

Unconstrained Thermal Hand Segmentation

Ewelina Bartuzi and Mateusz Trokielewicz

Warsaw University of Technology, Institute of Control and Computation Engineering
Nowowiejska 15/19, 00665 Warsaw, Poland

{e.bartuzi, m.trokielewicz}@elka.pw.edu.pl

Abstract

Robust segmentation of thermal hand images, a trivial task in controlled environments, can be difficult for unconstrained acquisition with a thermal camera, when the temperature of the object differs little from this of a background. This paper implements a method for segmenting hand images collected in the thermal spectrum involving a pre-trained, off-the-shelf data-driven model, fine-tuned with multiple databases of thermal hand images and corresponding, manually annotated ground truth masks. This allows superior performance for difficult samples, compared against conventional methods such as Otsu's thresholding and Gaussian Mixture Modeling. The segmentation accuracy for good quality samples is comparable with these traditional methods, while at the same time being more unconstrained acquisition-resistant. An improvement of up to 20% in accuracy measured by intersection over union is observed for difficult samples, such as hands that were partially colder than the background, as well as those with wristwatches and jewelry. This robust segmentation in various acquisition scenarios allows not only the correct localization of regions of interest for feature extraction, but also for extracting accurate hand geometry information. Together with the results of experiments, involving multiple training and testing scenarios on three different databases, we provide codes, model weights, and ground truth masks to ensure reproducibility and facilitate further research.

1. Introduction

Although multiple features of the human hand are already widely used for personal authentication and identification, they have heretofore been limited to fingerprints [13], palmpoints [26, 11, 20], geometric features [24, 30, 3], finger and hand vein patterns [15, 31, 29].

Perhaps the first to introduce a biometric verification system based on thermal information obtained from human

hand were Czajka and Bulwan [8], who employed heat distribution maps collected with a specialized thermal plate sensor, at that time considered an inexpensive alternative to thermal cameras (2013). With heuristic feature selection for determining the most discriminatory biometric features they were able to achieve an average EER (Equal Error Rate) of 6.67% on a group of 50 individuals. The method is considered a possible candidate for bi-modal recognition or a source of liveness cues for presentation attack detection. In [4], Bartuzi *et al.* presented a biometric system utilizing thermal features of the hand collected with the use of a specialized, high quality thermal infrared camera, albeit in an unconstrained, real-world scenario. Their study explored methods employing binarized statistical image features, Gabor wavelets, and convolutional neural networks, which are made translation-, rotation-, and scale-invariant. Authors also introduced stability maps for determining the most robust portions of the heat distribution, and achieved equal error rates of 0.36% (BSIF), 0.28% (Gabor wavelets), and 0% (CNN) for intra-session comparisons, and EERs ranging from 11% to 30% for inter-session comparisons. This shows that although thermal data obtained from the hand can represent the identity-discriminatory information, it is also subject to large variations in time.

It has also been shown that thermal features can be employed as liveness traits for construction of a presentation attack detection (PAD) method, that can be easily employed in a hand biometrics system and offer perfect accuracy in detecting fake representations [6]. Thermal images are also shown to be able to improve identification rates of a closed-set biometric system operating on visible light hand images. Datasets incorporating multi-spectral hand images, including thermal ones, have also been introduced [4, 5].

Image segmentation involves extracting the regions of interest, containing the representations of a given biometric characteristic, *i.e.*, the hand, while discarding the noise associated with the background and other possible intrusions. While the task is trivial for samples representing a hand against uniform background, yet complicated for other cases, including uncontrolled acquisition, during which it's

impossible to control the background. Also challenging are thermal images, which have recently emerged both as an identity-discriminatory, as well as a liveness-related cue. Here, problems include correctly segmenting hand regions that are colder than the background. Since some higher-level semantic information may be necessary, the use of a feature-learned approach employing deep convolutional neural networks (DCNNs) is justified.

This paper introduces a data-driven image segmentation method utilizing an off-the-shelf DCNN model fine-tuned for hand images collected in the thermal spectrum in an unconstrained acquisition scenario, including contributions:

- sample codes and network weights for thermal hand image segmentation models based on a re-trained off-the-shelf DCNN, fine-tuned with different datasets of thermal images and manually assessed ground truths,
- experiments showing a considerable improvement in hand segmentation accuracy and consistency over conventional methods such as Otsu's thresholding and Gaussian Mixture Modeling for challenging samples,
- datasets of ground truth masks for three publicly available databases used in this work.

2. Related work review

2.1. Hand segmentation methods

This section summarizes the most recent approaches to hand image segmentation that were proposed over the course of the last decade. A good review of older efforts in this field can be found in the work of Bu *et al.* [7]. For visible light images collected in a controlled or semi-controlled manner, *i.e.*, those exhibiting a hand against a uniform background, thresholding methods such as Otsu's algorithm [22] are usually used with good results, cf. [18]. However, when a challenging background is introduced, or a thermal image is used, the task becomes more complicated.

Munoz *et al.* explore hand segmentation utilizing fuzzy multiscale aggregation [21], which works for photographs obtained in visible light using a mobile device, *e.g.*, an iPhone, in a non-controlled environment. A database of images collected from 50 people is used. RGB images were converted into the HSV color space and the hue component was extracted for processing. An accuracy of 94.6% in F1 score is reported for the best approach, in which features from different color spaces are combined together.

Sierra *et al.* extend the experiments described above by proposing a segmentation method based on Gaussian multiscale aggregation applied to hand images coming from a synthetic database representing hands displayed on a variety of backgrounds, such as fabrics, carpet, fur, stone, *etc.* [9]. Employing multiscale gathering of the pixels accordingly with a similarity Gaussian function is said to outperform competing approaches, namely the Lossy Data Compres-

sion and Normalized Cuts, by offering accuracy of 88% to 96% of the F-measure metric, depending on the texture.

Mekyska *et al.* introduce the first database of thermographic hand images collected with an infrared camera TESTO 882-3, with thermal images acquired simultaneously with a visible light camera, (320x240 and 640x480). Thermal image segmentation is discussed as a non-trivial, person-dependent task – difficulties with segmenting hand with colder areas are considered. A segmentation method designed specifically for thermographic data is also introduced in the paper, employing Active shape Models trained with 50 manually labeled grayscale images collected in the thermal spectrum. No quantitative accuracy metrics are given, but the method is reported to produce incorrect results for hands with colder areas.

A pipeline for segmenting visible light hand images with complex backgrounds is introduced by Bu *et al.* in [7], employing two-stage procedure of first building a hand skin color model based on a neural network for coarse segmentation, and then refining the output by detecting hand boundaries via edge detection and voting techniques in different color spaces. The method is reported to achieve sensitivity and specificity of above 96% each.

Barra *et al.* present a hand-based biometrics system, in which hand shape is extracted from the RGB image using the conversion to HSV color space and multi-thresholding of the H plane of the resulting HSV image, followed by refining of the mask using morphological operators. No numeric values assessing the segmentation accuracy are given [3].

A short paper by Ungureanu *et al.* presents a hand segmentation approach employing deep learning methods for grayscale images with various backgrounds [28]. Visible light hand images from the CASIA and HKPU databases, are modified with additional, textural backgrounds, such as grass, wood, or textiles. Two deep neural network architectures are experimented with, namely the SegNet model and a 'U-shaped' model introduced by the authors that is designed with the equal number of parameters as the SegNet. The obtained accuracy is 99.55% and 99.72% of F1 score for SegNet and the U-shaped model, respectively.

In [4], segmentation methods based on Otsu's concept [22], as well as an approach utilizing Gaussian Mixture Models (GMM), are used to approximate the distributions of the hand and the background pixels. Although GMMs allow to obtain much better results, they are still imperfect in cases of especially cool fingers.

2.2. CNNs for semantic segmentation

Deep convolutional neural networks (DCNN) are already well known for their excellent performance in a variety of computer vision tasks, including semantic segmentation (pixel-level prediction), both for discerning objects from

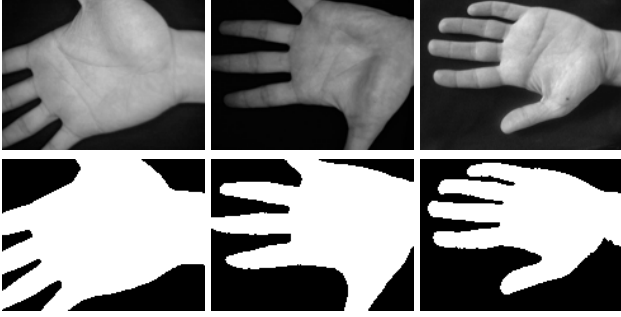


Figure 1: Samples from the *CASIA* database (**top**) and their respective ground truth binary masks (**bottom**).

backgrounds, as well as more complicated tasks, such as labeling traffic imagery for autonomous driving. A good review of current approaches to semantic segmentation can be found in [14].

These methods have already been applied in biometrics-related segmentation tasks, usually outperforming conventional, hand-crafted methods, especially on challenging datasets, such as noisy, visible light iris images [17], or post-mortem iris images [27], and was shown to offer superior performance [19]. Because of the DCNNs’ capability of automatic parameter finding, they can be trained to adapt to almost any task, although their limitation lies in the vast amounts of data needed for the training phase, which are especially difficult to obtain when fine-grained masks are involved, as they require a lot of resources to create. This issue can be partially alleviated by using a pre-trained model, which is later fine-tuned with a smaller dataset of target samples.

3. Experimental data

3.1. Datasets of hand images

We use three publicly available databases of hand images collected in different spectra: *CASIA*-PalmprintV1 [25], *Warsaw*-BioBase-Hand-Thermal-v1 [4], and *Tecnocampus* Hand Image Database (*THID*) [12, 10], further referred to as *CASIA*, *Warsaw*, and *THID*. *CASIA* contains 5501 images of hands of 312 subjects, collected in near-infrared spectrum, displayed on a black, rather uniform background. This is a semi-constrained acquisition, as the subjects were asked to place their hands inside a box equipped with a camera to limit outside illumination, but hand presentation varies significantly, cf. Fig. 1.

Images in the *Warsaw* database were acquired by the FLIR SC645 thermal sensor in a setup without any hand stabilization or positioning in a resolution of 640×480 pixels [4]. The non-uniformity of samples is mainly attributed to the varying temperature of the hands of the subjects, especially in cases with cool fingers, cf Fig. 2.

The following normalization procedure was performed on the samples prior to any experimentation:

$$I_{i,j} = \begin{cases} 0 & \text{if } T_{i,j} < T_{min} \\ 255 & \text{if } T_{i,j} > T_{max} \\ \frac{T_{i,j} - T_{min}}{T_{max} - T_{min}} \times 255, & \text{otherwise} \end{cases} \quad (1)$$

where $I_{i,j}$ denotes the value i, j -th pixel of the normalized image, $T_{i,j}$ corresponds to the value of i, j -th pixel in the thermal image, $T_{min} = 20^\circ\text{C}$ and $T_{max} = 40^\circ\text{C}$ is the experimentally determined temperature range, which can typically be observed in the image T (averaged over all images).

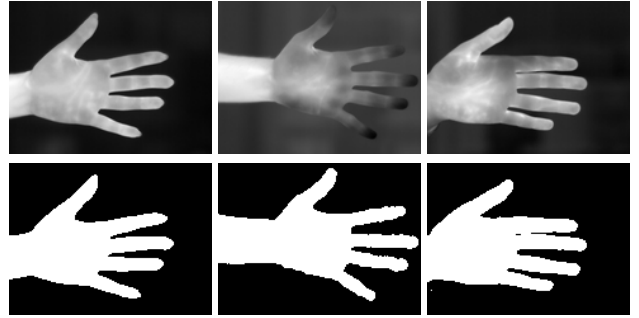


Figure 2: Samples from the *Warsaw* database following normalization (**top**) and their ground truth masks (**bottom**).

Finally, the *THID* database contains hand images collected from 100 subjects in visible light, near-infrared, and thermal spectra. In our experiments, we use only those images that were collected in the thermal range. The same normalization procedure as defined in Equation 1 is used for these samples, Fig. 3.

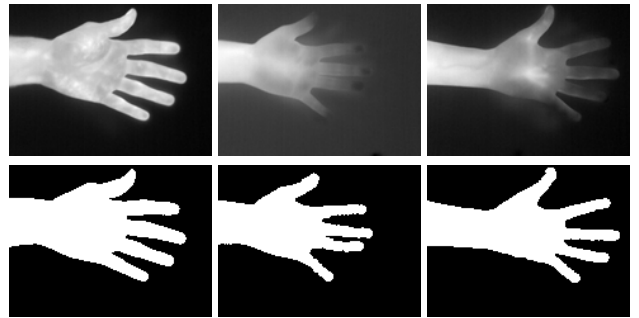


Figure 3: Same as in Fig. 2, but samples from the *THID* database are shown.

3.2. Ground truth binary masks

For each database, we have chosen a subset of samples for which ground truth, binary masks were manually prepared by annotating the hand region in each image. Since

preparing dense labels is a very time-consuming task, this was done for 734 images belonging to 85 classes for the THID database, and 731 images belonging to 70 classes for the Warsaw database. For the CASIA database, masks for all 5501 images were created, since most of them could be automatically obtained by thresholding, and the remaining portion of 'challenging' images was processed manually. Ground truth masks for selected samples coming from each dataset are shown in Figs. 1-3. These masks are available to interested researchers as one of the contributions of this paper, to stimulate further research in this area*

3.3. Image quality classes

All images from the *Warsaw* and the *THID* databases have been assigned a class that defines the overall character and quality of a particular image. These classes are defined and examples are given in Tab. 1.

4. Proposed methodology

4.1. Baseline conventional segmentation methods

For a complete assessment of the accuracy of the segmentation method proposed in this paper, we evaluate our approach against two conventional hand segmentation methods, namely:

- **Otsu's thresholding** and binarization; Otsu's method selects the threshold by maximizing the inter-class (object-background) variance without making any assumptions on the pixel intensity distributions [22];
- **Gaussian Mixture Models (GMM)**, which approximate the distributions of pixels belonging to the hand and those of the background [16];

4.2. Segmentation accuracy metrics

Two well-recognized accuracy metrics are employed for assessing the quality of the proposed segmentation method:

- **Intersection over Union**, a metric typically seen in segmentation tasks:

$$IoU = \frac{prediction \cap ground_truth}{prediction \cup ground_truth}$$

or

$$IoU = \frac{\sum_{i=1}^m \sum_{j=1}^n P_{ij} \wedge G_{ij}}{\sum_{i=1}^m \sum_{j=1}^n P_{ij} \vee G_{ij}}$$

- **E_1 error metric** as used in [23] and [1]:

$$E_1 = \frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n P_{ij} \oplus G_{ij}$$

*instructions on getting access to this data are provided at <http://zbum.ia.pw.edu.pl/EN/node/46>

where P_{ij} and G_{ij} denote the logical values of prediction mask and ground truth mask for the ij -th pixel, respectively, m, n is the image size in pixels, and \oplus denotes XOR (exclusive or) bitwise logical operator.

4.3. DCNN model architecture

For the purpose of this work, we take advantage of the SegNet architecture introduced in [2], which employs a fully convolutional encoder-decoder architecture. The encoder stage employs a VGG-16 model graph, whereas the decoder comprises several sets of convolution and upsampling layers, whose target is to retrieve spatial information from the encoder output, to yield a dense, pixel-wise output map of the same size as the input image. We then fine-tune the off-the-shelf weights of the SegNet model pre-trained on ImageNet with datasets of thermal hand images and their corresponding ground truth masks, cf. Sec. 3.

4.4. Training and evaluation



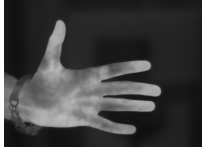




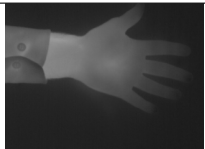

The training and testing experiments evaluate both the **within-dataset** (highlighted in blue) and **cross-dataset** performance of the proposed solution. In the first part of this study, the following five experiments are carried out, using data from two databases, namely *Warsaw* and *THID* databases:

- **training and testing on *Warsaw***
- **training and testing on *THID***
- training on *Warsaw*, testing on *THID*
- training on *THID*, testing on *Warsaw*
- **training and testing on both datasets**

Then, in the second part of the evaluation, the *CASIA* database is included in the training phase, and the network is initially trained on it before further training with *Warsaw*, *THID*, or both. Since *CASIA* contains a much larger number of images with corresponding ground truth labels than the other two datasets (5501 vs ≈ 730 , albeit these are not thermal images, but rather near-infrared ones), the goal is to help the network learn the typical shape of a human hand.

For the training and testing procedure in the **within-dataset** scenarios, 10 subject-disjoint train/test data splits were created by randomly choosing the data from approximately 80% of the subjects for training, and the data from the remaining 20% of the subjects for testing, for each of the experiments. The network is then trained with each train subset independently for each split, and evaluated on the corresponding test subset. All ten splits were made with replacement, making them statistically independent, and allowing for variance analysis of the results. As for the **cross-dataset** evaluations, the training is performed using using all available samples from the training dataset(s), whereas

Table 1: Definitions and examples for quality-based class assignment of the samples from the *Warsaw* and *THID* databases.

Class	Description	Warsaw	THID
I <i>warm</i>	images presenting a palm (and possibly a part of the wrist) against a colder background 33.79% of <i>Warsaw</i> 36.78% of <i>THID</i>		
II <i>warm with intrusions</i>	images similar to those from Class I, but with additional visible clothing and/or jewelry, wristwatches, etc. 33.11% of <i>Warsaw</i> 20.16% of <i>THID</i>		
III <i>cold</i>	images presenting hands with cooler regions, which temperature is similar to this of the background or lower 16.28% of <i>Warsaw</i> 23.98% of <i>THID</i>		
IV <i>cold with intrusions</i>	images similar to those from Class III, but with additional visible clothing and/or jewelry 16.82% of <i>Warsaw</i> 6.40% of <i>THID</i>		
V <i>heat shade</i>	images with heat-shade effect caused by hand movement during image acquisition none in <i>Warsaw</i> 12.68% of <i>THID</i>	—	

the testing employed the same 10 test splits obtained for the corresponding **within-dataset** experiment.

Training took 150 epochs in each trial, with stochastic gradient descent as the optimization method. Momentum of 0.9, learning rate of 0.001 decreased 10-fold after every 50 epochs, batch size of 4, and L2 regularization of 0.0001 were used. The data were shuffled after each epoch.

During testing, a prediction in the form of binary mask is obtained from the network for each of the images. For each predicted mask, Intersection over Union and E_1 error metrics are calculated between the prediction and the ground truth mask, which is available also for test portions of the data. These are then averaged to get the mean IoU and E_1 for each test split. The same procedure is repeated for evaluation of the conventional segmentation methods (cf. Sec. 4.1), which serve as baseline performance indicators, *i.e.*, they are evaluated on the exactly same sets of samples, as the corresponding test split of the DCNN-based method.

5. Experimental results

5.1. Average IoU and E_1 within- and cross-dataset

Fig. 4 shows the segmentation accuracy measured with IoU and averaged over all 10 train/test splits for the best performing model in both the within- and cross-database scenario. The winning within-dataset model was trained with data from all three available databases, namely: *Warsaw*, *THID*, and *CASIA*, and achieves mean

$$IoU_{CNN(within)}^{Warsaw} = 94.66\% \text{ and } IoU_{CNN(within)}^{THID} = 96.71\%$$

when tested on *Warsaw* and *THID* databases, respectively. In comparison, for the *Warsaw* dataset

$$IoU_{Otsu}^{Warsaw} = 83.75\% \text{ and } IoU_{Otsu}^{THID} = 88.15\%$$

$$IoU_{GMM}^{Warsaw} = 83.84\% \text{ and } IoU_{GMM}^{THID} = 91.38\%$$

for Otsu and GMM, respectively. For cross-dataset, the best DCNN models still outperform both conventional approaches, yet by a smaller margin:

$$IoU_{CNN(cross)}^{Warsaw} = 85.51\% \text{ and } IoU_{CNN(cross)}^{THID} = 94.57\%.$$

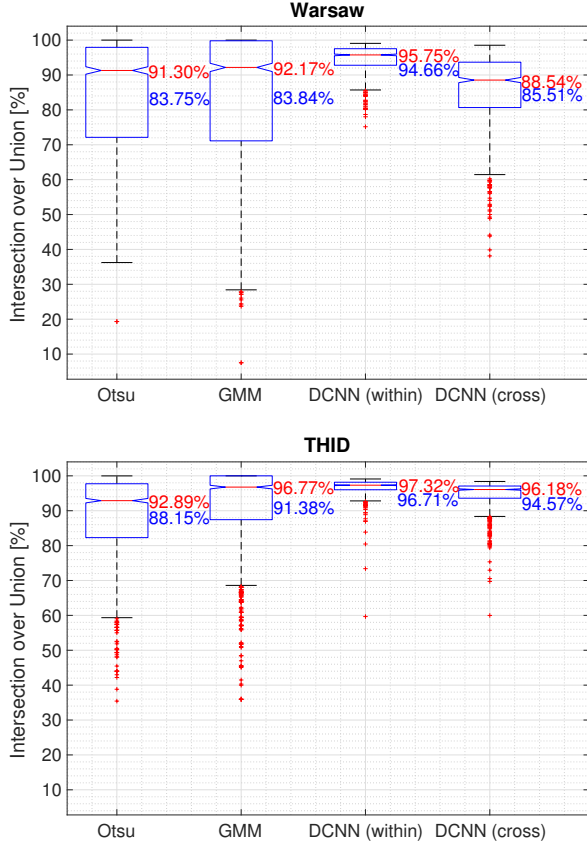


Figure 4: Boxplots representing Intersection over Union in 10 test splits, for the two conventional and the proposed method for both the **within-** and **cross-dataset** evaluation. Medians are shown in red, means in blue.

However, when analyzing the boxplots denoting the performance of each method, one may see that although the averaged gain on the *Warsaw* dataset is small ($< 2\%$) for the cross-dataset experiment, the method gives much more consistent, and thus predictable results, whereas for the traditional methods (Otsu and GMM) the segmentation accuracy is either very high or very low. We inspect this behavior closely in the following section by analyzing segmentation performance in relation to image quality.

5.2. Evaluation in respect to quality of the samples

To get better insight into conditions in which our DCNN model provides a significant advantage over the conventional approaches, we examine the performance of segmentation in respect to different ‘image quality’ classes, as introduced in Sec. 3.3. Table 2 gathers *IoU* scores in respect to the image type, the method, and the database involved. Blue rows denote the within-database evaluation, which we do not consider here, focusing only on the cross-database experiments, since they have much more potential value for the community. In addition to the numeric representation of

the results, we show selected samples from both databases together with their corresponding Otsu, GMM, and DCNN predictions, as well as ground truth masks, see Figs. 5-6.

For *warm* images from class I, Otsu and GMM give the best results. This is an expected behavior, since those samples always exhibit a hand that is well distinguishable from the background. The DCNN-based method is capable of achieving slightly lower performance on the *THID* dataset ($\approx 97\%$ vs $\approx 99\%$ for GMM), and lower (but still exceeding 90%, which can be considered good) for the *Warsaw* dataset. When looking at example predictions, the DCNN’s performance is predictable and coherent, and the network is not making any major mistakes, cf. Figs. 5 and 6, row 1.

As for the samples from class II: *warm with intrusions*, the GMM still offers best accuracy for the *Warsaw* dataset, but the DCNN method outperforms it for the *THID* database, albeit by almost 5%. The rather low accuracy obtained by the DCNN model in this case can be attributed to the fact that it tries to mask out the wrist in addition to the watch band or jewelry, but it does correctly approximate the exact shape of the palm, cf. Fig. 5, row 2.

Class III containing *cold* images of hands partially colder or of temperature similar to the background, is where the advantage of the proposed method becomes evident. The DCNN outperforms conventional methods by a large margin in both the *Warsaw* (82.20% vs less than 68%) and the *THID* databases (92.67% vs 82.44%). The predictions obtained from the proposed approach are shown in Figs. 5 and 6, row 3 (very good result on the *Warsaw* sample, and fairly good on the *THID* sample). In comparison, both conventional methods failed to correctly localize the fingers.

Similarly to class III, class IV also shows the clear advantage of the proposed solution over the Otsu or the GMM algorithms. Figs. 5 and 6, rows 4 show a consistent and accurate prediction given by the DCNN method, and again a failure of the conventional algorithms. *IoU*-wise, the DCNN offers a mean of 72.64% and 88.40% for the *Warsaw* and the *THID* datasets, respectively, whereas the best performing conventional method yielded 63.11% and 67.68%.

Finally, the *THID* database contains some samples with a blurring effect caused by hand movement during acquisition, which we call *heat shade*, cf. Fig. 6, row 5. These samples, however, do not seem challenging to any of the evaluated solutions, with all three methods offering good segmentation accuracy: 93.86% for Otsu, 96.00% for DCNN, and 96.84% for GMM.

In within-database experiments, the DCNN-based model trained on all three datasets (*Warsaw*, *THID*, and *CASIA*) significantly outperforms the conventional ones on samples from almost all classes, except for class I: *warm* images. This approach, however, requires access to the subset of the target database to fine-tune the model.

Table 2: Intersection over Union in five sample quality classes (cf. Sec. 3.3) averaged over 10 train/test splits in each experiment obtained for two conventional methods (cf. Sec. 4.1) and the proposed DCNN-based segmentation. Mean IoU and E_1 are given. Best and worst results for each test database for the cross-database experiments are marked in green and red, respectively. The within-dataset ones are blue, with best performing models bolded and highlighted in dark blue.

	Class I	Class II	Class III	Class IV	Class V	Mean IoU	Mean E_1
Otsu							
Warsaw	97.16%	90.03%	67.87%	62.11%	–	83.75%	4.86%
THID	96.13%	86.12%	77.51%	63.84%	93.86%	88.15%	3.61%
GMM							
Warsaw	98.11%	90.13%	65.43%	63.11%	–	83.84%	4.73
THID	98.76%	88.67%	82.44%	67.68%	96.84%	91.38%	2.62%
CNN-based method:							
Train on Warsaw, test on Warsaw	96.62%	93.08%	91.26%	85.91%	–	92.75%	2.14%
Train on both, test on Warsaw	97.42%	94.08%	92.48%	88.57%	–	93.68%	1.83%
Train on THID, test on Warsaw	84.46%	80.02%	77.19%	64.74%	–	78.51%	6.54%
Train on THID, test on THID	96.55%	93.69%	92.02%	88.51%	95.49%	94.42%	1.68%
Train on both, test on THID	97.07%	95.44%	91.74%	90.83%	95.34%	94.46%	1.67%
Train on Warsaw, test on THID	96.86%	93.64%	92.67%	88.40%	96.00%	94.57%	1.60%
Train on Warsaw + CASIA, test on Warsaw	97.35%	94.48%	92.38%	87.98%	–	93.98%	1.76%
Train on all three, test on Warsaw	97.59%	94.67%	93.58%	90.26%	–	94.66%	1.53%
Train on THID + CASIA, test on Warsaw	90.85%	85.57%	82.20%	72.60%	–	85.51%	4.37%
Train on THID + CASIA, test on THID	97.35%	94.72%	93.37%	90.90%	96.40%	95.48%	1.36%
Train on all three, test on THID	97.83%	95.85%	95.49%	93.66%	97.20%	96.71%	0.98%
Train on Warsaw + CASIA, test on THID	96.92%	92.42%	92.65%	88.23%	95.85%	94.30%	1.70%

6. Conclusions

In this paper we implement and evaluate a DCNN-based segmentation method for segmenting hand images collected in the thermal spectrum, with a re-trained off-the-shelf network. Experiments evaluating the proposed approach show that although our model achieves slightly lower performance than the conventional Otsu and GMM methods for *easy* samples – hand that is easily discernible from the background, it can still be considered as a state-of-the-art solution for segmenting thermal hand images thanks to its excellent predictions given for *difficult* samples, such as those with parts of hands or fingers colder than the background, or images with various intrusions, such as wristwatches or jewelry. This makes the proposed approach valuable not only for thermal-based biometric applications, but also for geometry-based approaches, which rely on accurate segmentation of the entire palm.

Apart from the higher average IoU performance of the proposed solution, it is also more consistent between the samples. Slightly lower accuracy obtained by the model for *easy* samples is mostly related to over-aggressive masking of the wrist region and in most cases should not be a disadvantage for systems relying on the palm region.

This is the first known to us study presenting a data-driven solution for hand localization within a thermal infrared image, which would successfully operate in difficult conditions associated with an unconstrained acquisition

scenario, as well as its evaluation against the conventional methods such as Otsu and GMM. We hope that this method, publicly available with network weights, sample codes, and ground truth masks for subsets of three databases used in this work, will be a valuable addition to the field.

References

- [1] M. Arsalan et al. Deep Learning-Based Iris Segmentation for Iris Recognition in Visible Light Environment. *Symmetry*, 9, 2017.
- [2] V. Badrinarayanan, A. Kendall, and R. Cipolla. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE TPAMI*, 39, 2017.
- [3] S. Barra et al. A hand-based biometric system in visible light for mobile environments. *Information Sciences*, 479:472 – 485, 2019.
- [4] E. Bartuzi, K. Roszczewska, A. Czajka, and A. Pacut. Unconstrained Biometric Recognition Based on Thermal Hand Images. *IEEE IWBF*, 2018.
- [5] E. Bartuzi, K. Roszczewska, M. Trokielewicz, and R. Białobrzeski. Mobibits: Multimodal Mobile Biometric Database. *BIOSIG*, 2018.
- [6] E. Bartuzi and M. Trokielewicz. Thermal Features for Presentation Attack Detection in Hand Biometrics. *BTAS*, 2018.
- [7] W. Bu, K. Wang, X. Wu, X. Cui, and Q. Zhao. Hand segmentation for hand-based biometrics in complex environments. *Journal of Software*, 8:2439–2446, 2013.
- [8] A. Czajka and P. Bulwan. Biometric verification based on hand thermal images. *ICB*, 2013.

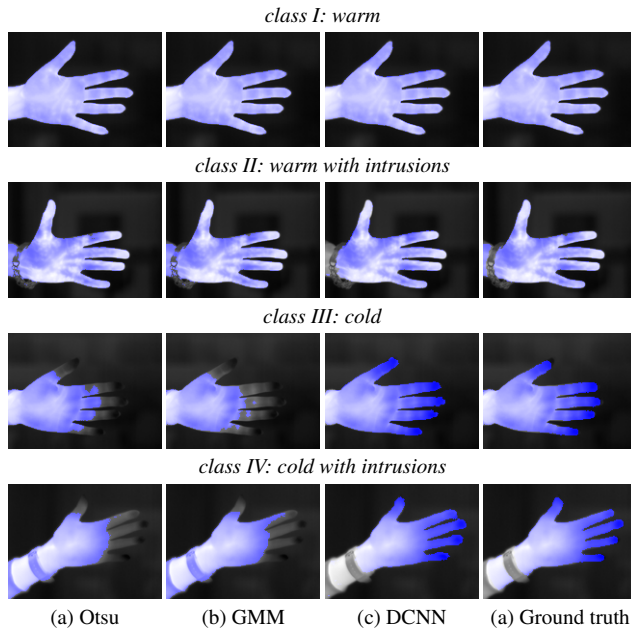


Figure 5: Results of the conventional (a, b) and proposed (c) segmentation for Warsaw samples, with manually annotated ground truth (d). Four image classes as in Sec. 3.3.

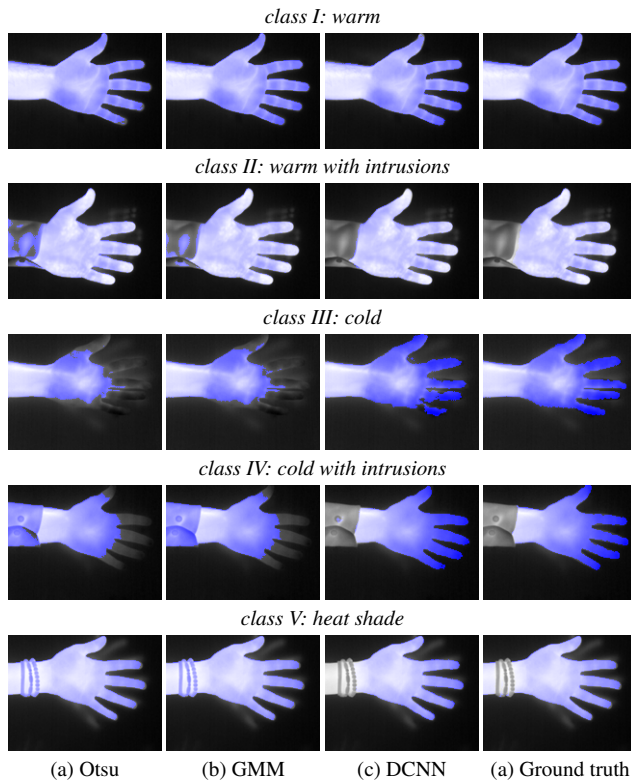


Figure 6: Same as in Fig. 5 but with THID samples.

[9] A. de Santos Sierra, C. Avila, J. Casanova, and G. del Pozo. Gaussian multiscale aggregation applied to segmentation in

hand biometrics. *Sensors*, 11(12):11141–11156, 2011.

[10] M. Faúndez-Zanuy et al. A new hand image database simultaneously acquired in visible, near-infrared and thermal spectrums. *Cognitive Computation*, 6:230–240, 2013.

[11] L. Fei, G. Lu, W. Jia, S. Teng, and D. Zhang. Feature Extraction Methods for Palmprint Recognition: A Survey and Evaluation. *IEEE TSMC*, 2018.

[12] X. Font-Aragones, M. Faundez-Zanuy, and J. Mekyska. Thermal hand image segmentation for biometric recognition. *IEEE MAES*, 28(6):4–14, 2013.

[13] F. Galton. *Finger Prints*. Macmillan and Company, 1892.

[14] A. Garcia-Garcia et al. A Review on Deep Learning Techniques Applied to Semantic Segmentation. <https://arxiv.org/abs/1704.06857v1>, 2017.

[15] J. Hashimoto. Finger Vein Authentication Technology and Its Future. *VLSI*, 2006.

[16] Z.-K. Huang and K.-W. Chau. A New Image Thresholding Method Based on Gaussian Mixture Model. *Applied Mathematics and Computation*, 205(2):899–907, 2008.

[17] E. Jalilian and A. Uhl. Iris Segmentation Using Fully Convolutional Encoder-Decoder Networks. *Deep Learning for Biometrics*, in: *Advances in Computer Vision and Pattern Recognition*, 2017.

[18] W. Jia, R. X. Hu, J. Gui, Y. Zhao, and X. M. Ren. Palmprint recognition across different devices. *Sensors*, 2012.

[19] D. Kerrigan, M. Trokielewicz, A. Czajka, and K. W. Bowyer. Iris recognition with image segmentation employing re-trained off-the-shelf deep neural networks. *ICB 2019*.

[20] A. Kumar. Toward More Accurate Matching of Contactless Palmprint Images Under Less Constrained Environments. *IEEE TIFS*, 14:34–47, 2018.

[21] A. Munoz et al. Hand biometric segmentation by means of fuzzy multiscale aggregation for mobile devices. *ETCHB*, 2010.

[22] N. Otsu. A threshold selection method from gray-level histograms. *IEEE TSMC*, 9(1):62–66, 1979.

[23] H. Proenca and L. A. Alexandre. The NICE.I: Noisy Iris Challenge Evaluation - Part I. *BTAS*, 2007.

[24] L. Stasiak. Support vector machine for hand geometry-based identity verification system. *Proc. SPIE*, 6347, 2006.

[25] Z. Sun, T. Tan, Y. Wang, and S. Z. Li. Ordinal palmprint representation for personal identification. *CVPR 2005*.

[26] K. Tiwari, C. J. Hwang, and P. Gupta. A palmprint based recognition system for smartphone. *FTC*, 2017.

[27] M. Trokielewicz and A. Czajka. Data-driven Segmentation of Post-mortem Iris Images. *IWBF*, 2018.

[28] A.-S. Ungureanu, S. Bazrafkan, and P. Corcoran. Deep Learning for hand segmentation in complex backgrounds. *IEEE ICCE*, 2018.

[29] H. Wan, L. Chen, H. Song, and J. Yang. Dorsal hand vein recognition based on convolutional neural networks. *IEEE BIBM*, 2011.

[30] E. Yoruk, E. Konukoglu, B. Sankur, and J. Darbon. Shape-based hand recognition. *IEEE TIP*, 15, 2006.

[31] Y. Zhou and A. Kumar. Human Identification Using Palm-Vein Images. *IEEE TIFS*, 6, 2011.