Bridging the Gap between Vitiligo Segmentation and Clinical Scores

Yanling Li* Steven Thng† Adams Wai-Kin Kong‡

May 28, 2024

Abstract

Quantitative evaluation of vitiligo is crucial for assessing treatment response. Dermatologists evaluate vitiligo regularly to adjust their treatment plans, which requires extra work. Furthermore, the evaluations may not be objective due to inter- and intra-assessor variability. Though automatic vitiligo segmentation methods provide an objective evaluation, previous methods mainly focus on patch-wise images, and their results cannot be translated into clinical scores for treatment adjustment. Thus, full-body vitiligo segmentation needs to be developed for recording vitiligo changes in different body parts of a patient and for calculating the clinical scores. To bridge this gap, the first full-body vitiligo dataset with 1740 images, following the international vitiligo photo standard, was established. Compared with patch-wise images, full-body images have more complicated ambient light conditions and larger variances in lesion size and distribution. Additionally, in some hand and foot images, skin can be fully covered by either vitiligo or healthy skin. Previous patch-wise segmentation studies completely ignore these cases, as they assume that the contrast between vitiligo and healthy skin is available in each image for segmentation. To address the aforementioned challenges, the proposed algorithm in this study exploits a tailor-made contrast enhancement scheme and long-range comparison. Furthermore, a novel confidence score refinement module is proposed to manage images fully covered by vitiligo or healthy skin. Our results can be converted to clinical scores and used by clinicians. Compared to the state-of-the-art method, the proposed algorithm reduces the average per-image vitiligo involvement percentage error from 3.69 % to 1.81 %, and the top 10 % per-image errors from 23.17 % to 8.29 %. Our algorithm achieves 1.17 % and 3.11 % for the mean and max error for the per-patient vitiligo involvement percentage, which is better than an experienced dermatologist's naked-eye evaluation.

---

*Y. Li is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore, 639798 (email: yanling001@e.ntu.edu.sg).

†S. Thng is with the School of Materials Science and Engineering, Nanyang Technological University, Singapore, 639798 and National Skin Center, Singapore, 308205 (email:steventhng@nsc.com.sg).

‡A.W.K. Kong is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore, 639798 (email: adamskong@e.ntu.edu.sg).

(a) Melanoma (b) Psoriasis (c) Vitiligo

Figure 1.1: The differences between melanoma [2], psoriasis, and vitiligo are as follows: Melanoma appears as dark spots, psoriasis lesions are typically red or pink patches covered by silvery scales, while vitiligo lesions are usually white patches. Melanoma typically presents as small spots, whereas psoriasis and vitiligo can spread over a large portion of the body's surface.

1. Introduction

Vitiligo is a skin disease caused by the loss of pigment. Accurate vitiligo assessment is important to diagnosis, prognosis, and treatment selection. Widely used clinical vitiligo assessment scores, such as the Vitiligo Extent Score (VES) [10], require dermatologists to estimate the vitiligo involvement percentage by the naked-eye. The scores are regularly estimated for tracking the change of vitiligo and adjusting the treatment promptly. Automatic evaluation is critical to alleviate the burden of dermatologists and avoid inter- and intra-accessor variability.

For skin cancers like melanoma [23] or basal cell carcinoma (BCC) [37], automatic segmentation methods mainly focus on patch-wise images. For example, the most widely used skin cancer dataset ISIC [13] contains thousands of patch-wise skin images that can be used to evaluate automatic segmentation methods. It can be observed from the images in ISIC that skin cancers, such as melanoma shown in Fig. 1.1(a), usually appear as small spots. Also, the clinical evaluation process of melanoma focuses on the skin patch. Therefore, it is reasonable that the automatic segmentation of melanoma images is primarily based on patch-wise images. However, skin diseases like psoriasis and vitiligo, as shown in Fig. 1.1(b) and (c), can affect a large portion of the body surface area, which differs from melanoma. Consequently, the clinical evaluation of vitiligo, such as Vitiligo Area Scoring Index (VASI) [17], is based on the percentage of the full-body surface involved in the disease. Aligning with the clinical evaluation, the vitiligo automatic methods should focus on full-body lesion segmentation. It should be highlighted that patch-wise segmentation is incapable of providing results for clinical usage.

Previous vitiligo automatic segmentation methods are either limited to patch-wise images [24, 25] or one particular body part, such as the trunk [33] or face [12, 39]. None of them consider the full-body evaluation, which makes the translation from automatic segmentation results to clinical vitiligo scores difficult. It makes the translation from automatic segmentation results to clinical vitiligo scores difficult. Full-body vitiligo segmentation is essential for clinical usage, but it is more complicated than patch-wise vitiligo segmentation. Firstly, full-body images have more complex ambient lighting conditions compared to patch-wise images. Secondly, vitiligo can spread over a large portion of the body's surface area. Some body part images are fully covered by vitiligo lesions, and these images are totally ignored in the patch-wise vitiligo dataset. In patch-wise segmentation, vitiligo lesions are always surrounded by healthy skin for comparison within the same image. For example, Fig. 1.2(a) shows images from a patch-wise vitiligo dataset [36], where vitiligo and healthy skin are present in the same image. The coexistence of vitiligo and healthy skin within the same image provides a reference for segmentation. Fig. 1.2(b) and (c) are hand images covered respectively only by healthy skin and only by vitiligo. It is challenging to localize the vitiligo and healthy skin in Fig. 1.2(b) and (c) due to the lack of a reference to the other category. However, such images (solely covered by vitiligo/healthy skin) do exist in real clinical applications, and they cannot be handled by patch-wise vitiligo segmentation methods. Moreover, patch-wise vitiligo segmentation fails to consider long-range comparisons, which are crucial for full-body vitiligo segmentation. Patch-wise segmentation mainly focuses on small patch images, while full-body segmentation requires referencing healthy skin from a broader context, as illustrated in Fig. 1.2(d). The vitiligo in the center of Fig. 1.2(d) needs to perform long-range comparisons with healthy skin in other regions.

In 2019, international vitiligo experts achieved a consensus for standardizing vitiligo photos [9] and recommended it for reproducible evaluations. In the standard, the international vitiligo experts defined standard positions of body parts to comprise a set of images for vitiligo patients, and the body surface area percentage of each image accounts for the total skin surface area. Per-image vitiligo involvement percentage can be obtained by calculating the number of predicted vitiligo and skin pixels from segmentation results. Combining the body surface area percentage with the per-image vitiligo involvement percentage, the per-patient vitiligo involvement percentage can be obtained. With a full-body dataset following the aforementioned international vitiligo photo standard (VPS) and an automatic full-body segmentation method, the segmentation result can be applied for clinical usage.

To bridge the gap between vitiligo segmentation methods and clinical scores, this paper makes the following contributions:

1. The authors point out that due to the discrepancy in clinical evaluations of skin cancers and vitiligo, segmentation of vitiligo should not follow the previous patch-wise scheme. Only full-body vitiligo segmentation, whose dataset follows the international vitiligo photo standard, can be applied for clinical usage. Thus, the first standard full-body vitiligo dataset is established, consisting of 1740 body part images. These images adhere to the international vitiligo photo standard.

2. A tailor-made contrast enhancement scheme is proposed to address the complicated light conditions of full-body images. To enable long-range comparisons between large lesions and healthy skin in the distance, the integrated successive dilation is employed in the proposed algorithm.

3. The authors reveal that in the full-body dataset, the pixel-wise intensity of vitiligo and

(a) 8 samples of patch-wise images

Figure 1.2: Comparisons between patch-wise vitiligo images (left) and body-parts vitiligo images (right). (a) vitiligo and healthy skin coexisted in all patch-wise images. (b) and (c) are hand images covered only by healthy skin and only by vitiligo, respectively. (d) the body-part vitiligo image needs long-range comparison for large lesion segmentation.

healthy skin has a large range of overlap due to complicated ambient light conditions. However, the mean intensity of vitiligo pixels is consistently higher than that of healthy skin pixels in the same image set. Based on this discovery, a novel confidence score refinement module is proposed to effectively handle images with pure healthy skin and pure vitiligo.

4. This paper conducts a comparison with a dermatologist's evaluation. The result from the proposed algorithm is better than an experienced dermatologist's naked-eye evaluation. The proposed algorithm achieves state-of-the-art performance in Intersection-over-Union, per-image involvement percentage, and per-patient involvement percentage.

In summary, the proposed algorithm improves the performance by including more healthy skin data and utilizing the proposed contrast enhancement, longer range comparison, and confidence score refinement schemes. This work presents the first vitiligo segmentation algorithm designed for full-body images, following the international vitiligo photo standard proposed by medical experts. Following the clinical standard ensures that the developed method is valuable for clinicians. This work is important as it addresses the real-world vitiligo automatic diagnosis problem: how to automatically diagnose vitiligo lesions when vitiligo and healthy skin do not coexist in the same image. Previous works have entirely ignored these cases and have performed poorly on images fully covered only by vitiligo or skin due to the lack of contrast. The proposed algorithm utilizes cross-referencing between images from the same patient to establish contrast between vitiligo and healthy skin, providing a practical solution for these cases. As a result, this work generates clinical vitiligo scores automatically and offers an effective way to bridge the gap between image segmentation methods and clinical applications. The proposed algorithm provides a slightly better estimation than the naked-eye estimation of an experienced dermatologist, reducing the workload of dermatologists. The research was approved under National Healthcare Group, Singapore, DSRB (Domain Specific Review Board) Protocol Number 2014/01015 on 26th April 2018.

The rest of the paper is organized as follows. Section II summarizes the previous related works and their drawbacks. Section III provides the details of the datasets used in this work. Section IV describes the proposed algorithm, and Section V presents the experimental results. Section VI concludes this work.

2. Related Work

This section gives an overview of the previous works related to full-body vitiligo segmentation. It begins with previous work in bridging the gap between medical imaging and clinical usage. Then the literature on automatic skin lesion diagnosis is reviewed. The difference in clinical evaluation metrics for skin cancer and vitiligo indicates that the automatic vitiligo segmentation methods should not follow the previous patch-wise scheme and full-body vitiligo segmentation should be established for clinical usage. In the last part of this section, previous vitiligo segmentation methods are also reviewed.

2.1. Efforts for Bridging the Gap Between Medical Imaging and Clinical Usage.

The emergence of artificial intelligence boosts the accuracy of image-based automatic diagnosis. However, the discrepancy between general image recognition methods and requirements of automatic diagnosis introduces gaps in clinical translation. These gaps may render developed methods ineffective for clinicians. The first concern is that the evaluation metrics of general image recognition methods are inadequate to describe patient conditions, and clinical evaluation is essential for assessing these methods. In organ delineation, Xu *et al.* [30] discovered that widely used evaluation metrics, such as the Dice score, are inadequate for describing the results in delineation. They suggested that the evaluation should incorporate physician-reported assessments. Similarly, Diwakar *et al.* [7] suggested that automatic diagnosis should be assessed using both statistical metrics and observations from medical experts. The second concern is variations in data collected from different institutions. These variations result in non-independently and identically distributed (non-iid) datasets among data collected by different institutions. To address this issue, researchers proposed innovative frameworks in federated learning [38] and feature ensemble techniques [27] to enhance the generalization capability for non-iid data in retinal and COVID-19 datasets, respectively. To better address the variations, medical scientists also proposed numerous consensus-based guidelines to standardize AI in healthcare fields. For instance, a consensus-based guideline known as the Image Biomarker Standardization Initiative was published [42]. This initiative aims to standardize each step from data acquisition to post-analysis techniques in image-based biomarker processing. Similarly, medical scientists also developed standards and guidelines for pathology research [29], particularly in the context of next-generation sequencing (NGS), to reduce variations in tissue acquisition and method validation for pathology imaging. Physicians specializing in hepatic tumors proposed a recommendation for image-guided ablation methods development for liver tumors [19]. Similarly, to standardize vitiligo segmentation, vitiligo experts proposed an international vitiligo photo standard [9]. The standard observed that only standard full-body segmentation can be translated to vitiligo clinical scores. However, previous vitiligo segmentation approaches primarily focused on patch-wise segmentation, and none of them followed this standard.

2.2. The Gap Between Segmentation Results and Clinical Evaluation Metrics in Skin Images

The majority of skin lesion segmentation datasets consist of patch-wise images. This is reasonable for cancerous lesions such as melanoma. Because the clinical evaluation process for these types of cancer primarily concentrates on the local skin region initially and subsequently determines whether the lesion is cancerous or not, patch-wise segmentation results for cancerous skin diseases [14–16, 18] closely align with the clinical evaluation metrics. Following the trends of these automated segmentation methods for cancerous skin diseases, patch-wise vitiligo [24, 25] segmentation methods were developed to alleviate the workload of dermatologists. However, these methods ignored the fact that clinical evaluation metrics for vitiligo are different from those for cancerous skin diseases. As shown in Fig. 1.1, vitiligo can affect a large skin surface, in contrast to the small-size skin cancer lesions. Therefore, the widely used clinical evaluation scores for vitiligo [10, 17] are based on the percentage of lesion involvement on the full-body skin surface. Due to the requirement of full-body clinical evaluation, patch-wise vitiligo segmentation results are not sufficient for clinical usage. Full-body vitiligo segmentation methods must be developed for clinical applications.

2.3. Previous Vitiligo Segmentation Datasets and Methods

For studying vitiligo segmentation, researchers have established several datasets. However, these datasets focus either on patch-wise images [4] [36], or one particular body part [39] [12]. With the above-mentioned vitiligo datasets, the majority of previous vitiligo segmentation methods deal with patch-wise images [4, 24, 25]. ISBI20 [21] segments vitiligo lesions by directly applying InceptionV2 [32] to patch-wise vitiligo images. Bian *et al.* [4] proposed a weakly supervised classification-segmentation workflow for patch-wise vitiligo image segmentation. Toh *et al.* [33] segmented trunk images using linear spectral clustering (LSC) superpixel [1]. Li *et al.* [39] proposed a face vitiligo segmentation method based on special synthetic face images and wild-position images. HybridSeg [12] directly applies YOLO3 [28] and UNet++ [41] as a detection-segmentation workflow to vitiligo segmentation. The previous methods focus on patch-wise or particular body-part segmentation, such that their results do not align with vitiligo clinical metrics.

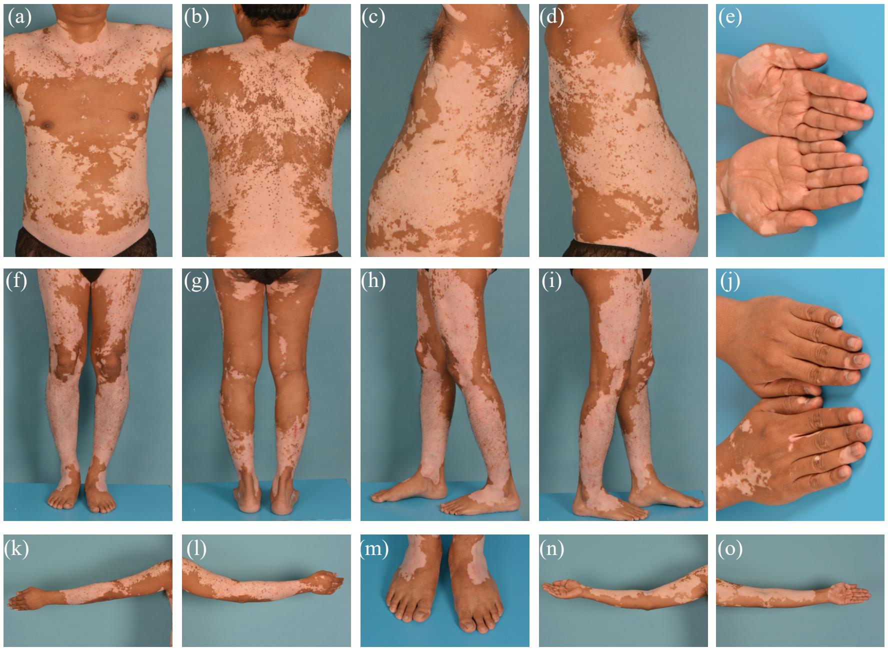


Figure 2.1: An example of an image set following the international vitiligo photo standard. There are 15 images in a set include a) trunk(front), b) trunk (back), c) trunk (left), d) trunk(right), e) palm, f) legs(front), g) legs(back), h) legs (left), i) legs (right), j) dorsum of the hand, k) left arm (back), l) right arm (back), m) feet, n) right arm (front), and o) left arm (front).

(a) Samples from SFVD

(b) Samples from NAFD

Figure 2.2: Some samples from (a) SFVD and (b) NAFD. SFVD is the clinical vitiligo full-body dataset, and it follows the international vitiligo photo standard. NAFD is the healthy non-standard body-part images from a public dataset [3].

3. Datasets

Two datasets are introduced in this section. The first one is the main dataset used for training and testing in this research. It is a full-body standard dataset that follows the

Table 3.1: Standard Full-body Vitiligo Dataset information, categorized by patients and by body-part name.[2]

by patient	patient id	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	no.of img sets	11	7	5	7	8	2	5	3	7	7	7	1	5	6	6	7	6	7	7	6

by body-part name	body-part name	no. of imgs	body-part name	no. of imgs	body-part name	no. of imgs
	(a) trunk (front)	111	(b) trunk (back)	113	(c) trunk (left)	112
	(d) trunk (right)	112	(e) palm	114	(f) legs (front)	120
	(g) legs (back)	120	(h) legs (left)	120	(i) legs (right)	120
	(j) dorsum of the hand	114	(k) left arm (back)	120	(l) right arm (back)	120
	(m) feet	104	(n) right arm (front)	120	(o) left arm (front)	120

international vitiligo photo standard. The second dataset consists of publicly available non-standard healthy skin images. The inclusion of additional images provides the segmentation model with a greater number of healthy skin samples to suppress false positives.

3.1. Standard Full-body Vitiligo Dataset (SFVD)

SFVD contains 1740 clinical body-parts images collected by the National Skin Center, Singapore, according to VPS [9]. 20 patients contributed to the dataset, and each patient has serial sets of images as recorded in Table 3.1. It is the first standard full-body vitiligo segmentation dataset and the largest vitiligo segmentation dataset in terms of image number. An example of an image set is shown in Fig. 2.1 and details of the image number for each body-part are recorded in Table 3.1. Each body-part surface area percentage is included in the international vitiligo photo standard [9]. The resolutions of the vertical and horizontal raw images are respectively 3000 by 4500 pixels and 4500 by 3000 pixels. These raw images were resized to 508 pixels for the longer side while maintaining the aspect ratio and padded to 508 pixels for the shorter side to create the final dataset. The annotation of the images was conducted by part-time workers and checked by a dermatologist. The masks of annotation are grayscale images, where pixel values of 0, 1, and 2 indicate background, healthy skin, and vitiligo, respectively. Underwear, personal accessories, and hair are marked as backgrounds in our dataset. These images were randomly split to 5-fold for cross-validation. Images from the same patient are always in either training or testing sets to avoid information leakage.

3.2. Normal Arm and Foot Dataset (NAFD)

The additional healthy skin dataset is introduced in this research for false positive suppression. NAFD consists of 60 upper arm images, 59 foot images, 64 forearm images, and 177 palm images. These images were collected from a healthy people public dataset [3], which does not follow the vitiligo photo standard. These images were processed in the same manner as SFVD. Some samples from SFVD are depicted in Fig. 2.2(a) and NAFD is depicted in Fig. 2.2(b).

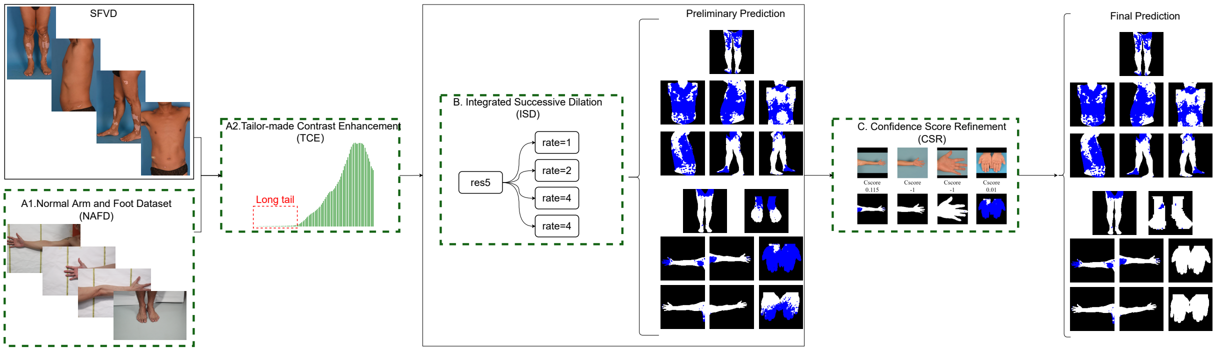


Figure 3.1: A block diagram of the proposed algorithm. The new modules are highlighted in the dark green dash boxes. The four new modules in the proposed algorithm are NAFD dataset, TCE, ISD, and CSR.

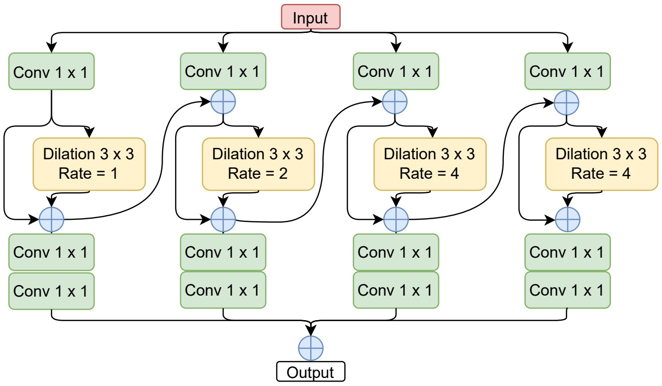


Figure 3.2: The architecture of the integrated successive dilation (ISD) module. ISD combines features from four parallel convolution branches with different dilation rates.

4. The Proposed Algorithm

The proposed algorithm is designed based on the characteristics of full-body vitiligo images and is summarized in Fig. 3.1. New blocks of the proposed algorithm are highlighted in green dashed boxes. On the input side, additional healthy skin images, denoted as NAFD in Fig. 3.1, are used to expose the segmentation model to a broader range of healthy skin variations. Furthermore, a tailor-made contrast enhancement scheme, labeled as TCE in Fig . 3.1, is proposed to provide the segmentation model with more reliable features. An integrated successive dilation module, marked as ISD in Fig. 3.1, is employed to perform long-range comparisons for large lesions. Lastly, a novel confidence score refinement module, marked as CSR in Fig. 3.1, is proposed to correct the preliminary segmentation. The research was approved under National Healthcare Group, Singapore, DSRB (Domain Specific Review

Board) Protocol Number 2014/01015 on 26th April 2018.

4.1. Normal Arm and Foot Dataset (NAFD) and the Tailor-made Contrast Enhancement (TCE)

In medical settings, the dataset is usually small compared to other generic datasets. Small datasets may lead to bias in the training data, especially for the settings in our study where each fold of the training-testing split is based on patients, instead of images. In each fold of the evaluation, the training dataset contains only skin images of around 16 patients. Although the international vitiligo photo standard has reduced some variations in the images, these small training sets make it hard to cover diverse human skin colors and textures. To alleviate this, we introduce NAFD to provide a wider range of skin colors and textures to the model in order to handle more diverse types of skin.

The segmentation performance of vitiligo is highly affected by light conditions. Images from the same patient captured in different light conditions have various contrast levels between vitiligo and skin. Low contrast between vitiligo and skin contributes to hard samples in segmentation. Contrast enhancement can alleviate this issue and provide a reliable feature. The widely used contrast enhancement [26] is not effective enough for vitiligo pixels due to long-tail pixel intensity problems. For example, in Fig. 2.1(a) to (d), the darkest pixel is from underwear or body hair. Standard contrast enhancement increases the contrast between vitiligo and underwear, rather than the contrast between vitiligo and healthy skin. Thus, we modified the widely used module by adding a long-tail check such that it can remove the long-tail background pixels and focus on the contrast between vitiligo and healthy skin. The enhancement is formulated as:

$$I_o = (I_i - Min_i) * \frac{Max_o - Min_o}{Max_i - Min_i} + Min_o \tag{4.1}$$

where I_i and I_o are the input and output pixel values in one of the RGB channels, respectively. The contrast parameters are sampled from intensity images and the same set of parameters are applied to each of the RGB channels. Max_i and Min_i are the maximum and minimum pixel values of the input intensity image and Max_o and Min_o are the maximum and minimum pixel values of the output intensity image. Max_o is uniformly sampled from [0.8, 1.0] and Min_o is uniformly sampled from [0, 0.2]. To obtain the value of Max_i and Min_i, two values ${Max_i}^*$ and ${Min_i}^*$are first uniformly sampled from [0.8, 1.0] and [0, 0.2], respectively. The pixels between the 0-1st percentile and between the 99-100th percentile are regarded as the long tail. The cut-off intensity of the long tail for both sides is denoted as $[L_{min}, L_{max}]$. To avoid losing too much pixel information during the enhancement, we set $Min_i = min\,(L_{min}, {Min_i}^*)$. The same operation is applied to the high-intensity end of the image. Thus, this module stretches or squeezes the input image to a random output range [Min_o, Max_o] with a function to reduce the effect of long-tailed pixels.

---

[2] Around 10% of the total image sets lack one or two images from the standard 15 body-part sets. Due to the costly nature of collecting medical images, we have included these sets in our dataset.

4.2. Integrated Successive Dilation (ISD)

In cases where there is low contrast between vitiligo and healthy skin in large vitiligo lesions, our baseline [40] model struggles to identify the interior portion of the lesion. In such cases, the struggle may be related to the lack of interior feature enhancement and long-range comparisons with healthy skin areas at a distance in our baseline model. Inspired by [11, 31], we utilize ISD to perform long-range comparisons to enhance the interior features of large lesions. ISD combines features from multiple convolutional layers with different dilation rates. As shown in Fig. 3.2, we use four parallel branches but the last two branches have the same dilation rates. This block is implemented right after the last layer of our backbone.

4.3. Confidence Score Refinement (CSR)

In the patch-wise vitiligo image studies, the vitiligo lesions are always surrounded by healthy skin in the same image, as shown in Fig. 1.2(a). Thus, the contrast between vitiligo and healthy skin can be used to distinguish vitiligo from healthy skin. However, some clinical vitiligo images following the international standard have only vitiligo or healthy skin. As a result, their contrast is not available for segmentation and diagnoses. The problem is common on body parts with small surface areas such as hands and feet. The experiments show that the state-of-the-art methods cannot handle these images well. Fig. 4.1 lists five SFVD images with the highest segmentation errors, revealing that the results with the highest errors are dominated by hand images with only vitiligo or with only healthy skin. To address this challenge, we proposed a novel confidence score refinement module.

Before presenting our refinement module, we first analyze the pixel intensity distribution of vitiligo and healthy skin. With complicated ambient light conditions on the skin surface area of full-body images, it is not guaranteed that the pixel-wise intensity of vitiligo is always brighter than healthy skin, especially when comparing vitiligo and healthy skin pixels from different body parts. Fig. 4.2 shows the boxplot for the intensity of vitiligo and healthy skin from 30 sets of images of patients. Each set consists of 15 body part images as shown in Fig. 2.1. In each set, the red boxplot represents the pixel-wise intensity distribution of vitiligo, while the blue boxplot represents the pixel-wise intensity distribution of healthy skin. The red box, along with the nearest blue box, represents the intensity distributions of vitiligo and healthy images in one image set. The first observation from Fig. 4.2 is that the intensity of pixel-wise vitiligo and healthy skin overlaps to a large extent, which supports that vitiligo is not always brighter than healthy skin. The variation in pixel-wise vitiligo intensity poses challenges in the segmentation of vitiligo in full-body images. However, it can be observed that the mean intensity of vitiligo, represented by a small triangle within the red boxplot, is consistently higher than that of the healthy skin, across all 30 image sets.

We were inspired by these observations and set the design principle for the confidence score refinement: segmentation may be more accurate if the mean intensity of the segmented vitiligo is higher than that of healthy skin. Although there is variation in the means of vitiligo and healthy skin, considering that the model may mistakenly predict healthy skin as vitiligo and vice versa, a greater discrepancy between the two means leads to a more confident

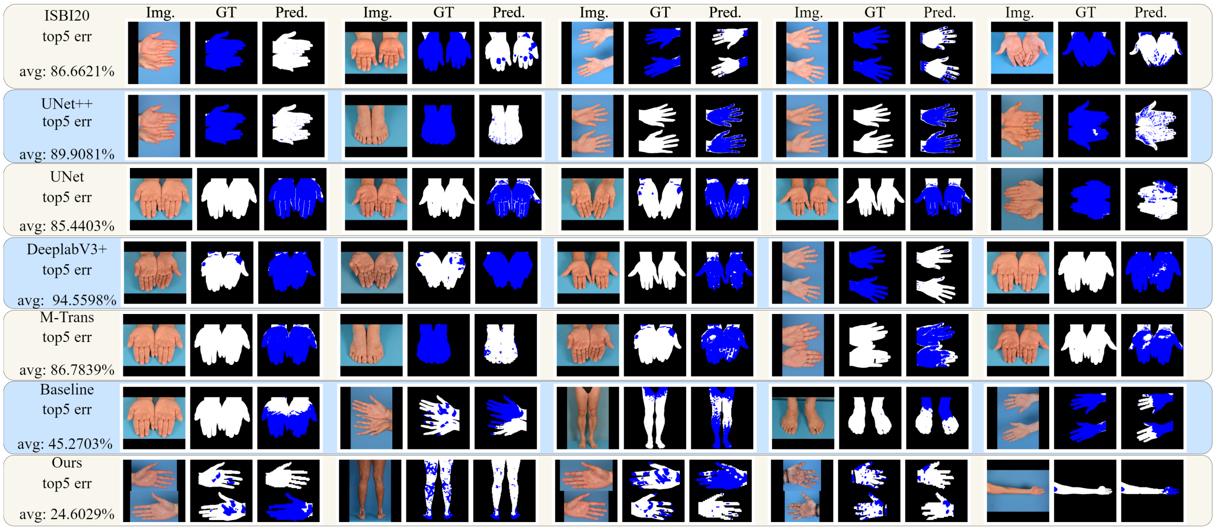


Figure 4.1: The top 5 prediction errors of the previous works, the baseline, and the proposed algorithm. *Img.*, *GT*, and *Pred.* stand for the original image, ground-truth, and segmentation results, respectively. *avg* on the left side of each method stands for the average percentage of the top 5 errors. Top error images are mainly from hand and foot images fully covered by only vitiligo or only healthy skin.

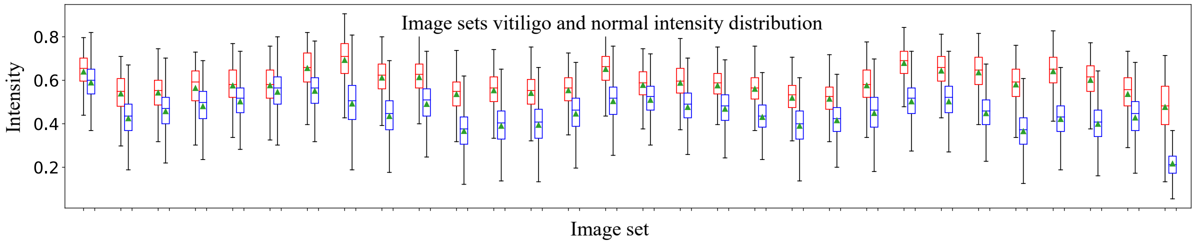
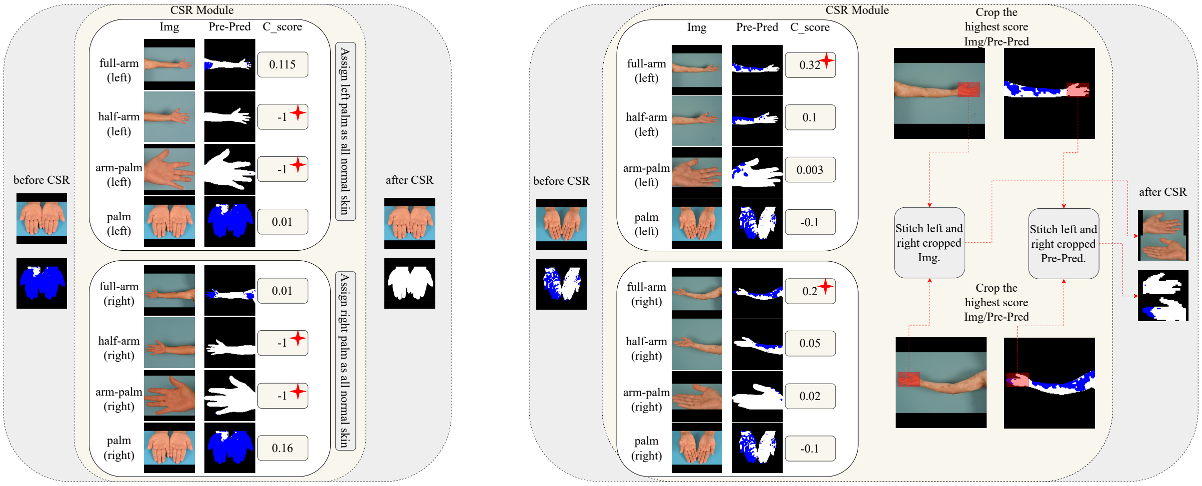


Figure 4.2: Boxplots of vitiligo and healthy skin intensity from 30 patient image sets. Each set consists of 15 body part images as shown in Fig. 2.1. In each set, the red boxplot represents the pixel-wise intensity distribution of vitiligo, while the blue boxplot represents the pixel-wise intensity distribution of healthy skin. The red box, along with the nearest blue box, represents the intensity distributions of vitiligo and healthy images in an image set.

prediction. Based on the design principle, we propose the confidence score refinement (CSR). The refinement applies to small body parts such as the palm, dorsum of the hand, and foot images as shown in Fig. 2.1(e), (j), and (m). We take palm refinement as an example and the steps are described as follows. Firstly, the trained model segments full-arm images, half-arm images, palm-arm images, and palm images. The half-arm images and palm-arm images are cropped from the full-arm images. The coordinates of the palm in the arm images are detected by the hand keypoint detection tool Mediapipe [22]. With all images and their predictions for full-arm, half-arm, palm-arm, and palm images, the confidence scores are calculated according to Eq. 4.2:

$$C_{score} = \begin{cases} -1 & \text{if all } fg \text{ pixels predicted as healthy skin} \\ 1 & \text{if all } fg \text{ pixels predicted as vitiligo} \\ \frac{1}{V}\sum_1^V I_v - \frac{1}{S}\sum_1^S I_s & \text{otherwise} \end{cases} \tag{4.2}$$

where fg stands for the foreground, V and S are numbers of predicted vitiligo and healthy skin pixels. I_v and I_s are the values of the corresponding pixels in a normalized grayscale testing image. C_{score} is calculated based on all pixels in the image. The principle of the refinement is that by cross-referencing all predictions related to palm regions, the most confident prediction is selected as the final prediction. The first step for the refinement is to construct $C_{scorelist}$ for the left and right palms. The $C_{scorelist}$ consists of four C_{score} values from the arm-palm list (ArP), which includes full-arm, half-arm, palm-arm, and palm. There are four refinement cases. Case 1 is when both left and right $C_{scorelist}$ have two or more C_{score} calculated as -1. It indicates that the model believes both palms should be all healthy skin. The reason is that all four predictions of ArP include the palm region. If at least two of these predictions are all healthy skin, then the final prediction of the palms should be set as all healthy skin. One example of case 1 is shown in Fig. 4.3(a). Case 2 has two or more C_{score} calculated as 1 in both left and right $C_{scorelist}$. It is similar to case 1, and the final prediction of both palms should be all vitiligo. The remaining two cases are based on selecting the highest score predictions. Case 3 is when the highest C_{score} in the left or the right $C_{scorelist}$ corresponds to the original palm prediction. The original palm prediction is kept unchanged in case 3. If the $C_{scorelist}$ of ArP does not fall into the aforementioned three cases, then it falls into case 4. Case 4 processes the left and right palms separately and combines them as the final palm prediction. For the left $C_{scorelist}$, the prediction corresponding to the highest C_{score}, denoted as $Pred_{left}$, is selected as the most confident prediction. Img_{left} is the image that corresponds to the $Pred_{left}$. Note that $Pred_{left}$ is the segmentation result, not a score. Then, hand detection is performed on Img_{left} to obtain the hand bounding box $bbox_{left}$. $Pred_{left}$ and Img_{left} are cropped by $bbox_{left}$. The cropped image and prediction are denoted as $Imgcrop_{left}$ and $Predcrop_{left}$. The right-side ArP is processed similarly to get $Imgcrop_{right}$ and $Predcrop_{right}$. The $Imgcrop_{left}$ and $Imgcrop_{right}$ are combined as a single image, which serves as the output image. Similarly, $Predcrop_{left}$ and $Predcrop_{right}$ are combined as the output prediction. Fig. 4.3(b) shows an example of case 4. The refinement deals with the same body parts of the same person but may correspond to different images, e.g. palms from front arm images and palm images. However, the ground-truth of vitiligo involvement percentage is kept almost unchanged before and after the refinement as shown in Fig. 4.4. The refinement for the dorsum of the hand is the same as the palm. To detect the foot keypoints, the Openpose [5] tool is used in front leg images.



(a) CSR-- Case 1 Example (b) CSR -- Case 4 Example

Figure 4.3: (a) An example of CSR case 1. Firstly, $C_{scorelist}$ is calculated for the left and right palms. The $C_{scorelist}$ contains C_{score} of full-arm, half-arm, palm-arm, and palm. In *case* 1, both left and right $C_{scorelist}$ have two or more -1. According to the refinement rule, the refined palm prediction is set as all healthy skin. (b) An example of CSR *case* 4. In *case* 4, the left and right palms are processed separately and then combined as the final palm prediction. For the left $C_{scorelist}$, the prediction corresponding to the highest C_{score}, denoted as $Pred_{left}$, is selected as the most confident prediction. Img_{left} is the image that corresponds to the segmentation result $Pred_{left}$. Then, hand detection is performed on Img_{left} to obtain the hand bounding box $bbox_{left}$. $Pred_{left}$ and Img_{left} are cropped by $bbox_{left}$. The cropped image and prediction are denoted as $Imgcrop_{left}$ and $Predcrop_{left}$. The four right palm images are processed similarly to get $Imgcrop_{right}$ and $Predcrop_{right}$. The $Imgcrop_{left}$ and $Imgcrop_{right}$ are combined as a single image, which serves as the output image. Similarly, $Predcrop_{left}$ and $Predcrop_{right}$ are combined as the output prediction.

5. Experiments

5.1. Experimental Settings

The GLNet [40], originally designed for satellite images containing scattered small objects and featuring a multi-layer smoothing function, is selected as the *Baseline*. This choice is made because both vitiligo and satellite images consist of numerous scattered small objects, unlike the large and salient objects typically found in PASCAL [8]. GLNet uses a ResNet50 with FPN as the backbone and employs the focal loss [20] with $\gamma = 6$ for training. We follow the learning rate decay scheme from [40]. For a fair comparison, we use ResNet50 as the backbone of UNet [35], UNet++ [41] and DeepLabv3+ [6]. LSC [33] and ISBI20 [21] use their original backbones because the backbone selection is a part of their methods for vitiligo segmentation. The backbone of M-trans [34] is built with their gated axial attention block. To obtain a more comprehensive evaluation, all previous methods and the baseline are

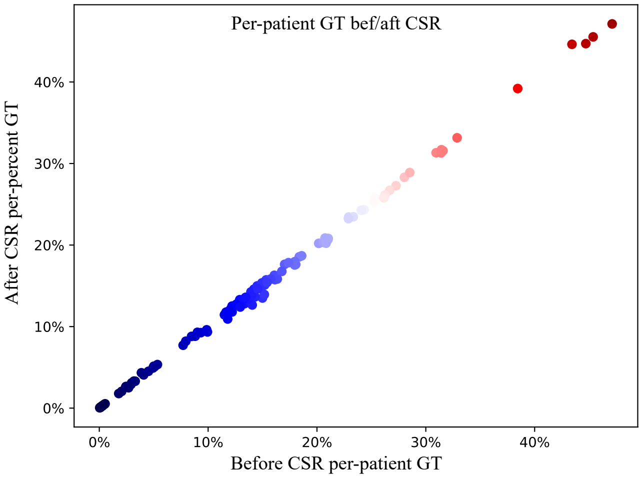


Figure 4.4: Per-patient ground-truth percentage before and after the refinement.

Table 5.1: Results of the previous methods and the proposed algorithm.

method	year	skin	vitiligo	mean
LSC [33]	2018	0.7960 (±0.04)	0.3678 (±0.03)	0.5819 (±0.02)
ISBI20 [21]	2020	0.8920 (±0.03)	0.6639 (±0.08)	0.7780 (±0.05)
UNet [35]	2018	0.9022 (±0.02)	0.6493 (±0.06)	0.7758 (±0.03)
UNet++ [12]	2022	0.9053 (±0.02)	0.6526 (±0.08)	0.7790 (±0.04)
Deeplabv3+ [6]	2018	0.8870 (±0.04)	0.6972 (±0.06)	0.7921 (±0.05)
M-trans [34]	2021	0.9322 (±0.02)	0.7417 (±0.08)	0.8370 (±0.04)
Ours	-	0.9223 (±0.01)	0.7758 (±0.05)	**0.8491 (±0.03)**

evaluated using a 5-fold cross-validation with patient-based data splitting to prevent overlap between the training and testing sets. 20% of the 1740 images were used as the testing set in each fold. Due to the training speed, only the global branch of GLNet is applied in our algorithm.

5.2. Evaluation and Ablation Studies from Engineering Perspectives

The proposed algorithm is evaluated following the experimental settings detailed in the previous section. Table 5.1 reports mean IoU (mIoU) for the foreground of the previous methods and the proposed algorithm. The mean values of the 5-fold cross-validation result are reported in the table and the standard derivations among 5 folds are recorded in the brackets. Our proposed algorithm achieves an mIoU of 0.8491 and outperforms state-of-the-art medical segmentation network M-Trans with an mIoU of 0.8370. Besides, compared to M-Trans, the proposed algorithm was trained in 4 hours using four GPU cards, significantly shorter than the M-trans training time of 33 hours on the same four GPU cards. We opted not to compare our results with Li *et al.* 's method based on synthetic images [39] due to

Table 5.2: Ablation study of proposed modules.

model	additional modules	skin	vitiligo	mean
Baseline	-	0.9201 (±0.01)	0.7560 (±0.05)	0.8381 (±0.03)
Baseline_N	NAFD	0.9211 (±0.01)	0.7591 (±0.05)	0.8401 (±0.02)
Baseline_NI	NAFD, ISD	0.9228 (±0.01)	0.7613 (±0.05)	0.8421 (±0.02)
Baseline_NIE	NAFD, ISD, TCE	0.9251 (±0.01)	0.7700 (±0.05)	0.8476 (±0.02)
Our result	NAFD, ISD, TCE, CSR	0.9223 (±0.01)	0.7758 (±0.05)	**0.8491 (±0.03)**

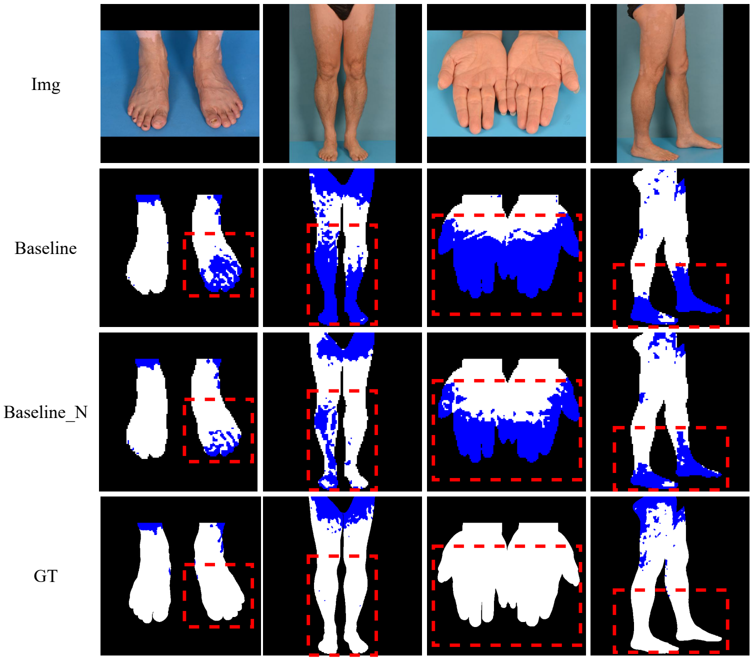


Figure 5.1: The comparison of predictions with and without NAFD. From top to down, they are test images, predictions without NAFD, predictions with NAFD, and ground truth denoted respectively as *Img*, *Baseline*, *Baseline_N*, and *GT*. Red dash boxes highlight the differences between predictions with and without NAFD.

two main reasons: Firstly, the synthesis method [39] is designed for face images, not for full-body images, and based on face detection. Secondly, while standard frontal face datasets are abundant for synthesis, it is challenging to find standard full-body images.

The ablation study of the proposed modules is recorded in Table 5.2. The baseline with NAFD is denoted as *Baseline_N*. By adding NAFD to the training set, the foreground mIoU is improved to 0.8401. NAFD provides a wider range of skin colors and textures to the model, allowing it to handle a more diverse range of skin types. The impact of NAFD is visualized in Fig. 5.1. The test images, the baseline predictions without NAFD (denoted as *Baseline_N*), the predictions with NAFD (denoted as *Baseline_N*), and the ground truth (denoted as *GT*) are shown from top to bottom. As highlighted by the red dashed boxes in Fig. 5.1, this demonstrates that false positives of vitiligo detection are reduced by providing more samples of healthy skin to the network. The third row of Table III

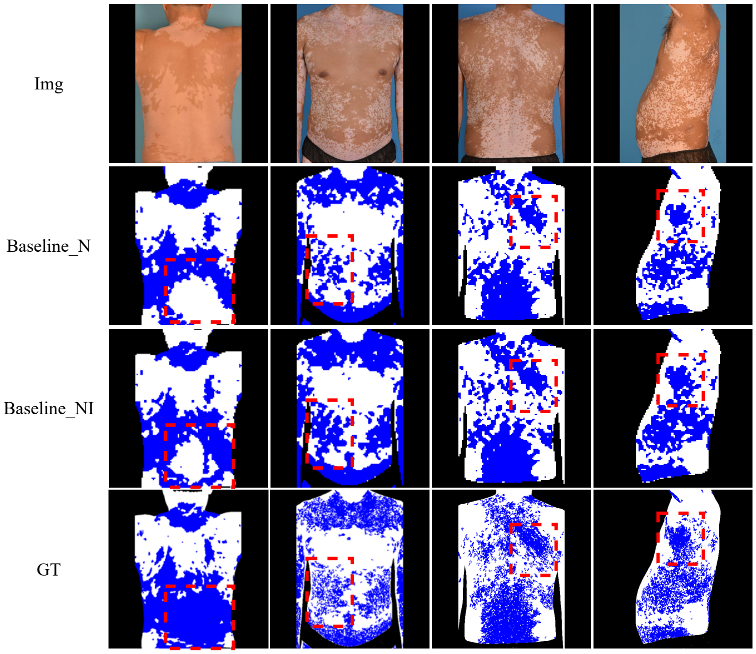


Figure 5.2: The comparison of predictions with and without ISD. From top to down, they are test images, predictions without ISD, predictions with ISD, and ground truth denoted respectively as *Img*, *Baseline_N*, *Baseline_NI*, and *GT*. Red dash boxes highlight the differences between predictions with and without ISD.

lists the results with NAFD and ISD [31], named *Baseline_NI*. ISD further enhances the performance to 0.8421 by enabling long-range comparisons for large lesions, as illustrated in Fig. 5.2. The mIoU is subsequently boosted to 0.8476 by incorporating the proposed contrast enhancement scheme (referred to as *Baseline_NIE*). This enhancement scheme improves both background/foreground segmentation and the segmentation of low-contrast images, as depicted in Fig. 5.3. The proposed enhancement scheme, which offers more reliable features through comparisons between healthy skin, background, and vitiligo pixels, reduces incorrect segmentation of background (e.g., patient's bra) and lesions with low contrast. The outcome of the proposed algorithm with the CSR model performing cross-reference to reduce segmentation errors of hands and feet is documented in the last row of Table III. It surpasses the baseline in vitiligo segmentation by a margin of 2%. Consequently, images that only have vitiligo or healthy skin can be more accurately segmented. Examples of predictions before and after the refinement process are presented in Fig. 4.3.

5.3. Evaluation from Clinical Perspectives

Vitiligo is a skin disease that can spread to a large ratio of the body's surface area. Additionally, the clinical evaluation metric of vitiligo involves calculating the percentage of vitiligo on all healthy and vitiligo-affected skin. The per-image vitiligo involvement percentage is defined as the ratio of vitiligo to all skin, including both healthy and vitiligo-affected skin.

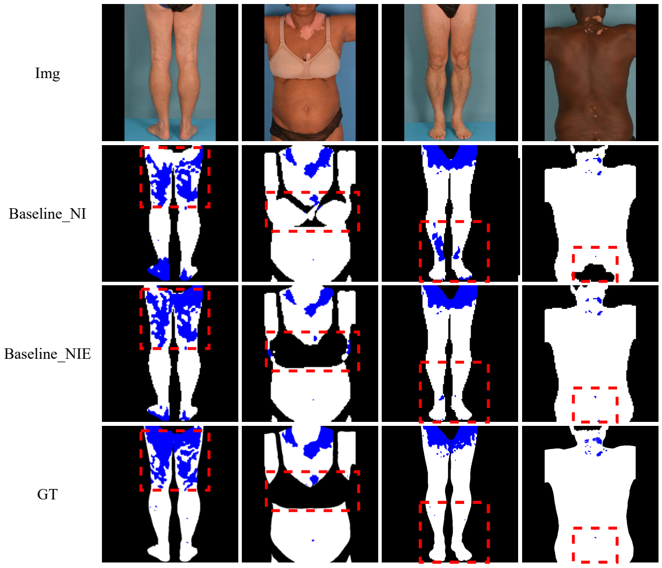


Figure 5.3: The comparison of predictions with and without the tailor-made contrast enhancement (TCE). From top to down, they are test images, predictions without TCE, predictions with TCE, and ground truth denoted respectively as *Img*, *Baseline_NI*, *Baseline_NIE*, and *GT*. Red dash boxes highlight the differences between predictions with and without TCE.

Table 5.3: Per-image vitiligo involvement percentage error.

Image	Mean error	Top 1 error	Top 10 error	Top 20 error
percentile	100	0.06	0.57	1.15
M-trans [34]	3.69%	98.84%	79.47%	69.43%
Our result	1.81%	34.14%	21.53%	18.27%

However, the IoU only measures the overlap between the prediction and the ground-truth, which fails to measure the previously mentioned vitiligo percentage.

For example, in Fig. 5.4, two predictions are recorded as (a) and (b). For each prediction, the predicted vitiligo is marked as a red box, the predicted/ground-truth skin as a black box, and the ground-truth vitiligo as a blue box. In Fig. 5.4(a), the predicted vitiligo to skin percentage is 13%, while the ground-truth vitiligo to skin percentage is 0.8%, resulting in a difference of 12.2% between the prediction and the ground-truth percentages. In Fig. 5.4(b), the difference in percentage between the prediction and the ground-truth is much larger at 49.6%. From a clinical perspective, prediction (b) is considered much less accurate than prediction (a). However, if we evaluate (a) and (b) solely based on IoU, both of them have the same IoU value of 5.3%.

This example demonstrates why it is necessary to evaluate the proposed algorithm using

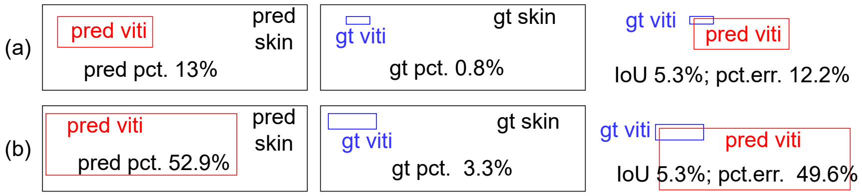


Figure 5.4: Vitiligo involvement percentage vs IoU evaluation. Two predictions are recorded as (a) and (b). For each prediction, the predicted vitiligo is marked as a red box, the predicted/ground-truth skin is marked as a black box, and the ground-truth vitiligo is marked as a blue box. For each prediction, from left to right are predictions, ground-truths, and the results of two evaluation metrics, IoU and percentage error.

both mIoU and the vitiligo involvement percentage. While mIoU measures spatial overlap, the vitiligo involvement percentage provides insight into the extent of the disease on the overall skin area. Evaluating the model using both metrics ensures a more comprehensive assessment of its performance.

5.3.1. Per-Image Vitiligo Involvement Percentage (PIVP)

PIVP is the initial metric chosen for assessing performance from a clinical standpoint. PIVP is defined as the ratio of vitiligo area over skin area on an image. The error between the estimation and ground-truth is an indicator of the performance of the algorithm. The average of the top 1, 10, and 20 errors are given in Table 5.3. The results show that the proposed algorithm reduces top errors in a huge range as compared to the previous best method, M-trans. The error distributions of M-trans, baseline, and the final result are recorded in Fig. 5.5(a), (b), and (c). For a better illustration, the 18% to 99% error range is zoomed in the same figures. The first observation is that our baseline outperforms the M-Trans in max error. The PIVP max error of M-Trans is 98.84% while the max error of the baseline is 62.57%. PIVP max error is further reduced to 34.14% with our proposed algorithm. Our proposed algorithm performs not only well in reducing top errors but also excellent in accurate segmentation. The left-most bars of Fig. 5.5(a), (b), and (c) indicate the predictions with less than 1% error. Ours has much more (around 7% of the total image number) accurate segmentation than M-trans.

Fig. 5.6(a) is the comparison of the error for every 10 percentiles of the images to the baseline. Fig. 5.6(b) is the same comparison with all the previous methods. In these two figures, the X-axis is the average per-image error of the top y percentile of total images, where y is the value from the Y-axis. Therefore, lower lines correspond to more accurate methods. Fig. 5.6(a) shows that our algorithm has an all-image average error of 1.814%, outperforming the baseline whose average error is 2.227%. Fig. 5.6(b) shows that our result outperforms all the previous methods. The figures and tables in this section show the proposed algorithm consistently outperforms the baseline and the previous methods.

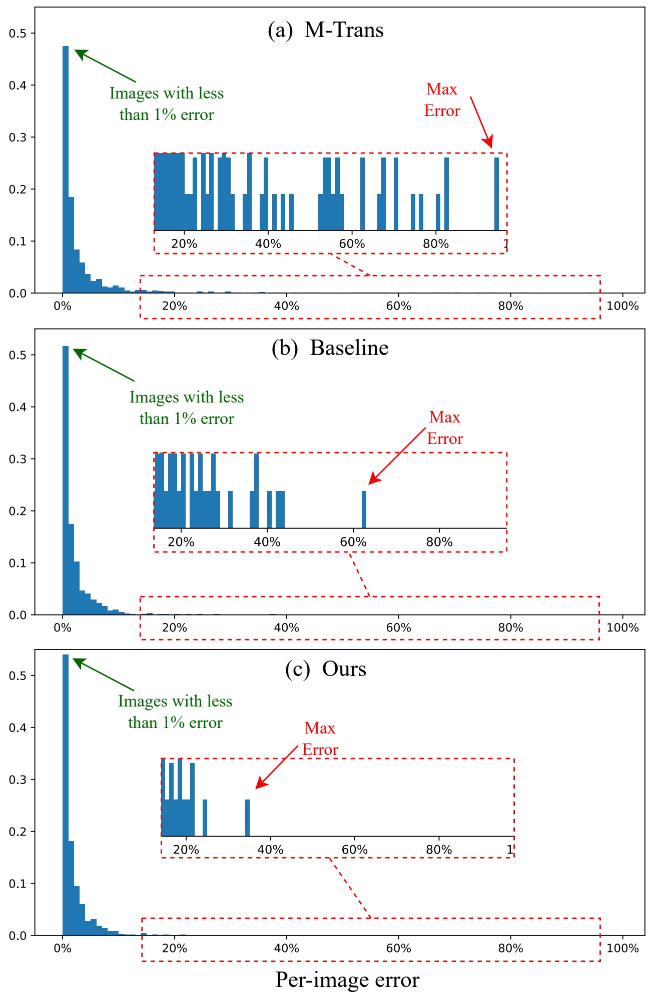


Figure 5.5: Error distribution of (a) the previous best method M-trans, (b) the baseline, and (c) the proposed algorithm. In each histogram, max error and accurate predictions (marked as "image with less than 1% error") are highlighted.

5.3.2. Per-Patient Vitiligo Involvement Percentage (PPVP)

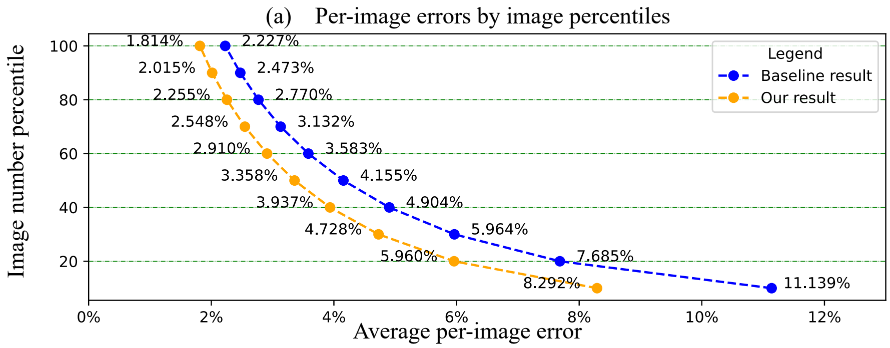
According to [9], our full-body vitiligo images can be used to calculate the Per-Patient Vitiligo Involvement Percentage (PPVP). The error of PPVP is defined as:

$$\sum_{b=0}^{B} w_b * \left|\frac{PV_b}{PS_b} - \frac{GV_b}{GS_b}\right| \tag{5.1}$$

where B is the number of body-parts images of the current image set and w_b is the body surface area percentage defined in the international vitiligo photo standard.[3] PV_b, PS_b,

---

[3] :Face images are discussed separately in the previous paper [39] and not included in this equation.



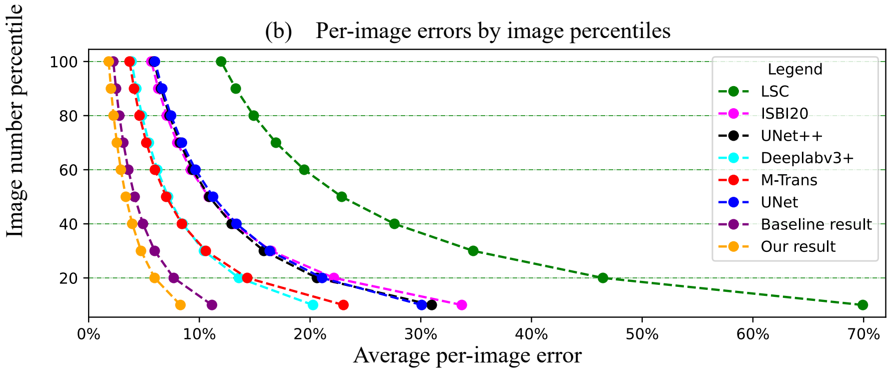


Figure 5.6: (a) and (b) are the comparison of the sorted per-image error by every 10 percentile for the proposed algorithm with the baseline and previous methods, respectively.

GV_b, and GS_b are the numbers of predicted vitiligo pixels, predicted skin pixels, ground-truth vitiligo pixels, and ground-truth skin pixels for image b, respectively. Two sets of images from each of the 5 folds are randomly selected for the dermatologist's naked-eye evaluation. In total, 10 sets of full-body images containing 150 images are used for full-body evaluation and they account for 10% of our total clinical images. The comparison between our result and an evaluation result of a senior dermatologist with 26 years of vitiligo diagnosis experience is recorded. The max and mean error of our result are 3.11% and 1.17%, while the max and mean error of the dermatologist are 3.17% and 1.24%. The comparison shows that the proposed algorithm performs better than the dermatologist.

6. Conclusion

In this paper, the authors pinpoint that patch-wise segmentation is not suitable for clinical usage and therefore, propose the first automatic vitiligo full-body segmentation algorithm for the standard vitiligo images. The first standard full-body vitiligo dataset, consisting of 1740 images, has been established for this research and the results have a good translation to clinical vitiligo scores. An additional non-standard healthy skin dataset, NAFD, has been added to the training set to reduce false positive prediction. The segmentation performance is subsequently improved by a newly proposed tailor-made contrast enhancement module and the integrated successive dilation module. Additionally, a novel confidence score refinement scheme is proposed to rectify the preliminary predictions of hand and foot images by cross-referencing multiple images within the same image set. The proposed algorithm significantly outperforms previous vitiligo segmentation methods and networks in the average of the top 1, 10, and 20 errors and the mean error. The vitiligo clinical score is obtained automatically for the first time with this work. The proposed algorithm provides a slightly better estimation than the naked-eye estimation of an experienced dermatologist, reducing the workload of dermatologists.

References

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, 34(11): 2274–2282, 2012.

[2] A. A. Adegun and S. Viriri. Deep learning-based system for automatic melanoma detection. *IEEE Access*, 8: 7160–7172, 2019.

[3] *BFLdataset github*. `https://github.com/BFLTeam`, note=Accessed:2023-02-23.

[4] Z. Bian, S. Xia, C. Xia, and M. Shao. Weakly supervised vitiligo segmentation in skin image through saliency propagation. *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM),* IEEE. 931–934, 2019.

[5] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. *Proceedings of the IEEE conference on computer vision and pattern recognition,* 7291–7299, 2017.

[6] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. *European Conference on Computer Vision,*

[7] M. Diwakar, A. Shankar, C. Chakraborty, P. Singh, and G. Arunkumar. Multi-modal medical image fusion in nsst domain for internet of medical things. *Multimedia Tools and Applications*, 81(26): 37477–37497, 2022.

[8] M. Everingham, S. A. Eslami, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. The pascal visual object classes challenge: a retrospective. *International journal of computer vision*, 111: 98–136, 2015.

[9] N. van Geel, I. Hamzavi, I. Kohli, A. Wolkerstorfer, H. W. Lim, J. M. Bae, H. Lui, J. E. Harris, A. G. Pandya, S. T. T. Guan, et al. Standardizing serial photography for assessing and monitoring vitiligo: a core set of international recommendations for essential clinical and technical specifications. *Journal of the American Academy of Dermatology*, 2019.

[10] N. van Geel, J. Lommerts, M. Bekkenk, A. Wolkerstorfer, C. A. Prinsen, V. Eleftheriadou, A. Taıeb, M. Picardo, K. Ezzedine, and R. Speeckaert. Development and validation of the vitiligo extent score (ves): an international collaborative initiative. *Journal of Investigative Dermatology*, 136(5): 978–984, 2016.

[11] R. Gu, L. Wang, and L. Zhang. De-net: a deep edge network with boundary information for automatic skin lesion segmentation. *Neurocomputing*, 468: 71–84, 2022.

[12] L. Guo, Y. Yang, H. Ding, H. Zheng, H. Yang, J. Xie, Y. Li, T. Lin, and Y. Ge. A deep learning-based hybrid artificial intelligence model for the detection and severity assessment of vitiligo lesions. *Annals of Translational Medicine*, 10(10), 2022.

[13] D. Gutman, N. C. Codella, E. Celebi, B. Helba, M. Marchetti, N. Mishra, and A. Halpern. Skin lesion analysis toward melanoma detection: a challenge at the international symposium on biomedical imaging (isbi) 2016, hosted by the international skin imaging collaboration (isic). *arXiv preprint arXiv:1605.01397*, 2016.

[14] M. H. Hesamian, W. Jia, X. He, and P. Kennedy. Deep learning techniques for medical image segmentation: achievements and challenges. *Journal of digital imaging*, 32: 582–596, 2019.

[15] N. Ibtehaz and M. S. Rahman. Multiresunet: rethinking the u-net architecture for multimodal biomedical image segmentation. *Neural networks*, 121: 74–87, 2020.

[16] R. Kaur, R. LeAnder, N. Mishra, J. Hagerty, R. Kasmi, R. J. Stanley, M. Celebi, and W. V. Stoecker. Thresholding methods for lesion segmentation of basal cell carcinoma in dermoscopy images. *Skin Research and Technology*, 23(3): 416–428, 2017.

[17] T. Kawakami and T. Hashimoto. Disease severity indexes and treatment evaluation criteria in vitiligo. *Dermatology Research and Practice*, 2011, 2011.

[18] P. Kharazmi, M. I. AlJasser, H. Lui, Z. J. Wang, and T. K. Lee. Automated detection and segmentation of vascular structures of skin lesions seen in dermoscopy, with an application to basal cell carcinoma classification. *IEEE journal of biomedical and health informatics*, 21(6): 1675–1684, 2016.

[19] R. Lencioni, T. de Baere, R. C. Martin, C. W. Nutting, and G. Narayanan. Image-guided ablation of malignant liver tumors: recommendations for clinical validation of novel thermal and non-thermal technologies-a western perspective. *Liver cancer*, 4(4): 208–214, 2015.

[20] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár. Focal loss for dense object detection. *Proceedings of the IEEE international conference on computer vision,* 2980–2988, 2017.

[21] M. Low, V. Huang, and P. Raina. Automating vitiligo skin lesion segmentation using convolutional neural networks. *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI),* IEEE. 1–4, 2020.

[22] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee, et al. Mediapipe: a framework for building perception pipelines. *arXiv preprint arXiv:1906.08172*, 2019.

[23] A. J. Miller and M. C. Mihm Jr. Melanoma. *New England Journal of Medicine*, 355(1): 51–65, 2006.

[24] H. Nugroho, M. H. Ahmad Fadzil, N. Shamsudin, and S. Hussein. Computerised image analysis of vitiligo lesion: evaluation using manually defined lesion areas. *Skin Research and Technology*, 19(1): 72–77, 2013.

[25] A. Nurhudatiana. A computer-aided diagnosis system for vitiligo assessment: a segmentation algorithm. *International Conference on Soft Computing, Intelligence Systems, and Information Technology,* Springer. 323–331, 2015.

[26] *Pytorch Color jittering.* `https://pytorch.org/vision/main/generated/torchvision.transforms.ColorJitter.html`, note=Accessed:2023-02-23.

[27] V. Ravi, H. Narasimhan, C. Chakraborty, and T. D. Pham. Deep learning-based meta-classifier approach for covid-19 classification using ct scan and chest x-ray images. *Multimedia systems*, 28(4): 1401–1415, 2022.

[28] J. Redmon and A. Farhadi. Yolov3: an incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.

[29] S. Roy, C. Coldren, A. Karunamurthy, N. S. Kip, E. W. Klee, S. E. Lincoln, A. Leon, M. Pullambhatla, R. L. Temple-Smolkin, K. V. Voelkerding, et al. Standards and guidelines for validating next-generation sequencing bioinformatics pipelines: a joint recommendation of the association for molecular pathology and the college of american pathologists. *The Journal of Molecular Diagnostics*, 20(1): 4–27, 2018.

[30] M. V. Sherer, D. Lin, S. Elguindi, S. Duke, L.-T. Tan, J. Cacicedo, M. Dahele, and E. F. Gillespie. Metrics to evaluate the performance of auto-segmentation for radiation treatment planning: a critical review. *Radiotherapy and Oncology*, 160: 185–191, 2021.

[31] J. Su, J. Li, Y. Zhang, C. Xia, and Y. Tian. Selectivity or invariance: boundary-aware salient object detection. *Proceedings of the IEEE/CVF international conference on computer vision,* 3799–3808, 2019.

[32] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. *Proceedings of the IEEE conference on computer vision and pattern recognition,* 2818–2826, 2016.

[33] J. Toh, S. Bhoi, V. Tan, S. Chuah, A. Jhingan, A. Kong, and S. Thng. Automated scoring of vitiligo using superpixel-generated computerized digital image analysis of clinical photographs: a novel and consistent way to score vitiligo. *British Journal of Dermatology*, 179(1): 220–221, 2018.

[34] J. M. J. Valanarasu, P. Oza, I. Hacihaliloglu, and V. M. Patel. Medical transformer: gated axial-attention for medical image segmentation. *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021 Proceedings, Part I 24,* Springer. 36–46, 2021.

[35] M. Vardhana, N. Arunkumar, S. Lasrado, E. Abdulhay, and G. Ramirez-Gonzalez. Convolutional neural network for bio-medical image segmentation with hardware acceleration. *Cognitive Systems Research*, 50: 10–14, 2018.

[36] *Vitiligo Images Kaggle.* `https://www.kaggle.com/datasets/shaikhshahid/vitiligo-images`, note=Accessed:2023-02-23.

[37] C. Wong, R. Strange, and J. Lear. Basal cell carcinoma. *Bmj*, 327(7418): 794–798, 2003.

[38] A. Xu, W. Li, P. Guo, D. Yang, H. R. Roth, A. Hatamizadeh, C. Zhao, D. Xu, H. Huang, and Z. Xu. Closing the generalization gap of cross-silo federated medical image segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,* 20866–20875, 2022.

[39] L. Yanling, A. W.-K. Kong, and S. Thng. Segmenting vitiligo on clinical face images using cnn trained on synthetic and internet images. *IEEE Journal of Biomedical and Health Informatics*, 25(8): 3082–3093, 2021.

[40] S. Zhang, L. Song, C. Gao, and N. Sang. Glnet: global local network for weakly supervised action localization. *IEEE Transactions on Multimedia*, 22(10): 2610–2622, 2019.

[41] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang. Unet++: redesigning skip connections to exploit multiscale features in image segmentation. *IEEE transactions on medical imaging*, 39(6): 1856–1867, 2019.

[42] A. Zwanenburg, M. Vallières, M. A. Abdalah, H. J. Aerts, V. Andrearczyk, A. Apte, S. Ashrafinia, S. Bakas, R. J. Beukinga, R. Boellaard, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*, 295(2): 328–338, 2020.