearman’s Rank Correlation Coefficient (SRCC) of Different Models on SCIN and F17K Datasets. F17K refers to the Fitzpatrick17k dataset.

Model	SCIN	F17K
Combined MLP Regressor	0.66	0.75
Combined XGB Regressor	0.65	0.73
Combined XGB Classifier	0.58	0.61
Combined MLP Classifier	0.43	0.46
F17K MLP Regressor	0.54	0.69
SCIN MLP Regressor	0.62	0.49
F17K XGB Regressor	0.53	0.67
SCIN XGB Regressor	0.61	0.48
SCIN MLP Classifier	0.53	0.45
F17K MLP Classifier	0.47	0.58
SCIN XGB Classifier	0.54	0.43
F17K XGB Classifier	0.46	0.59

1.2.1 Parallel Coordinate Plot

Figure 1.1, a parallel coordinate presents a visual comparison of the model performances across different criteria and the overall SRCC.

1.2.2 Performance Metrics

The performance of the final MLP regressor model on individual criteria is shown in Table 1.2. This table provides a comprehensive view of the model’s strengths and weaknesses, highlighting areas for improvement. The model was evaluated on 475 good quality Fitzpatrick images that were synthetically distorted using the distortion pipeline.

Table 1.2: Performance Metrics for Each Distortion Criteria

Criteria	MAE	R ²	SRCC	Cohen’s Kappa
Background	0.9684	0.2595	0.5422	0.4399
Lighting	0.5726	0.6440	0.8028	0.7913
Focus	0.4042	0.7385	0.8622	0.8568
Orientation	0.9895	0.1824	0.4735	0.4102
Color calibration	0.4905	0.7334	0.8622	0.8583
Resolution	0.3642	0.7656	0.8722	0.8726
Field of view	0.5474	0.5976	0.7710	0.7660
Overall	0.6195	0.5646	0.7507	0.7396

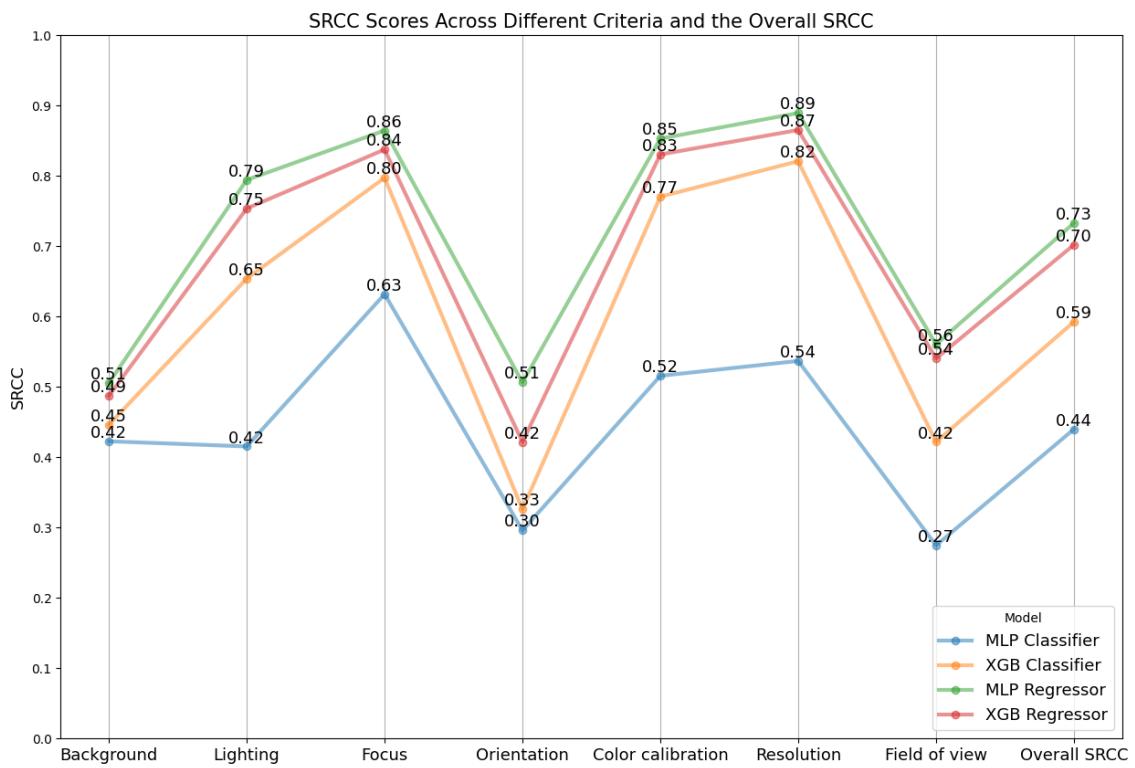


Figure 1.1: Parallel coordinate plot showing the best SRCC values for the four different models across the seven criteria and the overall SRCC. This plot highlights the performance of the MLP Regressor.

1.2.3 Confusion Matrices

In addition to numerical metrics, confusion matrices¹ were created for each criterion, as shown in Figure 1.2. These matrices display where the model makes correct predictions and where it makes mistakes, showing a detailed view of its accuracy for each type of distortion. The comparison between actual and predicted scores helps identify specific areas where the model performs well and areas that need improvement. Furthermore, the confusion matrices also reveal any biases the model might have toward certain severity ranges, indicating whether it tends to predict only low or high severity levels, or if its predictions are skewed in some way.

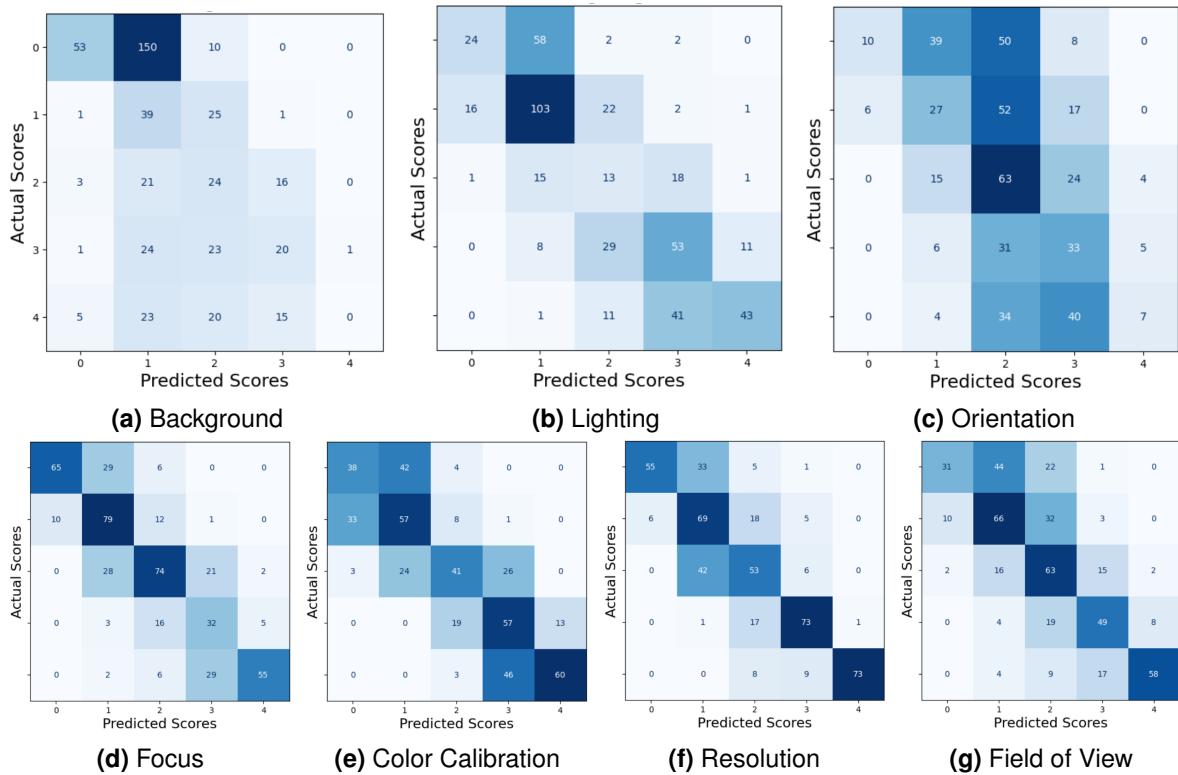


Figure 1.2: Confusion matrices for the MLPRegressor model evaluated on the 475 images from the Fitzpatrick dataset. Each matrix corresponds to a specific distortion criterion and shows the actual scores on the y-axis and the predicted scores on the x-axis. Darker shades indicate higher counts, highlighting where the model's predictions match the actual values and where discrepancies occur.

¹from utils.visualization import plot_all_confusion_matrices

1.3 Model Predictions

To better understand the model's performance on the two test sets (70 synthetic distorted images and 200 authentic images), radar charts² were created. These charts show the criteria on the outside, with severity ranges going from the center (0) to the outer edge (1), indicating high distortion for each criterion. These visualizations provide a clear and simple view of the model's performance, showing its strengths and areas for improvement. They help identify specific cases where the model predicted the correct severity and where it struggled. The radar charts also make it easy to see which distortions the model handles well and which need more attention.

1.3.1 Visualizations for Synthetic Distorted Images

These visualizations, as shown in Figure 1.3, help to compare the model's predictions with the actual distortions introduced by the pipeline. This approach clearly demonstrates the model's ability to handle various types of distortions.

The first column shows the original image, the second displays the distorted image, the third contains the actual labels, and the fourth presents the model's predictions. This setup makes it easy to compare the model's predictions with the actual distortions.

1.3.2 Visualizations for Authentic Images

The visualizations, as shown in Figure 1.4, compare the model's predictions with human-labeled scores. This method highlights the model's performance in real-world scenarios, showing its strengths and areas for improvement.

The first column shows the image, the second column displays the human-labeled scores, and the third column presents the model's predictions. This comparison helps show how well the model's predictions align with the human evaluations.

²from utils.visualization import plot_results

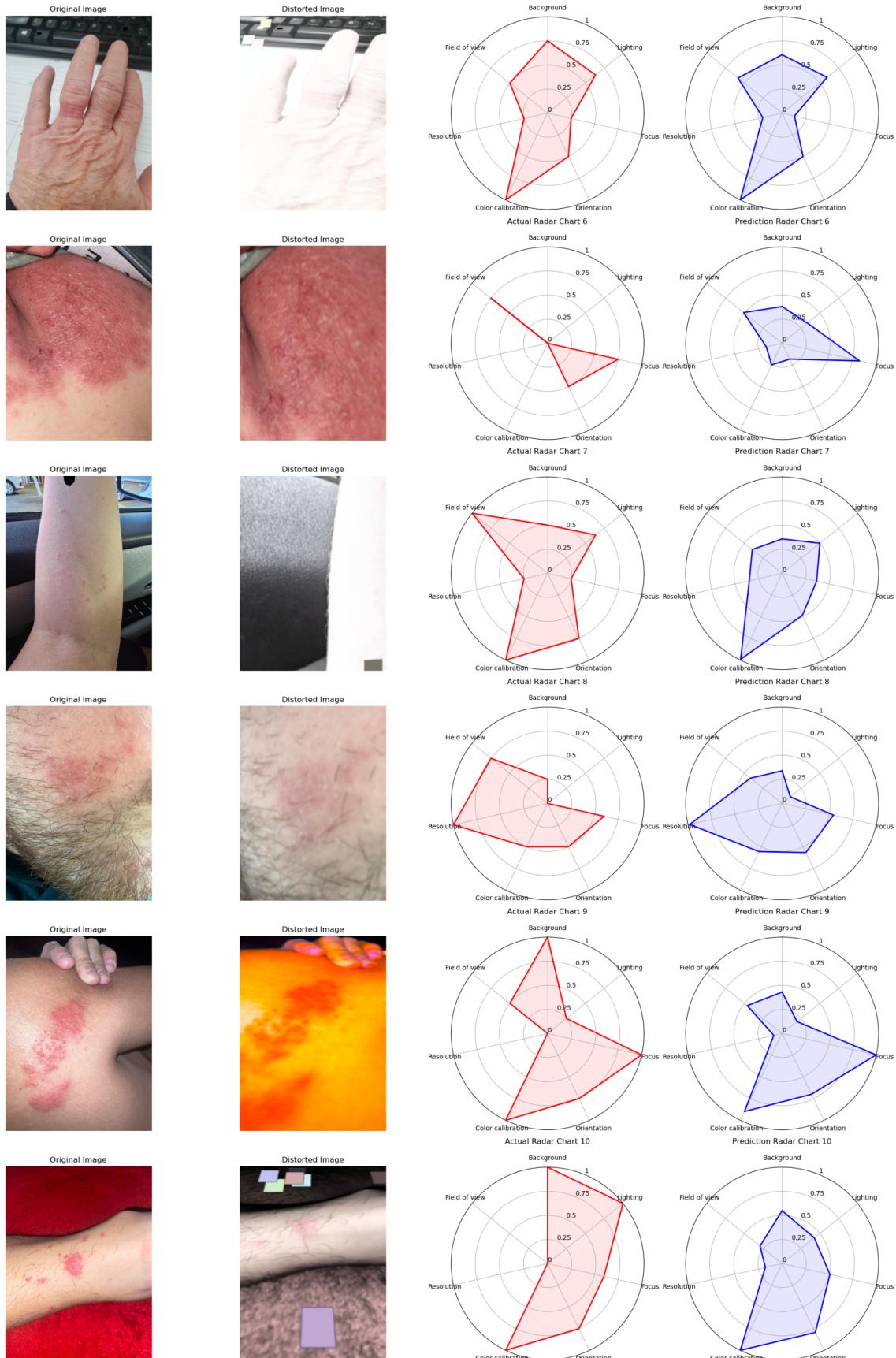


Figure 1.3: Visualizations for the MLP Regressor model on 70 synthetic distorted images. The four-column layout shows the original image, the distorted image, the actual labels, and the model's predictions.

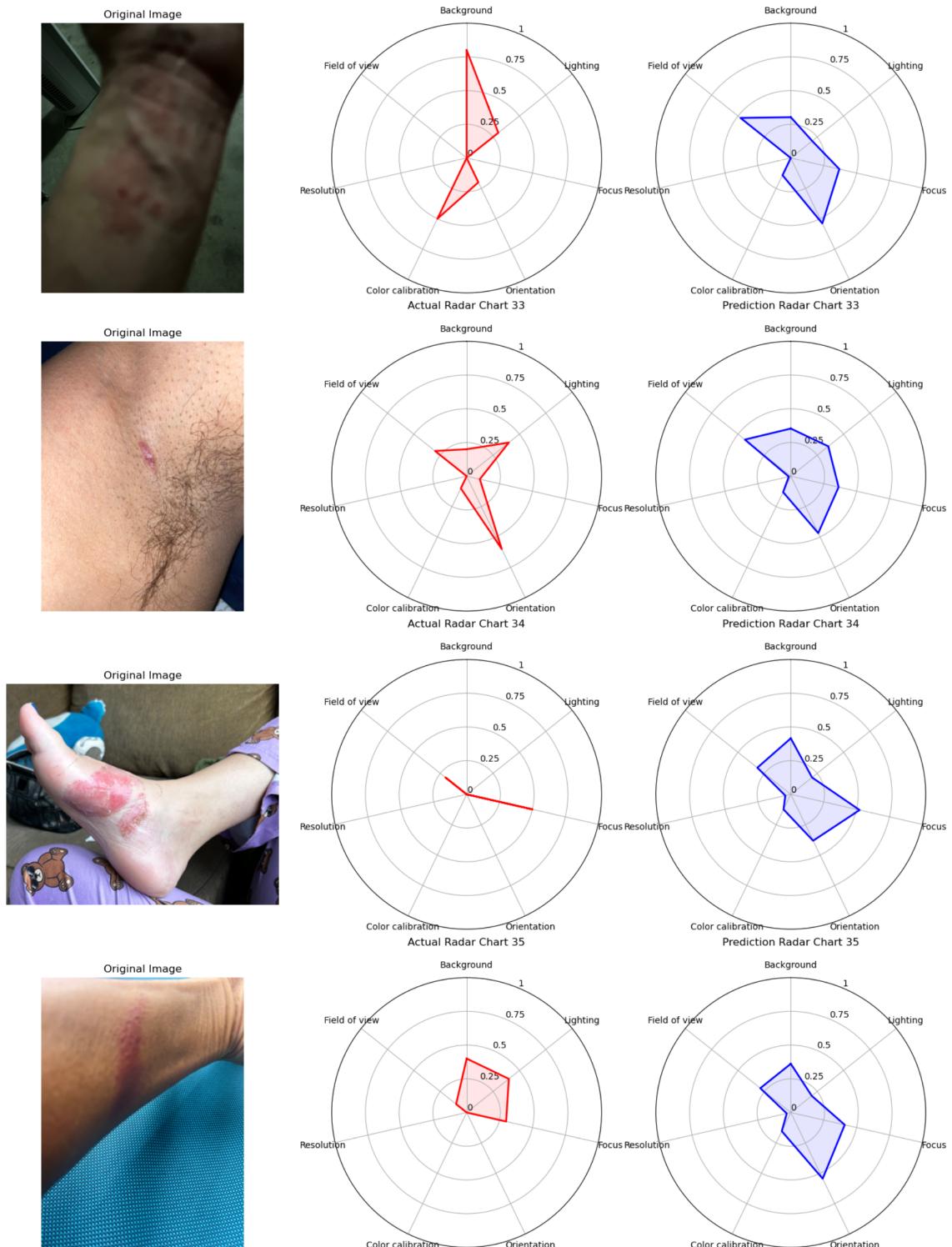


Figure 1.4: Visualizations for the MLP Regressor model on 200 authentic images. The three-column layout shows the image, the human-labeled scores, and the model's predictions.

1.4 Assessing Training and Testing Images Quality

To verify the quality of the images used for training and see how they change after synthetic distortion, radar charts were created. These charts show the quality of the original training images and how they are affected by the distortions. Additionally, the quality of both the synthetic and authentic test images is assessed using the same method. These radar charts show a simple visual representation of the quality and the level of distortion across the seven criteria.

1.4.1 Training Images Quality

SCIN Dataset

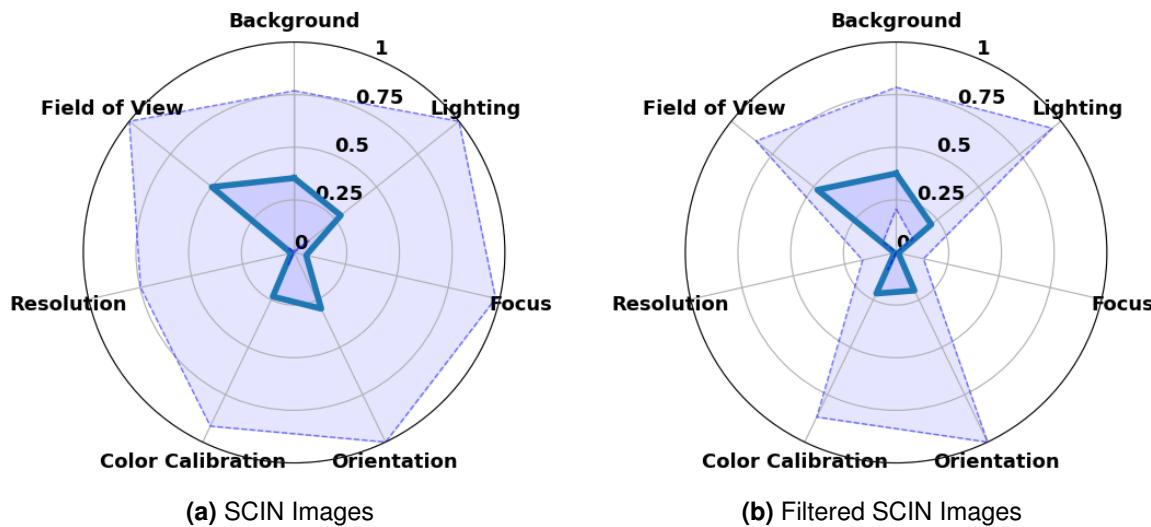


Figure 1.5: Radar charts for the SCIN dataset. (a) Original images from the SCIN dataset (10'379 images). (b) Filtered good quality images (475 images).

Fitzpatrick Dataset

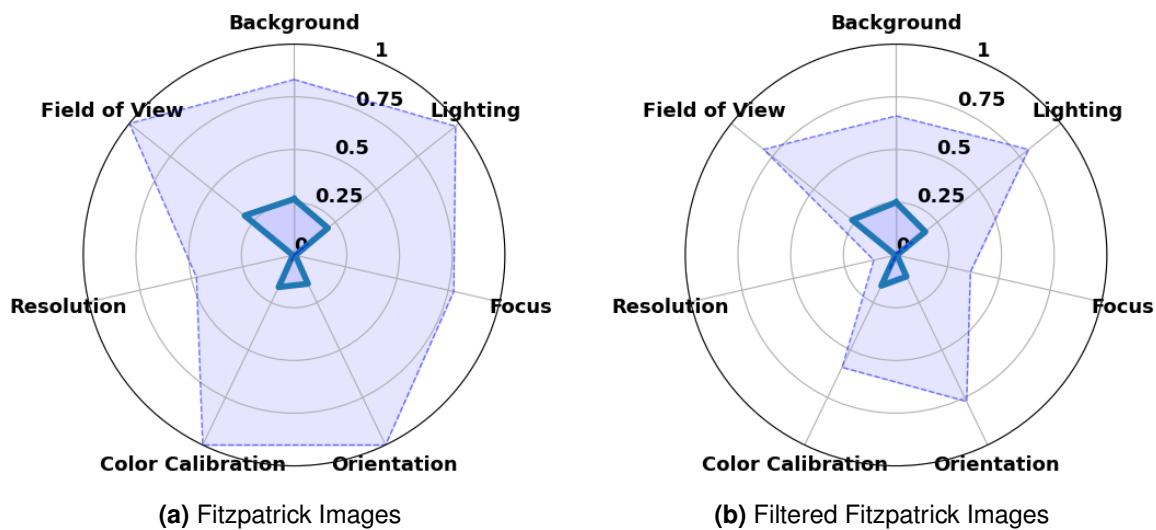


Figure 1.6: Radar charts for the Fitzpatrick dataset. (a) Original images from the Fitzpatrick dataset (16'577 images). (b) Filtered good quality images (475 images).

Combined Dataset

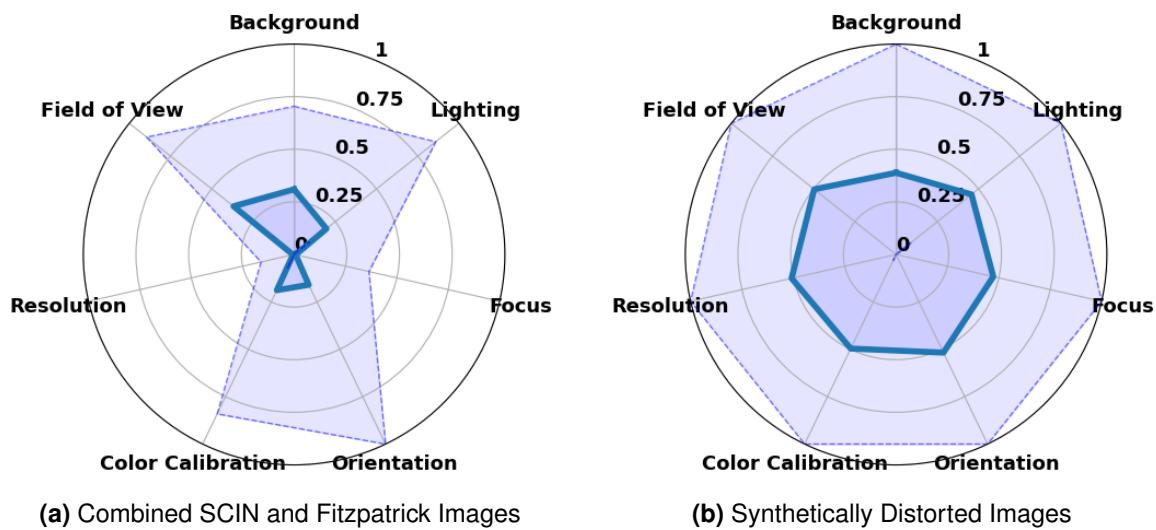


Figure 1.7: Combined dataset analysis. (a) Combined SCIN and Fitzpatrick images (950 images). (b) Synthetically distorted images.

1.4.2 Test Images Quality

Synthetic Test Images

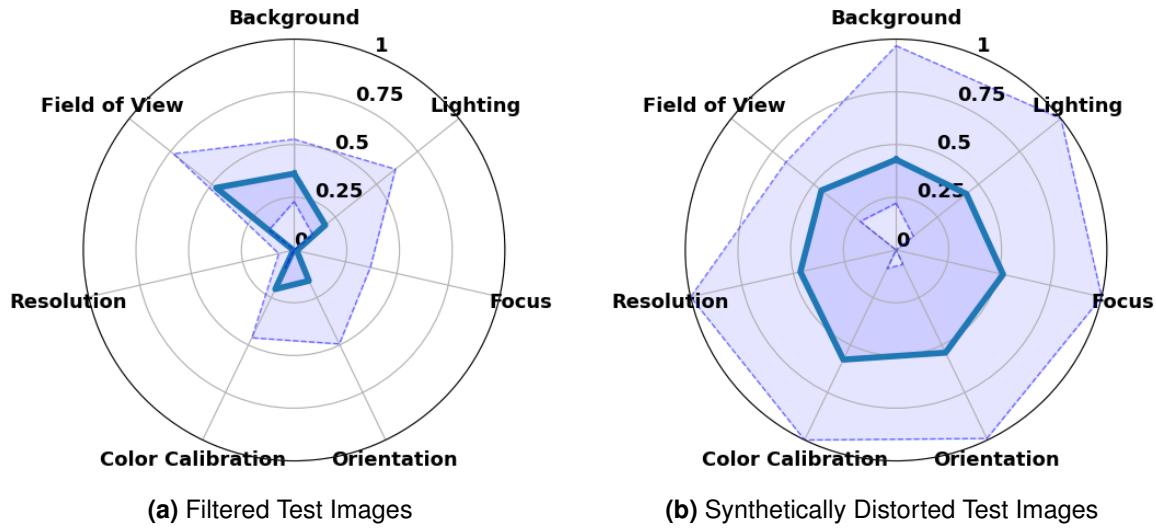


Figure 1.8: Synthetic test set analysis. (a) Filtered good quality test images (70 images, independent of training set). (b) Synthetically distorted test images.

Authentic Test Images

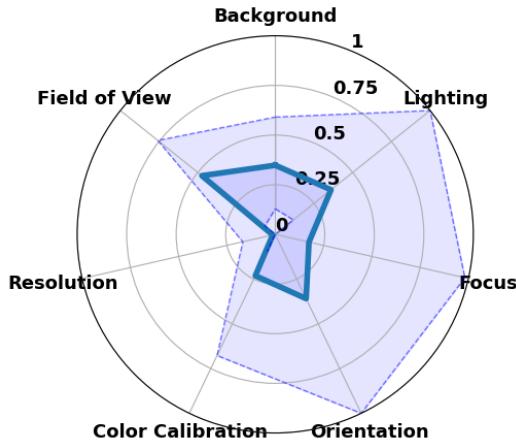


Figure 1.9: Authentic test set from the SCIN dataset, independent of the training images, showing real-world distortions.

Chapter 2

Discussion

2.1 Interpretation of Results

Examining the confusion plots shows that the bottom row (Figure 1.2d to 1.2g) shows good performance, where the diagonal indicate correct predictions with only minor fluctuations. In contrast, the top row (Figure 1.2a to 1.2c) has more noticeable issues. For instance, Figure 1.2a shows that there are rarely predictions on the higher severity for background distortion. This is because, in the distortion pipeline, if the background proportion is less than 10%, no color blocks are added, resulting in a 0 value for background distortion. This indicates many images were given a 0 for background distortion. Improving the training dataset to include more images with background could address this issue.

Orientation predictions, as shown in Figure 1.2c, is generally unsure and tend to cluster in the middle. This might be due to the various perspective changes (top, bottom, right, left) applied, making the model predict around the middle as it detects perspective distortions but not precisely which way and how strong.

Lighting predictions, as shown in Figure 1.2b, are reasonably accurate, but errors may occur because the criteria include two opposite types of distortion: brightening and darkening the image. This could lead to mispredictions as the inherent distortions look opposite but have the same values for the criteria.

These observations highlight the strengths of using a combined dataset, making the model more robust. Testing with datasets containing more background, such as the SCIN dataset, shows higher background scores, validating this hypothesis. Conversely, the Fitzpatrick dataset, with images taken in controlled settings or with dermatoscopes, shows better field of view predictions due to less background, supporting the combined dataset's robustness. This detailed analysis helps to understand where the model performs well and where improvements are needed.

Furthermore, the final model was tested on the original training images filtered for good quality. The radar charts in Figure 4.9 provide a visual representation of distortion levels across seven quality criteria. These charts confirm that the SCIN images exhibit more distortion compared to the Fitzpatrick17k images, which is expected due to the controlled environment in which the Fitzpatrick17k images were taken. The absence of distortion in resolution and focus across both datasets confirms the effectiveness of the filtering process.

Furthermore, the final model was tested on the original training images that were filtered as good quality images. This test was done to confirm that the images are indeed of good quality. Fig xx shows radar charts for the mean distortion levels and standard deviations across seven quality criteria for the 475 good quality SCIN, Fitzpatrick17k, and combined images. These charts provide a visual representation of the distortion levels across the seven criteria, with values ranging from 0 (center) to 1 (outer edge). The standard deviations indicate the variability in distortion levels for each criterion.

The radar charts reveal that the SCIN images have more distortion compared to the Fitzpatrick17k images, with the combined images falling in between. This suggests that the SCIN images have more distortions than the Fitzpatrick17k images, which is expected since the Fitzpatrick17k images were taken in a controlled environment. Additionally, both the SCIN and Fitzpatrick17k images show no distortion in resolution and focus, confirming that the filtering process was effective in selecting good quality images.

2.2 Answering the Research Questions

2.3 Comparison with Related Work

2.4 Reflection

2.5 AI Tools Used

In this work, several AI tools were used. ChatGPT was used to compress and summarize content. Additionally, it was used to optimize sentences and sections to make them more reader-friendly. Furthermore, GitHub Copilot was used in the development environment. It primarily helped in developing the Python scripts and models. These tools made the work more efficient and helped improve the overall quality of the thesis.

Chapter 3

Conclusion and Future Work

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>
- 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>
- 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>
- 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>
- 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>
- 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>
- 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>
- 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.](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

A.1 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.

include
risk man-
age-
ment and
project
planning
pdfs

search
database
and
maybe
add
KonIQ-
10k and
other from
ARNIQA!

- **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 .

A.2 Degradation Types

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

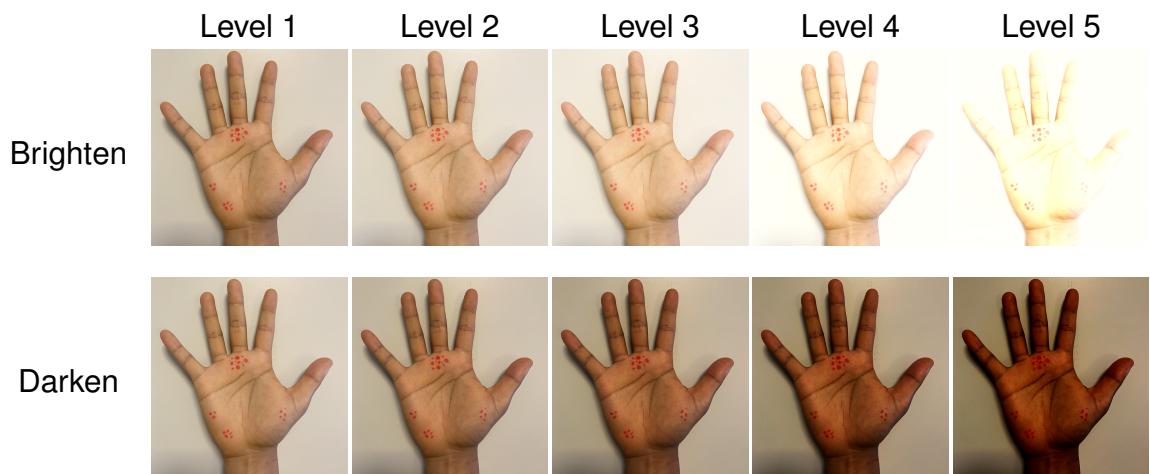


Figure A.1: Visualization of the degradation types belonging to the *Brightness change* group for increasing levels of intensity.

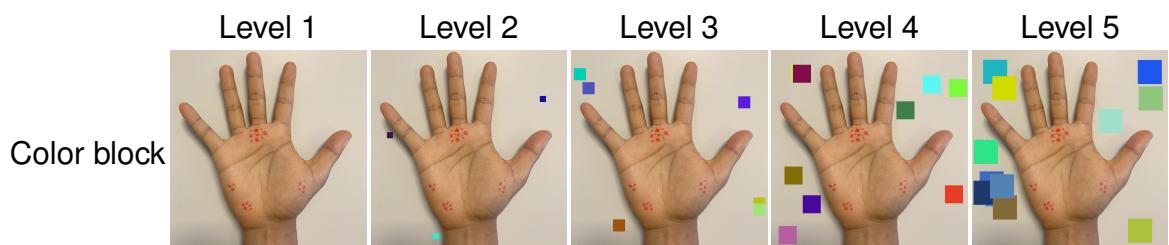


Figure A.2: Visualization of the degradation types belonging to the *Background color* group for increasing levels of intensity.

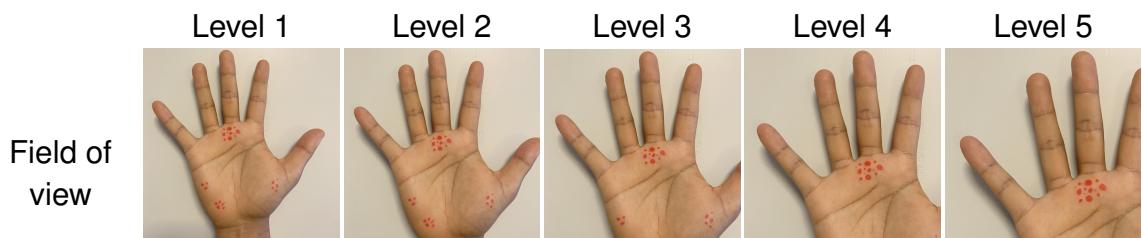


Figure A.3: Visualization of the degradation types belonging to the *Field of View* group for increasing levels of intensity.

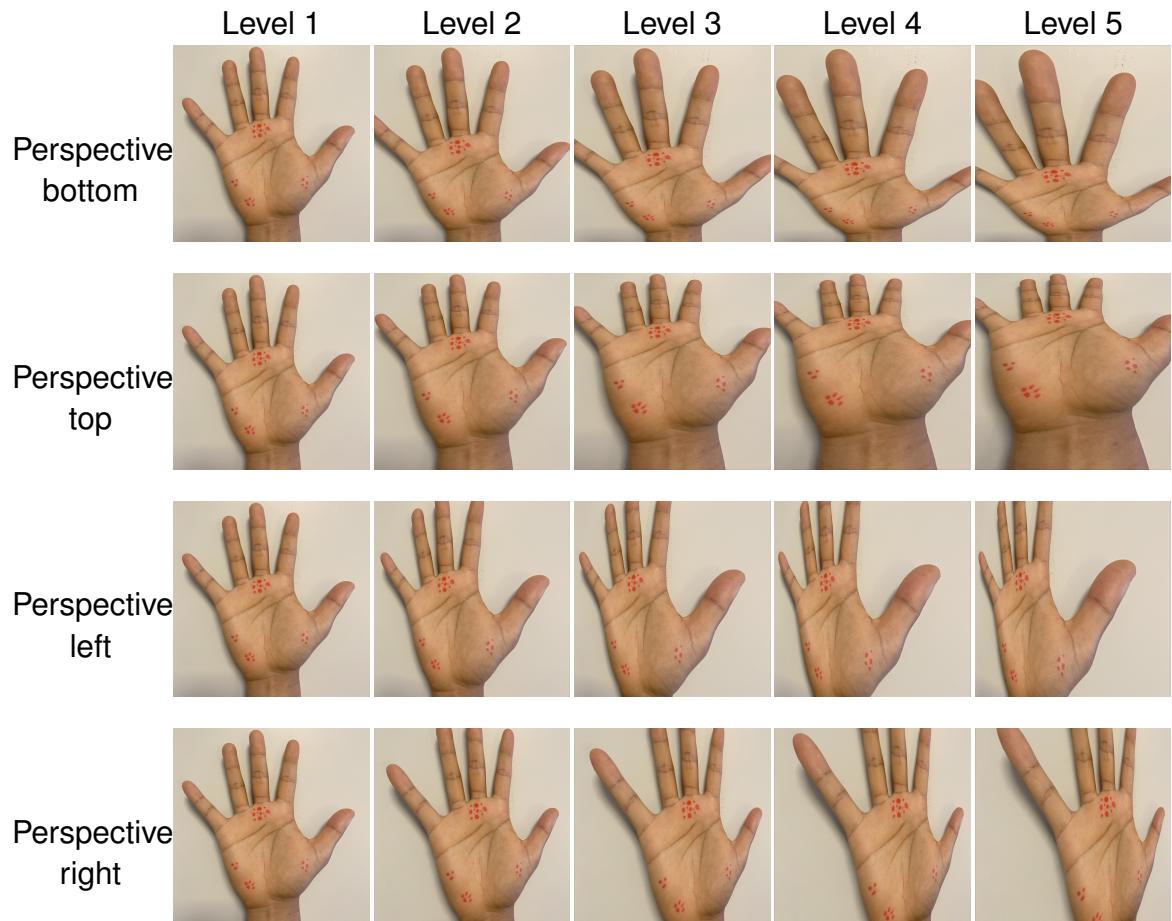


Figure A.4: Visualization of the degradation types belonging to the *Image orientation* group for increasing levels of intensity.

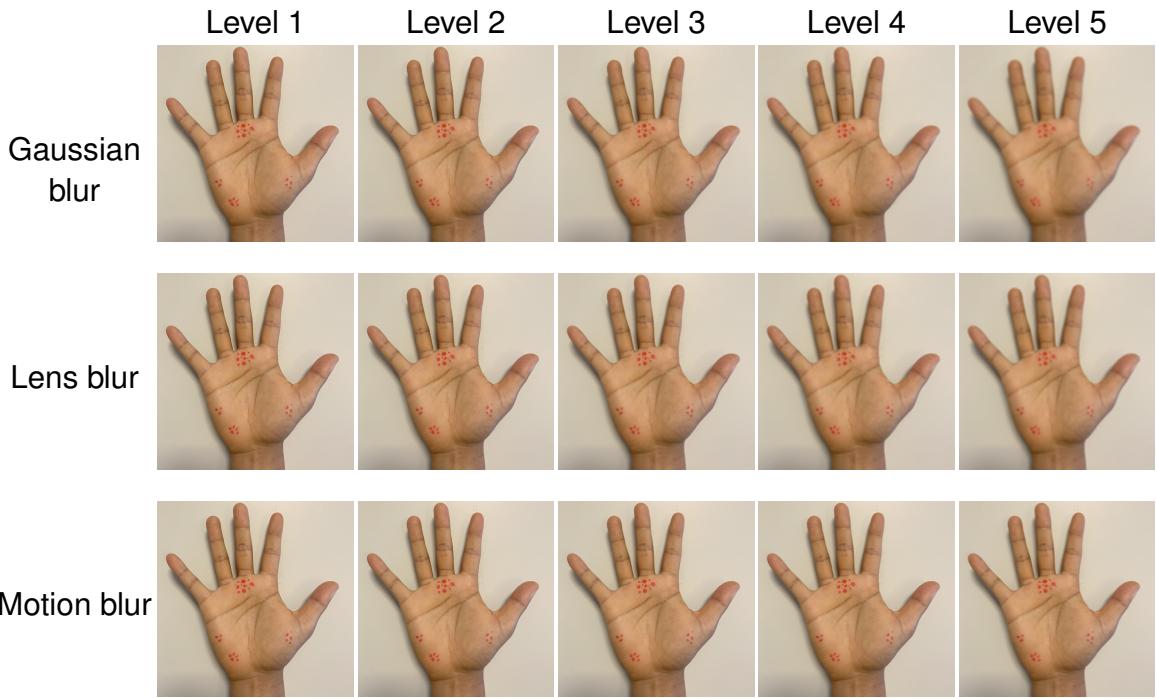


Figure A.5: Visualization of the degradation types belonging to the *Focus* group for increasing levels of intensity.

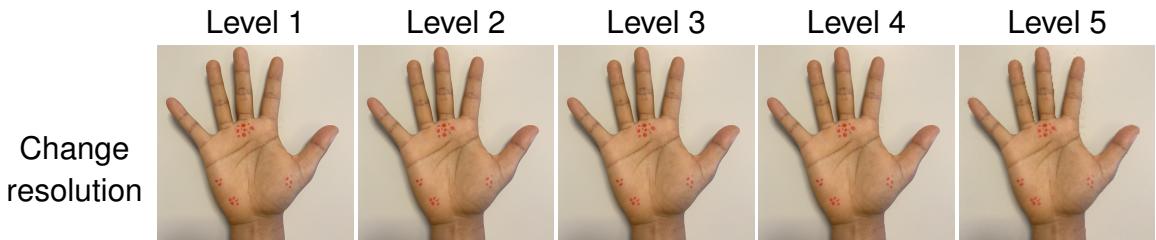


Figure A.6: Visualization of the degradation types belonging to the *Resolution* group for increasing levels of intensity.

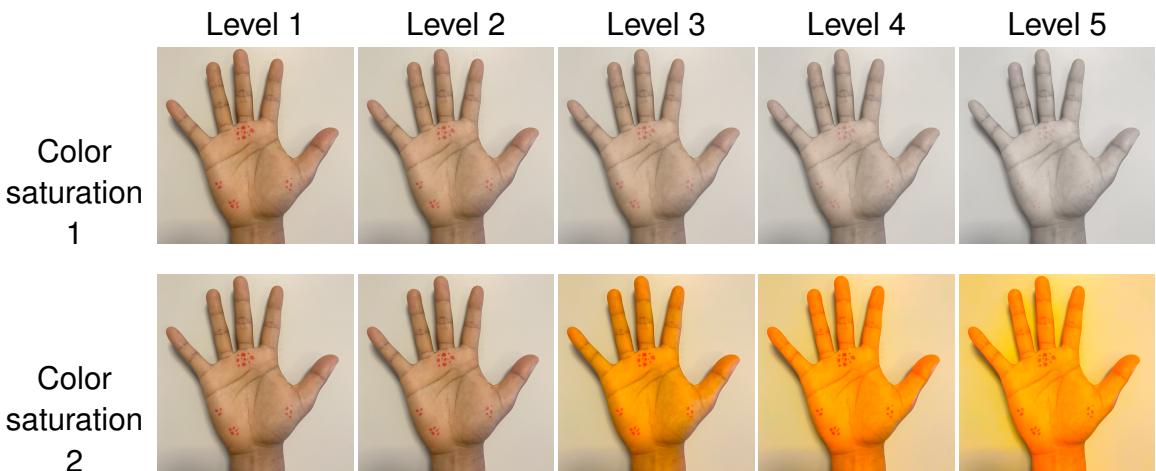


Figure A.7: Visualization of the degradation types belonging to the *Color calibration* group for increasing levels of intensity.