Improving Hand Landmark Detection in Infrared Images for Leprosy Diagnosis Using Colorization and Image Transformations

Zofia Rogacka-Trojak[1]

Supervisor(s): Jan van Gemert[1], Thomas Markhorst[1], Zhi-Yi Lin[1]

[1]EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
January 26, 2025

Name of the student: Zofia Rogacka-Trojak
Final project course: CSE3000 Research Project
Thesis committee: Jan van Gemert, Thomas Markhorst, Zhi-Yi Lin, Kaitai Liang

Abstract

Hand landmark detection in infrared (IR) images is essential for early leprosy diagnosis in developing countries like Nepal, helping to prevent serious complications and disability. However, current hand landmark detection models, such as Google's detection models comprised in the MediaPipe framework, often struggle with this task due to domain mismatch. While these models are trained on RGB images, the data for this research consists of greyscale IR images. This study addresses this challenge by exploring image transformation and colorization techniques to enhance MediaPipe's hand landmark recognition accuracy on IR images. Preprocessing was chosen over retraining the existing model due to limited computational resources and the lack of labeled target domain data, which makes the retraining infeasible.

Two preprocessing pipelines were developed to address different image characteristics: images with visible hand edges but varying colors of the hand, and images where hands blend in with the background, making the edges difficult to distinguish. The transformations include turning an image into its negative, colorization, contrast enhancement using Contrast Limited Adaptive Histogram Equalization (CLAHE), and masking to remove occlusion.

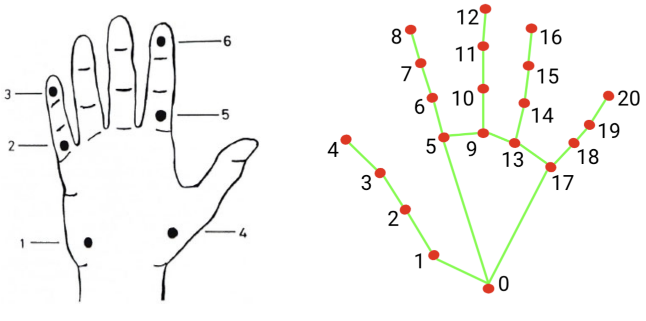
To evaluate the effectiveness of these techniques, accuracy has been calculated using Percentage of Correct Keypoints (PCK) metric and were compared against two baselines: a lower bound (MediaPipe performance on unchanged IR images) and an upper bound (MediaPipe performance on similar RGB images). Preliminary findings indicate that colorization significantly improves recognition for hands with sharp color transition, while contrast enhancement boosts edge definition for hands that blend into the background. By combining these approaches, the overall accuracy of hand landmark detection improved up to 25%, depending on the threshold value, particularly for the targeted open palm-up hand position.

These results demonstrate that preprocessing techniques can effectively reduce the input domain mismatch, enhancing automated leprosy diagnosis and supporting early detection efforts in low-resource settings.

1 Introduction

Despite being one of the oldest recorded diseases [1], leprosy, known as Hansen's disease, is still a modern problem, especially in underserved regions such as Nepal. While the treatment process is relatively straightforward, early diagnosis is crucial to prevent disabilities among those infected.

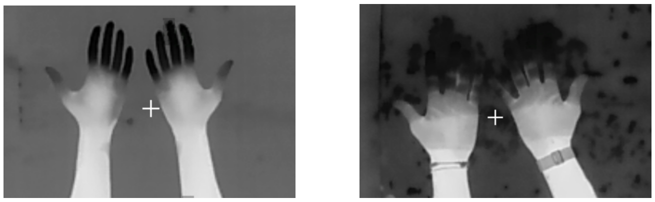
As proposed in [2], due to reduced blood flow caused by leprosy, a preliminary classification can be done by infrared imaging of patients' hands. By analyzing temperature differences between areas of interest, as shown in Figure 1a, and the rest of the hand, it is possible to identify whether a subject has leprosy. The current approach depends on manual observation of these differences, however, automating this process could significantly enhance the speed, accuracy, and overall effectiveness of diagnosis.



(a) (b)

Figure 1: Overlap between sensory testing areas (a) [3] and MediaPipe landmarks (b) [4] shows that the model's detected points can serve as reliable annotations, demonstrating its suitability for this research.

In a previous study [5], the MediaPipe model [6] was utilized to automate landmark detection, as the model's predicted landmarks, shown in Figure 1b, align with the areas of interest. However, challenges arose when analyzing greyscale infrared images, where the model struggled to detect hands, likely because it was trained exclusively on RGB images [7]. Moreover in infrared images, hands exhibit sharp color transitions between hot and cold areas, resulting in a patchy appearance, as shown in Figure 2a. This contrasts with the smooth and uniform look typically seen in RGB images. Additionally, some infrared images suffer from poor edge visibility around the fingers, as illustrated in Figure 2b, making it difficult to distinguish the fingertips and finger boundaries clearly. This visual disparity between RGB and infrared images represents the domain gap that limits the model's ability to recognize hands accurately.



(a) (b)

Figure 2: Infrared greyscale images provide two challenges for hand landmark detection models: (a) hands with sharp color transition and (b) edge-obscured fingers, decreasing their performance.

This research investigates colorization techniques and image transformations to bridge the domain gap between infrared and RGB images for leprosy diagnosis. The ultimate objective is to evaluate how effectively these techniques can enhance hand landmark recognition on infrared images, thereby improving the potential for automated leprosy diagnosis in resource-limited settings.

Preprocessing techniques were chosen over retraining the existing model due to two critical constraints: limited computational resources and the absence of labeled target domain data. Retraining was deemed infeasible because the original model was trained on 116,000 images [7], and achieving similar high performance would likely require a dataset of comparable size, which is unattainable in this context.

The rest of the paper is structured in the following way. Section 2 introduces a short literature survey of the related work. Section 3 describes the methodology of the research with the main hypothesis, dataset collection and proposed method. Section 4 elaborates on the evaluation metric, baselines and the experiments. Section 5 discusses responsible research practices and their relevance to the project. Section 6 engages in a broader discussion of the findings, while Section 7 concludes the work and outlines directions for future research.

2 Related Work

This section reviews relevant literature on hand landmark detection, image colorization, and domain mismatch, providing context for the current study.

2.1 Hand Landmark Detection

Hand landmark detection has seen significant advancements in recent years, especially with models like MediaPipe's Gesture Recognizer [8]. However, this model was trained exclusively on RGB images [7], which likely explains its suboptimal performance when applied to infrared greyscale images. Infrared images differ significantly from RGB images in terms of color representation, which creates a domain gap that affects model accuracy.

The MediaPipe Gesture Recognizer operates in two steps: first, it detects the palm using the BlazePalm Detector [7], and then it applies a method based on [9] to detect finger keypoints. This two-step process might cause misclassification, as it is possible for only part of the hand to be correctly detected—such as the palm, but with misplaced finger keypoints. This issue highlights the challenges of using the model for infrared hand landmark detection as infrared images can show hands gradually blending in with the background.

2.2 Image Colorization

Image colorization is a well-established area of research [10], [11], [12]. However, most existing techniques focus on the colorization of near-infrared (NIR) images, which are not directly applicable to our work with thermal infrared images. For instance, the method proposed by Limmer and Lensch [13] colorizes near-infrared images using deep convolutional networks, but this approach does not address the unique challenges posed by thermal patterns on hands in infrared.

Furthermore, many existing colorization models are trained on images that were artificially turned into greyscale [14], which may limit their ability to perform well on infrared images. The domain gap between artificial greyscale images and actual infrared images presents additional challenges for colorization techniques, as infrared images contain distinct patterns that do not align well with RGB data.

In contrast, technique proposed by Zhang et al. called ECCV16 [12] uses class rebalancing to introduce greater color diversity into colorization tasks. This might be helpful in our case, as it could help the model handle the sharp color transitions and varied appearance of infrared images. In their study [12], the model was able to produce images that fooled 30% of participants into thinking the images were naturally colored. Additionally, Zhang et al.'s newer model called SIGGRAPH17 [15] outperformed earlier approaches in classification tasks and may offer insights for improving colorization accuracy in the context of infrared images.

2.3 Domain Gap and Its Impact

The domain gap between infrared and RGB images is a key issue for this research. Models trained on RGB images, such as MediaPipe's Gesture Recognizer, struggle to adapt to infrared images due to differences in visual features, such as the lack of color and unusual patterns on the hands. Addressing this domain mismatch is crucial for improving the accuracy of hand landmark detection.

3 Methodology

3.1 Hypothesis and Problem Statement

The primary hypothesis of this research is that the disproportion in MediaPipe's performance on RGB hand images compared to greyscale infrared images arises from a mismatch between the training data used to develop the MediaPipe model and the input images provided during detection. This performance gap can be reduced by transforming infrared images through a series of processing steps to resemble RGB images as closely as possible in appearance, including color inversion and colorization.

3.2 Data Collection

To evaluate the accuracy and performance of hand landmark detection on transformed images, a dataset of annotated images was created. The images were designed to replicate the diagnostic setup used in Nepal, where patients presented their hands in an open palm-up gesture. The dataset includes images of both male and female hands to introduce variability in hand size and shape.

To simulate the temperature differences observed in patients with leprosy, where the palms are typically warmer (whiter tone on the image) than the fingers (darker tone), participants dipped their fingers in water maintained at approximately 11 °C before imaging. This setup mimics the characteristic thermal patterns of the diagnostic process.

The images were captured using an infrared camera UTi721M [16] in the grayscale mode ”White Hot”, replicating the technology currently employed in Nepal. The corresponding RGB images were also captured using a smartphone camera for comparison purposes.

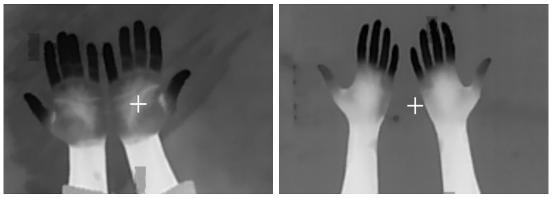
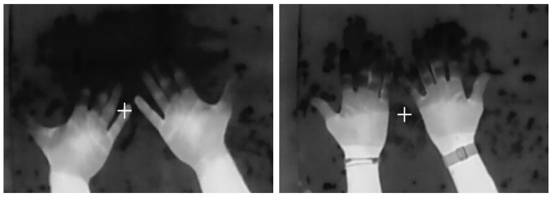
Challenge	Proposed Solution	Example Images
Hands with Sharp Color Transition	• Remove occlusion. • Transform into negative. • Use the SIGGRAPH17 model for colorization.	
Edge-Obscured Hands	• Remove occlusion. • Use CLAHE to enhance edge definition.	

Table 1: The dataset presents two main challenges: hands with sharp color transition and edge-obscured hands which should be treated separately.

After data acquisition, each image (both RGB and infrared) was annotated manually. A total of 21 hand landmarks were annotated for each hand, based on the landmark definitions provided by the MediaPipe framework, as illustrated in Figure 1b.

Dataset Characteristics and Challenges

Analysis of the dataset revealed distinct subsets of images, each presenting unique challenges. The summary of these challenges, along with proposed solutions and example images, can be seen in Table 1.

Subset 1: Hands with Sharp Color Transition

For most participants in the data collection process, the hands exhibit two extreme color tones: the palm typically appears significantly brighter, sometimes even white, while the fingers are much darker. This contrast is characterized by a sharp color transition rather than a gradiental transition, with distinctly defined edges. For this subset of images, the hypothesis is that inverting the colors, followed by an application of a colorization model, could enhance MediaPipe's performance. Examples of this subset can be found in the upper row of Table 1.

Subset 2: Edge-Obscured Hands

In this subset, fingers subjected to cold temperatures appear significantly darker, with the background also being dark. This results in confluent colors and poorly defined edges, making it difficult even for human observers to detect the edges. The primary objective for this subset is to enhance edge definition to improve both human and algorithmic recognition. Examples of this subset can be found in the lower row of Table 1.

3.3 Proposed Method

Pipelines to Address Dataset Challenges

The final setup consists of two transformation pipelines, followed by a comparison and assessment process to determine which landmarks and coordinates should be kept.

First Pipeline: Optimized for Hands with Sharp Color Transition

The first pipeline is designed for images where the hands are distinctly visible with a brighter tone against a darker background and a sharp color transition.

- **Object Removal:** Unnecessary objects, such as temperature indicators (displaying the highest and lowest recorded temperatures), are removed from the image. These objects often overlap with the palm or finger regions, obstructing visibility. Object removal is achieved by filtering out red and green colors and replacing the affected pixel values with those of surrounding pixels.
- **Smoothing:** The image is slightly smoothed using Gaussian smoothing to eliminate residual artifacts introduced during object removal.
- **Color Inversion:** The processed image is converted to its negative, aligning its appearance more closely with the training data used for the colorization model.
- **Colorization:** The transformed image is colorized using the SIGGRAPH17 colorization model proposed by Zhang et al. [15].

An example of transformation process for this pipeline can be seen in Figure 3.

Second Pipeline: Optimized for Edge-Obscured Hands

The second pipeline was specifically designed to address challenges in images where the fingers blend into the surrounding background, making them difficult to distinguish.

- **Object Removal:** Similar to the first pipeline, irrelevant objects are removed to ensure unobstructed hand visibility.
- **Contrast Enhancement:** The Contrast Limited Adaptive Histogram Equalization (CLAHE) method is applied to enhance image contrast, effectively highlighting finger edges. It already has been proven to improve

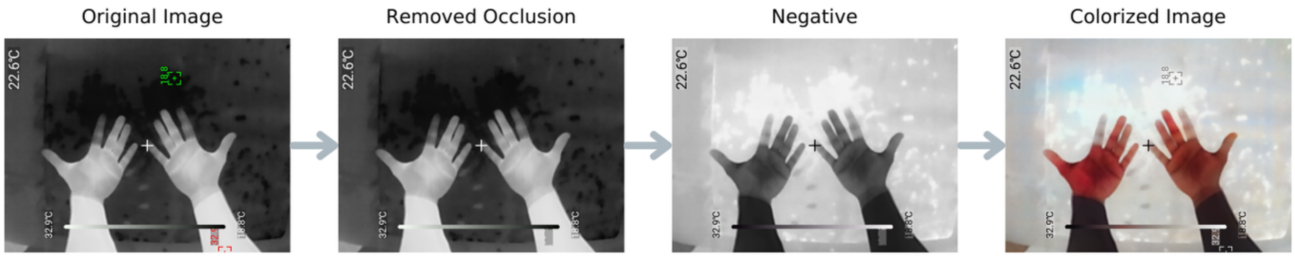


Figure 3: The first pipeline enhances infrared images by removing occlusion, turning the image into a negative, and applying colorization to improve landmark detection accuracy.

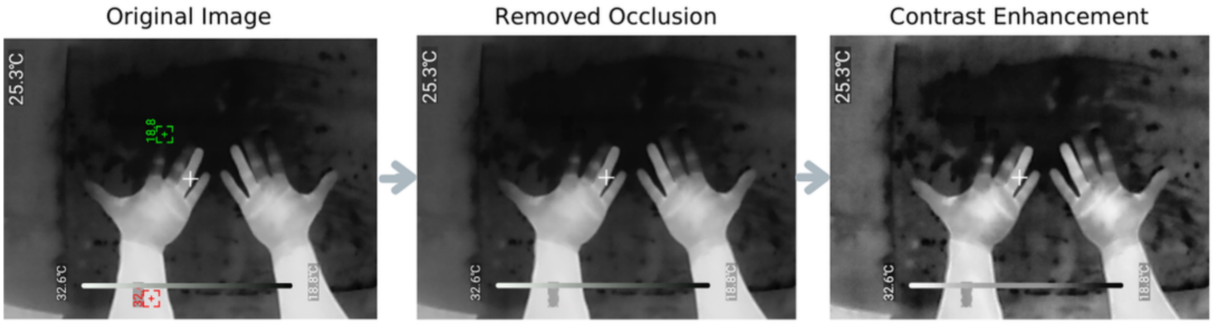


Figure 4: The second pipeline enhances images with blended fingers by removing occlusion and using CLAHE to improve contrast, making finger edges more distinguishable for accurate landmark detection.

the detectability of face features on black-and-white images [17] and has been proposed as a way to lighten the dark areas of images.

An example of transformation process for this pipeline can be seen in Figure 4.

Landmark Selection

Assessing the correctness of results is a critical aspect of this research. Correctness is defined as the accurate detection of all 21 hand landmarks for both hands in the image, assuming a standard palm-up position. A common challenge is that certain transformations may enable the accurate recognition of landmarks for one hand while failing on the other. This inconsistency requires a methodology to combine results from different processing pipelines.

Combining Results from Multiple Pipelines

The Gesture Recognizer [8] operates in two steps [7], which could lead to only partially correct hand recognition, e.g., with the palm accurately detected but with misplaced fingers. Due to this, the results can not be combined based on whether the model detected the hand. Instead, another feature of Gesture Recognizer is utilized. Namely, the detected gesture based on the location of landmarks. During the diagnostic process, patients are typically asked to present their hands in an open palm-up position. This characteristic is used to evaluate the preliminary accuracy of the recognized landmarks. If the detected landmarks correspond to an ”Open Palm” gesture as identified by the Gesture Recognizer, it indicates a higher likelihood that the results are accurate. By focusing on results where the ”Open Palm” gesture is recognized, it becomes possible to filter out incorrect detections and merge results from multiple pipelines more effectively.

Currently, both pipelines are executed, and their outputs are fed into the Gesture Recognizer. The results are then combined based on the following rules:

- **Both Pipelines Detect Both Hands, Labeled as ”Open Palm”:** In this scenario, the landmarks are combined based on the following criteria. For the index, middle, ring, and little fingers, the landmarks corresponding to the longest detected fingers are selected. For the thumb, the landmark farthest to the right is chosen for the right hand, while the farthest left is used for the left hand.
- **Both Pipelines Detect Both Hands, Some Labeled as ”None”:** If only one left or right hand is labeled as ”Open Palm,” the landmarks from that result are chosen, and the corresponding ”None” labels are disregarded.

  If both (e.g., left) hands are labeled as ”None,” the landmarks with the lower confidence score for ”None” are selected, assuming that lower confidence in the ”None” label indicates that the landmarks may more closely resemble an ”Open Palm” gesture.
- **Not All Hands Are Detected:** If the pipelines detect opposite hands (e.g., one detects the right hand and the

other detects the left), the results are combined to include both hands.

If both pipelines detect the same hand, the landmarks are selected based on the criteria outlined above.

If one pipeline detects both hands and the other only detects one, the results are compared as described above to determine if a substitution is needed for one hand.

- **No Hands Are Detected:** In this case, an empty list is returned.

This approach ensures a systematic and confidence-based combination of results, optimizing the accuracy and reliability of landmark detection.

4 Experiments

4.1 Software Details

A Python 3.10 environment was developed to facilitate the implementation of the transformation pipelines, landmark detection, accuracy measurement, and result evaluation. The transformation pipelines incorporate methods and functions from the OpenCV library [18], as well as two colorization models proposed by Zhang et al. [12], [15], sourced from the corresponding GitHub repository [19]. For landmark detection and gesture analysis, the MediaPipe model known as Gesture Recognizer [8] was utilized. Accuracy calculations and result visualizations were performed using standard Python libraries, including Matplotlib for generating charts and graphs.

4.2 Evaluation Metric

The evaluation of landmark detection accuracy in this research is carried out using the Percentage of Correct Keypoints (PCK) metric. PCK is commonly used in human pose estimation tasks [20] to assess the accuracy of predicted keypoints in relation to ground truth keypoints. The standard approach in PCK evaluation involves comparing the Euclidean distance between predicted and ground truth landmarks.

In this study, an adaptive version of PCK is employed, where the acceptable distance is scaled relative to the size of the hand. Specifically, the hand size is determined by measuring the distance from the wrist (point 0 in Figure 1b) to the tip of the middle finger (point 12 in Figure 1b), denoted as L_{hand}, ensuring that the evaluation treats every hand equally regardless of its size. The distance threshold for each landmark is calculated as the length of the hand multiplied by a specified threshold δ, such that:

$$\tau = L_{\text{hand}} \times \delta \tag{1}$$

where δ is a predefined threshold value, and τ represents the acceptable distance for the landmark.

The landmarks are normalized to a 0-1 range based on the image dimensions, with coordinates scaled relative to the image width and height. The Euclidean distance between predicted and ground truth landmarks is then computed in this normalized space. The normalized Euclidean distance for each landmark is given by:

$$d_{\text{norm},i} = \sqrt{(x_{\text{pred},i} - x_{\text{gt},i})^2 + (y_{\text{pred},i} - y_{\text{gt},i})^2} \tag{2}$$

where $(x_{\text{pred},i}, y_{\text{pred},i})$ are the predicted coordinates of the i-th landmark, and $(x_{\text{gt},i}, y_{\text{gt},i})$ are the ground truth coordinates of the i-th landmark, both normalized with respect to the image dimensions.

To determine whether a landmark is considered correct, the distance is compared to the threshold τ, also computed with the use of normalized ground truth landmarks. If the distance $d_{\text{norm},i}$ is smaller than or equal to the threshold τ, the prediction is considered correct. The PCK at a given threshold is defined as:

$$\text{PCK}(\delta) = \frac{1}{N} \sum_{i=0}^{N} \mathbf{1}(d_{\text{norm},i} \leq \tau) \tag{3}$$

where N is the total number of landmarks, $\mathbf{1}(\cdot)$ is the indicator function, which is 1 if the condition holds and 0 otherwise, and $d_{\text{norm},i}$ is the normalized Euclidean distance for the i-th landmark and τ corresponds to acceptable distance computed from Equation 1.

This approach ensures that the PCK metric is invariant to variations in hand size, allowing for fair comparisons across different hand sizes and image conditions.

4.3 Baselines

To assess the effectiveness of the transformation pipelines, two baseline performance measures were established:

Lower Baseline: This represents the accuracy of the Gesture Recognizer on the unaltered greyscale infrared images.

Upper Baseline: This was derived from RGB images of hands captured in positions resembling those in the infrared dataset as closely as possible. While these RGB images could not be captured under identical conditions due to technical limitations of the infrared camera, they closely resemble them. The RGB images were acquired using a different device and lacked the temperature scale and temperature indicators present in the infrared images. Despite these differences, the RGB images provide an upper bound for assessing the potential performance of transformed infrared images.

By comparing the accuracy of transformed images against these baselines, it becomes possible to evaluate whether the transformations significantly improve landmark detection and how closely they approach the performance observed for RGB images.

4.4 Results and Analysis

This subsection presents the intermediate results of each transformation, along with their analysis and proposed improvements, which collectively led to the final pipelines.

First Pipeline Steps and Results

Colorization

The first part of the experiment involved applying a colorization process to the images. Two models, ECCV16 [12] and SIGGRAPH17 [15] were chosen for this task. These models are Convolutional Neural Networks (CNN) designed with a feed-forward pass. While ECCV16 is a fully automated approach, SIGGRAPH17 offers a guided option. For this experiment, both models were used in their automated setup.

At each threshold value that influenced the acceptable distance used in PCK, the score for ECCV16 and SIGGRAPH17 was consistently lower than the lower bound, as can be seen on Figure 6. Upon closer inspection of the images, it was concluded that the models incorrectly colorized the images. Both models were trained on RGB images artificially converted to greyscale [12], which differed significantly from the infrared images in the evaluation dataset. In artificial greyscale images of hands, the hand typically appears in a uniform dark grey tone, with certain areas exhibiting darker details. These darker regions correspond to the natural folds and creases in the skin. However, in the evaluation dataset, the images often showed the opposite pattern, with hands appearing brighter overall and lacking these subtle contrasts. This discrepancy caused the models to either colorize the hands in unnaturally bright tones or fail to apply meaningful colorization, leaving the hands nearly untouched. Therefore, the next step was to transform the images into their negatives.

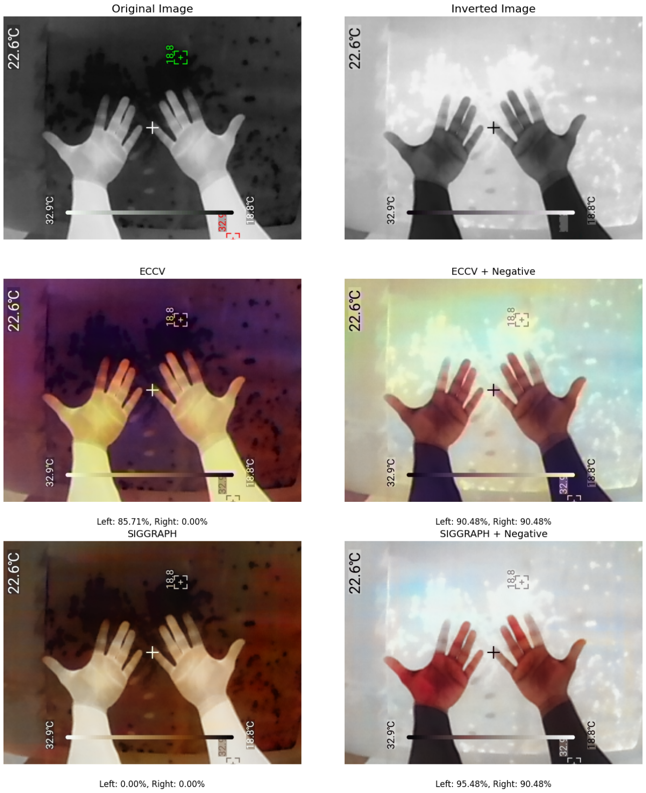

Original Image
Inverted Image
ECCV
ECCV + Negative
Left: 85.71%, Right: 0.00%
Left: 90.48%, Right: 90.48%
SIGGRAPH
SIGGRAPH + Negative
Left: 0.00%, Right: 0.00%
Left: 95.48%, Right: 90.48%
22.6℃
32.9℃
18.8℃

Figure 5: Inverted images improved alignment with the models' training data, enhancing colorization accuracy and hand feature visibility, as seen in accuracy score with threshold value 0.05.

This adjustment improved the results. ECCV16 achieved score around the lower bound for each threshold value and SIGGRAPH17 improved the PCK by 3-12%, depending on the threshold, in comparison with the lower bound. Notably, SIGGRAPH17 performed better than ECCV16, even though the images produced by ECCV16 appeared more colorful and initially seemed better colored. Those differences in colorization can be seen on Figure 5.

The difference in results can be explained by the findings in [15]. ECCV16 uses class rebalancing to introduce diversity in color, often resulting in overly aggressive colorization. This causes it to perform poorer on the classification task as it produces oversaturated colours, whereas SIGGRAPH17 colourization pallet more closely resembles realistic colours [15]. Consequently, the subsequent steps were performed using only the SIGGRAPH17 model.

Removing Occlusions

The next step was to eliminate objects occluding the view, as shown in Figure 3. As mentioned in [21], inter-class partial occlusion reduces the robustness of classifiers compared to humans and degrades the performance of detectors. Therefore, temperature indicators (red and green squares) were masked, and the affected pixels were replaced with neighboring pixel values. This step resulted in significant raise in the accuracy, with up to 20% final increase compared to lower baseline, as seen on Figure 6.

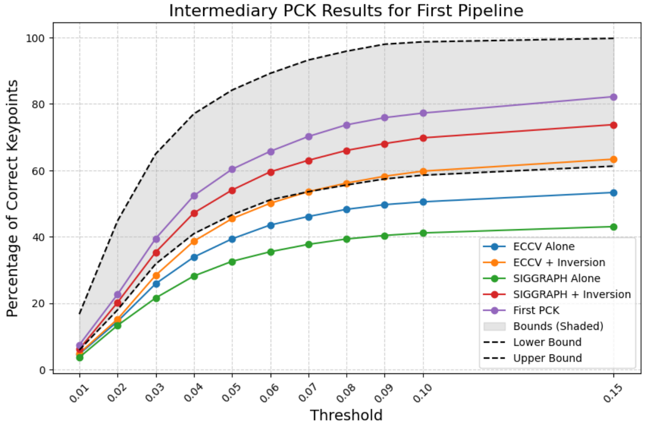


Figure 6: Intermediary PCK results for Pipeline 1 illustrates how image inversion and colorization improved accuracy by aligning the dataset with the models' training conditions.

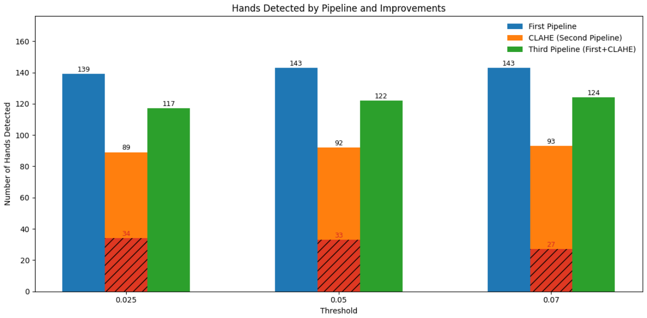
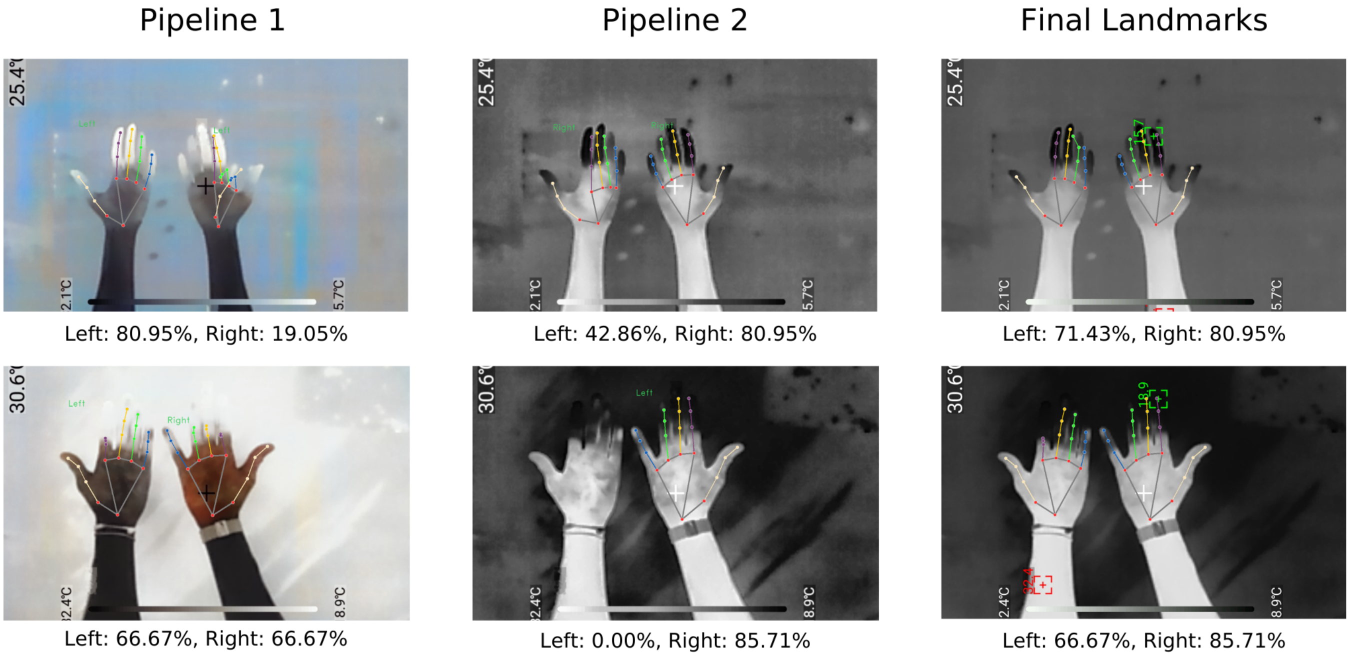


Figure 7: Comparison of hands detected by each pipeline. CLAHE improved accuracy for 17.5% of the dataset but reduced overall detection rates when combined with the first pipeline.



Left: 80.95%, Right: 19.05%

Left: 42.86%, Right: 80.95%

Left: 71.43%, Right: 80.95%

Left: 66.67%, Right: 66.67%

Left: 0.00%, Right: 85.71%

Left: 66.67%, Right: 85.71%

Figure 8: Examples of combining landmarks from both pipelines. The annotated images illustrate how the integration of results improves the detectability and accuracy of hand landmarks.

Overall Results for Pipeline 1

Pipeline 1 significantly improved the overall result. Approximately 55% of hands were detected with perfect accuracy, that is the model detected correctly both the palm and all the finger landmarks. Moreover, for another 25% it managed to correctly classify the palm landmarks. However, upon closer inspection, it was determined that the pipeline failed on images where the hand appeared in a uniform color, with fingers so dark that they blended into the background.

Second Pipeline Steps and Results

To address the problem of poor edge definition, the hypothesis was to improve the visibility of finger boundaries. Contrast Limited Adaptive Histogram Equalization (CLAHE) was applied, as it has been successfully used in enhancing face and pose detection [22], [23] also with MediaPipe [24].

Upon closer inspection, as shown in Figure 7, the CLAHE method identified fewer hands than the first pipeline. However, it achieved a higher accuracy score for 27-34 hands in each attempt, which represents about 17.5% of the dataset. For these hands, CLAHE improved the accuracy by at least 10% per hand. Combining CLAHE with the first pipeline, however, resulted in a lower number of detected hands. While the first pipeline identified 139-143 hands, depending on the threshold, the combination with CLAHE detected only approximately 117-124 hands.

Closer inspection revealed that the addition of CLAHE to the first pipeline caused the model to stop detecting hands or to detect them with worse results, especially in cases where there was a sharp colour transition between the white palm and black fingers, instead of a gradual gradient. This outcome aligns with the expectation that CLAHE enhances edge definition. As mentioned in [25], color plays a crucial role in object detection, and it was decided that hands with sharp color transitions and those with nearly invisible edges should be treated separately. Consequently, CLAHE became the foundation for the second pipeline.

The next step involved removing occlusions from the images, following the same procedure as in the first pipeline. This step slightly improved the result, with the final score for the second pipeline reaching the lower bound for each attempt, as can be seen on Figure 9c

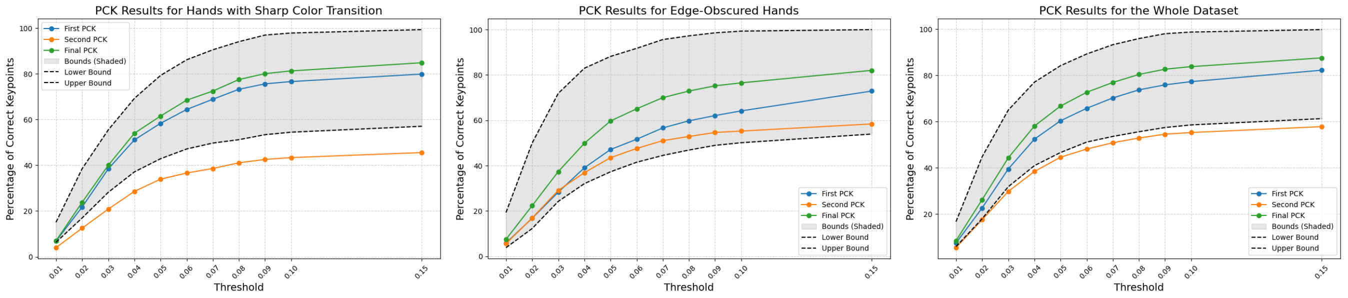
Although this result did not exceed the lower bound, it is significant for diagnostic purposes. Some regions of interest for diagnosis are located on the fingers, and their detectability was improved by this pipeline.

Combining the Results From the Pipelines

After developing both pipelines and identifying their respective strengths, the results were combined based on the recognized gestures. This approach leveraged the advantages of each pipeline to address the specific challenges in the dataset. More visual examples of the solutions are shown in Figure 8, which illustrates the significant improvements achieved through this combination.

The final setup was tested three times: once on the entire dataset, once on the subset of images presenting the first challenge (sharp color transitions), and once on the subset with the second challenge (edge-obscured views).

For the subset with sharp color transitions, as seen on Figure 9a, the first pipeline proved the most impactful. The second pipeline consistently performed much lower than the lower bound, while the first pipeline achieved results much closer to the final combined score. Nonetheless, the second pipeline contributed slightly to improving the overall score.



(a) (b) (c)

Figure 9: Results for different subsets of images show that for sharp color transition (a) first pipeline is the most impactful. While for obscured-edges (b) the combination of both pipelines is the most promising. For the whole dataset (c) mainly first pipeline contributes to the result, but combination slightly increases the accuracy.

For the edge-obscured subset, as seen on Figure 9b, the second pipeline was essential. It outperformed the lower bound and achieved results close to the first pipeline. However, only the combination of both pipelines produced significantly better results, showing the largest improvement between the final score and the highest score from any single pipeline.

For the whole dataset, Figure 9c, the performance graph resembled that of the sharp color transition subset, indicating that most images in the dataset faced challenges related to sharp color transitions. Overall, the combined approach improved the score by up to 25%.

These results demonstrate that the model effectively increases the detectability of hand landmarks. However, it is evident that accuracy remains a challenge: while the combination improves detection rates, the precise localization of keypoints is still suboptimal.

5 Responsible Research

This section outlines the ethical considerations, as well as the measures taken to ensure the reproducibility and transparency of the research.

5.1 Data Collection

Informed consent was obtained from all participants before image collection, allowing the use and sharing of the data for research purposes. To introduce variability and reduce bias, images were collected from both male and female participants. All images were captured under the same conditions, which are detailed in the document provided with dataset. Each participant followed a predefined set of instructions in the same order, ensuring the collection of multiple images for identical setups—for example, one finger of each hand submerged in cold water, with a rubber mat serving as the background. Additionally, all collected data were anonymized to ensure participant privacy.

5.2 Transparency and Reproducibility

To ensure transparency and reproducibility, the evaluation dataset and research code repository have been made publicly available. Moreover, the Methodology and Experiments sections provide detailed descriptions of the tools utilized, the intermediate steps, and the rationale behind the development of the pipelines. These details ensure that the experimental setup can be accurately replicated.

5.3 Additional Considerations

All collected data, except for the train set used to select transformations, was included in the final evaluation of the pipeline's performance. Importantly, all test set data were used to present the results, with no exclusion of images that yielded poor outcomes. This ensures that the reported results accurately reflect the method's performance across the entire test set, providing a comprehensive and unbiased assessment.

Given the limited size of the dataset, the pipelines were specifically designed to address the two largest subsets of images, which represent the most general image categories. This approach was taken to prevent overfitting and avoid tailoring the solution too closely to the specific dataset, ensuring the method's applicability to broader contexts.

6 Discussion

This research demonstrates the feasibility of reducing the performance gap in hand landmark detection between RGB and greyscale IR images using colorization and image transformation techniques. By integrating two pipelines—contrast enhancement via CLAHE and colorization model—the method improved detection accuracy by up to 25% across thresholds. These transformations enhanced hand visibility, making features more distinguishable for both the model and human observers.

6.1 Strengths

The primary strength of this approach lies in its ability to detect hands in challenging conditions, such as when edges are poorly defined or fingers blend with the background. Moreover, the dual-pipeline strategy—contrast enhancement via CLAHE and colorization—addresses diverse scenarios within the same image, enabling more accurate landmark detection. This flexibility ensures the method adapts well to different settings, improving overall performance.

6.2 Limitations

The method remains reliant on input image quality, which is the primary cause of errors. In 15 images, darker fingers led to inaccurate fingertip landmark predictions, while in 19 images, poor image quality resulted in missed or severely misaligned landmarks. Despite the overall improvements, further refinement is needed to enhance accuracy in challenging scenarios.

Additionally, the logic for combining pipeline results is simplistic because it was not the primary focus of this research. Future work could explore more sophisticated methods for integrating results to further enhance detection accuracy and reliability. Expanding the dataset to include more diverse conditions is also crucial for validating the method's generalizability and improving accuracy in varied IR imaging setups.

7 Conclusions and Future Work

This study investigated the effectiveness of colorization and image transformations in improving hand landmark detection in greyscale infrared images. The results demonstrate that these transformations enhance detection accuracy. The contrast enhancement proved to be the most useful for the hands with obscured edges, while colorization increased the performance on the hands with sharp color transition. However, due to the diverse nature of the images, only a combination of distinct transformation pipelines proved to yield the highest performance, with up to 25% increase above the lower bound.

The primary application of this research lies in its potential for early leprosy diagnosis. The transformations developed in this study could be integrated into a system for automatic hand landmark retrieval. Such a system would capture infrared images, apply the appropriate transformations, and leverage existing hand landmark detection models without requiring extensive retraining.

7.1 Future Research Directions

Future work should address the limitations identified in this study, particularly those related to image quality. Investigating methods to mitigate the impact of low-quality inputs could further improve detection accuracy. Additionally, developing more sophisticated logic for combining pipeline outputs could enhance performance.

Another critical issue is testing the generalizability of this approach on datasets collected in varied conditions. Expanding the dataset to include diverse scenarios and settings will be essential to assess the adaptability of the proposed method in real-world applications.

References

[1] W. M. Media, "Leprosy: An old disease with modern challenges," *Walsh Medical Journal*, 2024, accessed: December 15, 2024. [Online]. Available: https://www.walshmedicalmedia.com

[2] A. L. Cavalheiro, D. T. Costa, A. L. Menezes, J. M. Pereira, and E. M. Carvalho, "Thermographic analysis and autonomic response in the hands of patients with leprosy," *Anais brasileiros de dermatologia*, vol. 91, no. 3, pp. 274–283, 2016.

[3] D. M. McAuley, P. A. Ewing, and J. K. Devasundaram, "Effect of hand soaking on sensory testing," *International Journal of Leprosy and Other Mycobacterial Diseases*, vol. 61, no. 1, pp. 16–19, 1993.

[4] ResearchGate, "An evaluation of hand-based algorithms for sign language recognition - scientific figure on researchgate," https://www.researchgate.net/figure/MediaPipe-Hands-21-landmarks-13_fig1_362871842, 2024, accessed: 15 Dec 2024.

[5] I. Schemkes, "Semi-automatic temperature analysis based on real-time hand landmark tracking in infrared videos: A model to support research into the potential of infrared thermography for leprosy diagnosis," Master's Thesis, Delft University of Technology, Delft, Netherlands, July 2024. [Online]. Available: http://repository.tudelft.nl/

[6] Google AI, "Mediapipe hand landmarker," https://ai.google.dev/edge/mediapipe/solutions/vision/hand_landmarker, accessed: 15 Dec 2024.

[7] F. Zhang, V. Bazarevsky, A. Vakunov, A. Tkachenka, G. Sung, C.-L. Chang, and M. Grundmann, "Mediapipe hands: On-device real-time hand tracking," *Google Research*, 2020, https://google.github.io/mediapipe/solutions/hands.

[8] Google AI. (2025) Mediapipe gesturerecognizer api documentation. [Online]. Available: https://ai.google.dev/edge/api/mediapipe/python/mp/tasks/vision/GestureRecognizer

[9] T. Simon, H. Joo, I. Matthews, and Y. Sheikh, "Hand keypoint detection in single images using multiview bootstrapping," *Carnegie Mellon University*, 2017, https://arxiv.org/abs/1705.03468.

[10] F. Baldassarre, D. G. Morín, and L. Rodés-Guirao, "Deep koalarization: Image colorization using cnns and inception-resnet-v2," 2017. [Online]. Available: https://arxiv.org/abs/1712.03400

[11] K. D. Urmeneta and V. M. Romero, "Ensemble image colorization using convolutional neural network," in *2022 9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, 2022, pp. 355–360.

[12] R. Zhang, P. Isola, and A. A. Efros, "Colorful image colorization," in *ECCV*, 2016.

[13] M. Limmer and H. P. A. Lensch, "Infrared colorization using deep convolutional neural networks," *IEEE Transactions on Image Processing*, vol. 25, no. 6, pp. 2744–2757, 2016. [Online]. Available: https://ieeexplore.ieee.org/document/7458124

[14] S.-Y. Chen, J.-Q. Zhang, Y.-Y. Zhao, P. L. Rosin, Y.-K. Lai, and L. Gao, "A review of image and video colorization: From analogies to deep learning," *ACM Computing Surveys*, vol. 51, no. 6, pp. 1–34,

2018. [Online]. Available: https://dl.acm.org/doi/10.1145/3124391

[15] R. Zhang, J.-Y. Zhu, P. Isola, X. Geng, A. S. Lin, T. Yu, and A. A. Efros, "Real-time user-guided image colorization with learned deep priors," *ACM Transactions on Graphics (TOG)*, vol. 9, no. 4, 2017.

[16] Uni-Trend, "Uti721m thermal imaging camera," https://thermal.uni-trend.com/product/uti721m/, 2024, accessed: 15 Dec 2024.

[17] P. Musa, F. Rafi, and M. Lamsani, "A review: Contrast-limited adaptive histogram equalization (clahe) methods to help the application of face recognition," 10 2018, pp. 1–6.

[18] OpenCV Team. (2025) Opencv library. [Online]. Available: https://opencv.org/

[19] R. Zhang, "Colorization: Learning to color images from scratch," 2016, accessed: 2025-01-06. [Online]. Available: https://github.com/richzhang/colorization

[20] OECD.AI Policy Observatory. (2025) Percentage of correct keypoints (pck).

[21] K. Saleh, S. Szénási, and Z. Vámossy, "Occlusion handling in generic object detection: A review," *CoRR*, vol. abs/2101.08845, 2021. [Online]. Available: https://arxiv.org/abs/2101.08845

[22] G. Benitez-Garcia, J. Olivares Mercado, G. Aguilar-Torres, G. Sanchez-Perez, and H. Perez-Meana, "Face identification based on contrast limited adaptive histogram equalization (clahe)," vol. 1, 04 2012.

[23] P. Musa, F. A. Rafi, and M. Lamsani, "A review: Contrast-limited adaptive histogram equalization (clahe) methods to help the application of face recognition," in *2018 Third International Conference on Informatics and Computing (ICIC)*, 2018, pp. 1–6.

[24] D. Parashar, O. Mishra, K. Sharma, and A. Kukker, "Improved yoga pose detection using mediapipe and movenet in a deep learning model," *Revue d'Intelligence Artificielle*, vol. 37, pp. 1197–1202, 10 2023.

[25] A. Singh, A. Bay, and A. Mirabile, "Assessing the importance of colours for cnns in object recognition," *arXiv preprint arXiv:2012.06917*, 2020, zebra AI, Zebra Technologies, London, United Kingdom, {firstname.lastname}@zebra.com. [Online]. Available: https://arxiv.org/abs/2012.06917

A Appendix

A.1 Use of LLMs

During the research project, ChatGPT was used to check the grammatical correctness and clarity of the text. The prompts utilized included: "Could you fix any grammatical errors? [TEXT]", "Do any phrases sound out of place? Please list anything that does not seem right [TEXT]", and "Could you summarize the text and say it in your own words? [TEXT]". The last prompt was specifically used to ensure that the thoughts and ideas were clearly communicated.

Moreover, for additional support in grammatical issues Grammarly was used.