# Facial landmark detection on thermal data via fully annotated visible-to-thermal data synthesis

Khawla Mallat, Jean-Luc Dugelay

# Facial landmark detection on thermal data via fully annotated visible-to-thermal data synthesis

Khawla Mallat and Jean-Luc Dugelay
Department of Digital Security
EURECOM
Sophia-Antipolis, France
{mallat,dugelay}@eurecom.fr

## Abstract

*Thermal imaging has substantially evolved, during the recent years, to be established as a complement, or even occasionally as an alternative to conventional visible light imaging, particularly for face analysis applications. Facial landmark detection is a crucial prerequisite for facial image processing. Given the upswing of deep learning based approaches, the performance of facial landmark detection has been significantly improved. However, this uprise is merely limited to visible spectrum based face analysis tasks, as there are only few research works on facial landmark detection in thermal spectrum. This limitation is mainly due to the lack of available thermal face databases provided with full facial landmark annotations. In this paper, we propose to tackle this data shortage by converting existing face databases, designed for facial landmark detection task, from visible to thermal spectrum that will share the same provided facial landmark annotations. Using the synthesized thermal databases along with the facial landmark annotations, two different models are trained using active appearance models and deep alignment network. Evaluating the models trained on synthesized thermal data on real thermal data, we obtained facial landmark detection accuracy of 94.59% when tested on low quality thermal data and 95.63% when tested on high quality thermal data with a detection threshold of 0.15×IOD.*

## 1. Introduction

Facial landmark detection (FLD) consists in locating predefined landmarks, such as eye contours, eye brows, nose, lips. These detectors provide a shape representation of the face that captures transformations due to facial expressions and/or head movement. FLD has drawn a lot of attention recently as it became an essential requirement to perform facial image processing, e.g. face alignment and frontalization [20, 34], 3D face reconstruction [21, 20], emotion recognition [25] and lip reading [5]. Facial image processing itself has evolved to explore different spectra other than visible light given the additional information that they can provide, notably near-infrared [9, 35] and thermal spectrum [14, 15, 27]. Particularly, thermal spectrum has been proved beneficial for many applications such as face recognition in darkness [23, 22], facial medical imaging [29], and emotional state analysis [27]. However, FLD on thermal data has not been extensively explored yet, and to our knowledge there are no public facial landmark detectors available that are designed for thermal spectrum. Thermal imagery provides data with lower spatial resolution and contrast when compared with visible imagery, and it also lacks textural and geometrical information. Therefore, applying the advances of FLD designed for visible data to thermal spectrum may be challenging. Also, the lack of public thermal face databases available with facial landmark annotations prevents thermal spectrum from benefiting from the recent advances in deep learning that have yielded to remarkable improvements in FLD performance, including when tested in-the-wild.

In this paper, we present a novel concept that aims to tackle the lack of annotated data in spectra that are less studied than visible spectrum through inter-spectral conversion, particularly in thermal spectrum for FLD task. This proposed concept will enable broader exploration of thermal image processing. Thereby, we provide thermal face databases with full facial landmark annotation through artificial visible-to-thermal data synthesis using existing visible face databases designed for FLD, notably LFPW [2] and Helen [19] databases. We explore the possibility of training different FLD models on the synthesized thermal face data to be robust when tested on real thermal data. In particular, we used active appearance models [6] and deep alignment

978-1-7281-9186-7/20/$31.00 

network [18] to train our facial landmark detectors.

The remainder of this paper is organised as follows. Section 2 presents the previous work in FLD mainly focused on thermal spectrum. Section 3 describes the databases we selected to synthesize thermal face databases and the employed landmark annotation standard, followed by a presentation of the proposed approach to perform visible-to-thermal face synthesis. Section 4 introduces the two selected approaches that are used for FLD. Section 5 reports the experimental setup and the evaluation protocol followed by results and discussion. Conclusions are presented in Section 6.

## 2. Related work

FLD in visible spectrum has been extensively studied during the few last decades and it had witnessed great progress. Early works, based on classic parameterized approaches, include active appearance models [6] and constrained local models [7]. Later on, FLD approaches based on cascaded shape regression [4, 33] were introduced. Recently, approaches based on deep learning have achieved impressive results, notably Deep Alignment Network [18] and Style Aggregated Network [8]. A thorough survey of existing techniques of FLD on visible images and videos can be found in [32].

Very few works have focused on FLD on thermal data despite the attention that is being drawn to the usage of thermal imagery in face analysis tasks. First attempts were aiming to perform single landmark detection. Tzeng et al. [30] used video frames to detect nostrils through tracking the temperature variation due to respiration. Wang et al. [31] trained a SVM to perform binary classification of the eye region based on Haar-like features. Alkali et al. [1] located the temperature maxima as it is commonly situated in the inner corner of the eyes.

More recent works focused on the face region as a whole and aimed to detect multiple facial landmark points. Kopaczka et al. [14] trained an active appearance model using HOG and SIFT features to perform face tracking in thermal videos. This work has been extended [15] by incorporating the active appearance model into a deep convolutional network to provide it with a prior shape information. These two approaches were trained on a fully annotated thermal face database [16] collected by University of Aachen. This database provides high spatial resolution data at $1024\times768$ pixels, with high contrast and noise equivalent temperature difference (NETD) lower than 30mK, meaning that the sensor with which the data is acquired is able to identify very small differences of temperature as 30mK or lower. These data specifications result in extremely high quality thermal data much higher than the data provided by the currently available thermal databases and the affordable thermal sensors available on the market. The high quality of the training data of the FLD model mentioned above yields to a drastic decrease of landmark detection accuracy when tested on low or medium quality thermal data that is being used nowadays for research and commercial purposes.

## 3. Thermal face database synthesis

In this section, we describe the selected visible face databases provided with landmark annotation that are used in this paper. Then, we detail our method to perform visible-to-thermal data synthesis in order to obtain a synthesized thermal face database with full facial landmark annotations. Finally, we present some samples of the generated thermal faces.

### 3.1. Face Databases with full facial landmark annotation

We present, here, the selected databases and the landmark annotation used in this paper.

**Helen [19]:** Helen database contains 2330 face images collected from Flickr. The database includes a large set of variations including pose, lighting, expression, occlusion, and individual differences. The facial landmarks were annotated manually using Amazon Mechanical Turk after an initialisation performed using STASM algorithm.

**LFPW [2]:** The Labeled Face Parts in-the-wild database contains 1035 images collected from the web (Flickr, Google, Yahoo...). LFPW database covers the same variations as Helen database. The Labeling and facial landmark annotation were performed by three Amazon Mechanical Turk members.

Facial landmark annotations, used in this work for these databases, were obtained from those released in the context of the *300 Faces in-the-Wild Challenge: the first facial landmark localization Challenge [26]*, which attempted to mitigate the mismatched original annotation criterions present in the Helen and LFPW databases, with 194 and 29 selected landmark points, respectively. This mismatch in dimensionality motivated the application of a shared semi-supervised approach to FLD followed by manual correction, resulting in a common, consistent 68 facial points annotation. These annotations, which have been widely used as the *de facto* benchmark for landmark detection, were thus used as reference in the work presented here.

### 3.2. Visible-to-thermal data synthesis

Data synthesis from visible to thermal spectrum was carried out using cascaded refinement networks (CRN) trained using contextual loss, enabling it to be inherently scale and rotation invariant. Choosing CRN as the basic block for our image synthesis model was motivated by the fact that it considers multi-scale information and requires training a limited number of parameters resulting in high resolution

image generation. CRN is a convolutional neural network that consists of inter-connected refinement modules, each module consists of only three layers. The first module considers the lowest resolution space (4×4 in our case). This resolution is duplicated in the successor modules until the last module (128×128 in our case), matching the target image resolution.

During the training phase of our CRN network, we used contextual loss $CX$ that aims to compare regions with similar semantic details while preserving the context of the entire image. Our loss function can be modeled as a combination of two losses: style loss and content loss, as defined by Gatys et al. [10]. The style loss is computed between the generated thermal image and the ground truth thermal image. Minimizing the style loss yields to generate artificial images with the same properties as the target thermal image. The content loss is computed between the input visible image and the generated thermal image. The content loss aims at preserving details as facial attributes, while tolerating some local deformations that are required to perform the visible to thermal style conversion. Both losses are computed at embedding level, extracted using VGG19 [28]. The total loss can be formulated as follow:

$$\mathcal{L}(G, I_{Vis}, I_{Th}) = \alpha_1 \mathcal{L}_{CX}(\Phi_{l_s}(G(I_{Vis})), \Phi_{l_s}(I_{Th})) + \alpha_2 \mathcal{L}_{CX}(\Phi_{l_c}(G(I_{Vis})), \Phi_{l_c}(I_{Vis})) \quad (1)$$

Where $I_{Vis}$, $I_{Th}$ and $G$ denote the input visible image, the ground truth thermal image and the generator (i.e. visible to thermal synthesis model), respectively. $\Phi_{l_c}$ and $\Phi_{l_s}$ refer to the VGG19 embeddings extracted at content layers level and style layers level, respectively. $\alpha_1$ and $\alpha_2$ were adjusted using Grid Search.

To train the visible-to-thermal data synthesis model, we used the VIS-TH database [24] that contains paired face images acquired simultaneously in visible and thermal spectrum collected from 50 subjects. This database provides images of 160×120 spatial resolution and NETD<100mK. VIS-TH database contains 21 variations including expression, pose, occlusion and illumination variations. During training, we excluded one variation as it was acquired in total darkness, which yields to 1000 pairs of face images. The training was run for 40 epochs with a learning rate of 1e-4.

To obtain the synthesized databases from visible to thermal, we fed the images of HELEN and LFPW databases to our trained model, that returns the thermal version of the input image. We illustrate, in figure 1, some samples of the synthesized thermal face images and their original counterpart. We note that the synthesized thermal images present a realistic pattern of thermal signature. Some details, such as hair, eye brows and teeth, are converted into high pixel values reflecting regions with lower temperature compared to the face region. In addition, nose region is generated slightly darker as the nose is usually colder than the rest of the face because it is mainly composed of cartilage tissue. Also, eyes contours are generated lighter than the rest of the face, which reflects realistic thermal signature as the high temperatures are situated around the eye region. The synthesized images also present some artifacts as we can observe, in few samples, dark patterns at arbitrary regions of the face.

## 4. Facial landmark detection

In this section, we describe the two selected methods of FLD that will be trained on the synthesized thermal face databases.

### 4.1. Active appearance model

The first approach, used in this work, is based on Active Appearance Model [6] as it is the baseline approach for landmark detection. Active appearance models (AAM) were introduced by Cootes et al. [6] for facial image processing. AAM is a statistical appearance method aiming to model the shape of the face and its appearance as probabilistic distributions that can be generalized nearly to any face. To train the FLD model, AAM requires a set of face images with annotation points defining the facial landmarks. In the training phase, Procrustes analysis is applied to align the set of landmarks and the statistical shape and appearance model variations are extracted using principal component analysis. Unseen faces can be represented by a linear combination of the mean shape and the appearance from the training data with weighted shape and appearance vector.

As to faithfully replicate the AAM approach used to train the FLD model provided by Aachen University [14], we have trained a dense HOG feature-based AAM model fitted using the Inverse-Compositional algorithm.

### 4.2. Deep alignment network

The second selected approach is deep alignment network (DAN) [18] as it is the state-of-the-art in FLD for visible images and it has been evaluated on thermal data in [15]. DAN is based on multi stage neural network that performs an iterative process of refinement of landmark positions. Each stage of the DAN network is a feedforward neural network that provides a prediction of the refined facial landmark location. Each stage is trained until the validation error stabilises. We have used a two stage DAN, between the two stages a similarity transform is applied to re-align the image to the average face shape. A learning rate of 1e-3 is used with Adam optimizer on mini batch sizes of 64.

## 5. Experimental setup and results

In this section, we present firstly our two baseline FLD models. Then, we detail our experimental setup. Finally, we

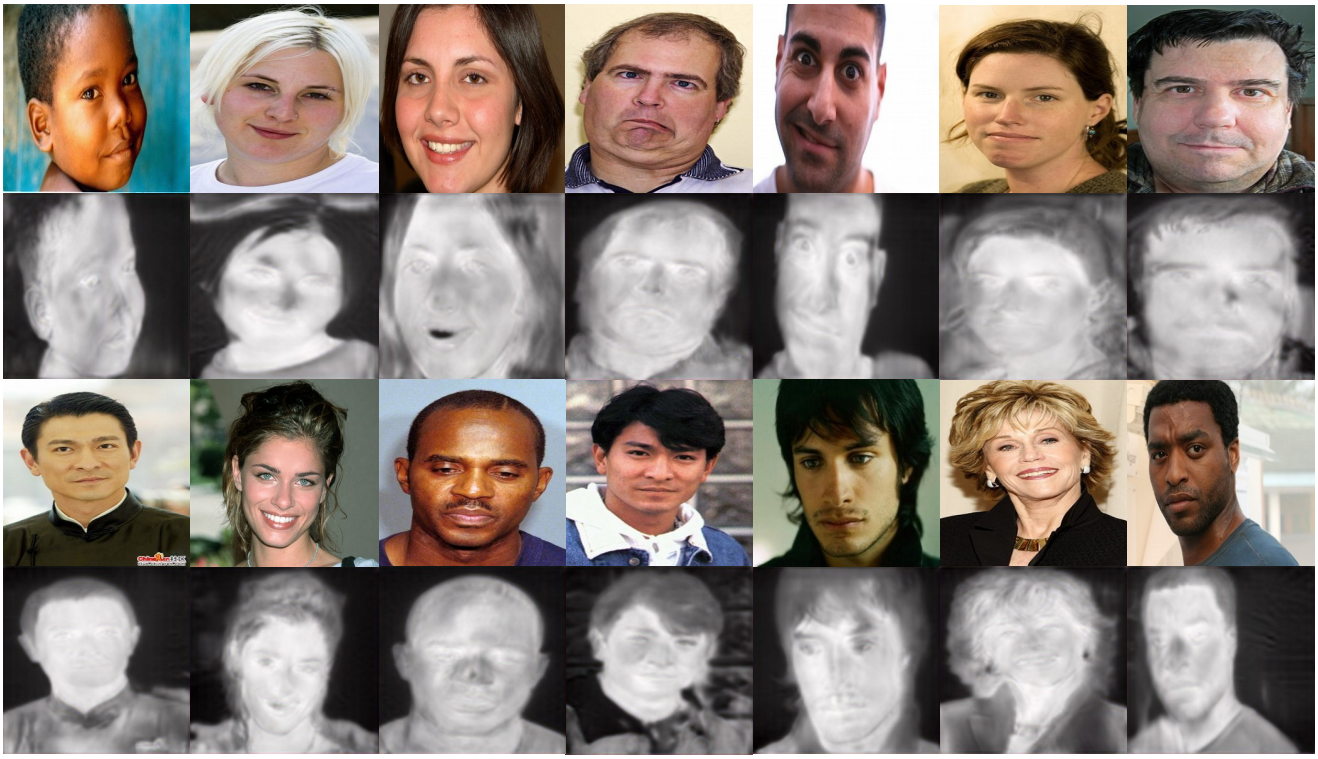


Figure 1: Samples of synthesized thermal images from HELEN and LFPW databases.

introduce our evaluation protocol followed by the reported results.

### 5.1. Baseline models

We consider as baseline models the facial landmark detectors, described in section 4, trained on high quality database provided by Kopaczka et al. [14] from University of Aachen. We will refer, in this paper, to active appearance model and deep alignment network, both trained on Aachen database, as *'AAM-Aachen'* and *'DAN-Aachen'*, respectively. The Aachen database includes high resolution thermal face images that are manually annotated [16]. Video sequences were acquired using a thermal camera with a NETD$<$30mK and spatial resolution of 1024$\times$768 pixels. 695 frames were extracted and manually annotated with 68 point landmarks. To train the AAM model described in section 4.1, the face images were mirrored and 1272 images were selected for the training phase, as described in [14].

### 5.2. Experimental setup

The two selected approaches for FLD, described in section 4, are trained on the synthesized thermal face databases Helen and LFPW separately. We refer to AAM models trained on the synthesized thermal data from Helen and LFPW as *'AAM-Helen'* and *'AAM-LFPW'* and to DAN models as *'DAN-Helen'* and *'DAN-LFPW'*, respectively.

Following the protocol defined in the context of *300 Faces in-the-Wild Challenge: the first facial landmark localization Challenge [26]*, we have used 2000 face images of Helen database and their corresponding facial landmark annotation files for training. Whereas for LFPW database, we have used 811 face images for training our models.

### 5.3. Evaluation protocol and results

The evaluation of FLD performance is assessed by comparing the estimated landmark coordinates to the ground truth. The normalized root mean square error (NRMSE), is computed, point-to-point, to assess the average localization error. NRMSE is considered as a standard metric to evaluate FLD performance [11] and it consists of the euclidean distance between the predicted landmarks and the ground truth landmarks normalized by a predefined distance. Conventionally, the normalization is performed with regards to the Inter-Ocular Distance (IOD), as stated in [11], which is the distance between the two eye outer corners. The normalization process is essential to obtain performance measurement independent of the face size or the image resolution.

The NRMSE, referred to as $E$, is obtained as follow:

$$E_k = \frac{\sqrt{((x,y)_k - (\bar{x},\bar{y})_k)^2}}{d_{norm}} \qquad (2)$$

where $(x,y)_k$ denote the ground truth coordinates and

$(\bar{x}, \bar{y})_k$ the estimated coordinates of the $k^{th}$ landmark point. $d_{norm}$ indicate the normalization distance.

The FLD performances can also be expressed in terms of detection rate as follow:

$$D = \frac{\sum_{k=1}^{K} \sum_{i=1}^{N} [\delta : E_k^i \leq threshold]}{N \times K} \tag{3}$$

$$where\ \delta = \begin{cases} 1 & \text{if } E_k^i \leq threshold \\ 0 & \text{otherwise} \end{cases}$$

where $K$ denotes the total number of the facial landmarks, and $N$ the number of test images. The threshold indicates the NRMSE value under which a landmark point is considered correctly localized.

#### 5.3.1 Evaluation on low quality thermal face data

To evaluate the FLD model on low quality thermal data, CSMAD database [3] is chosen since it provides aligned images in visible and thermal spectrum acquired simultaneously. CSMAD database provides thermal images of spatial resolution of 320×240 and NETD<70mK. This database is designed for face presentation attack, however, it is possible to select, for our evaluation, only the bona fide samples resulting in 423 images. The choice of this database is motivated by the fact that this database can simplify the annotation of the thermal images. The annotation process was performed automatically using DLIB [13] facial landmark detector on the visible set of the database and then corrected manually. Since the visible and thermal sets are aligned, the landmarks detected on the visible set are considered as the ground truth landmark points for the thermal set.

Given that this database also provides samples in visible spectrum, we trained the FLD approaches on the original visible face databases Helen and LFPW. FLD performance on the original visible database will be considered as a reference. The comparison of the performance obtained using thermal based model with the visible based model will quantify the discrepancy between the two spectra in terms of FLD.

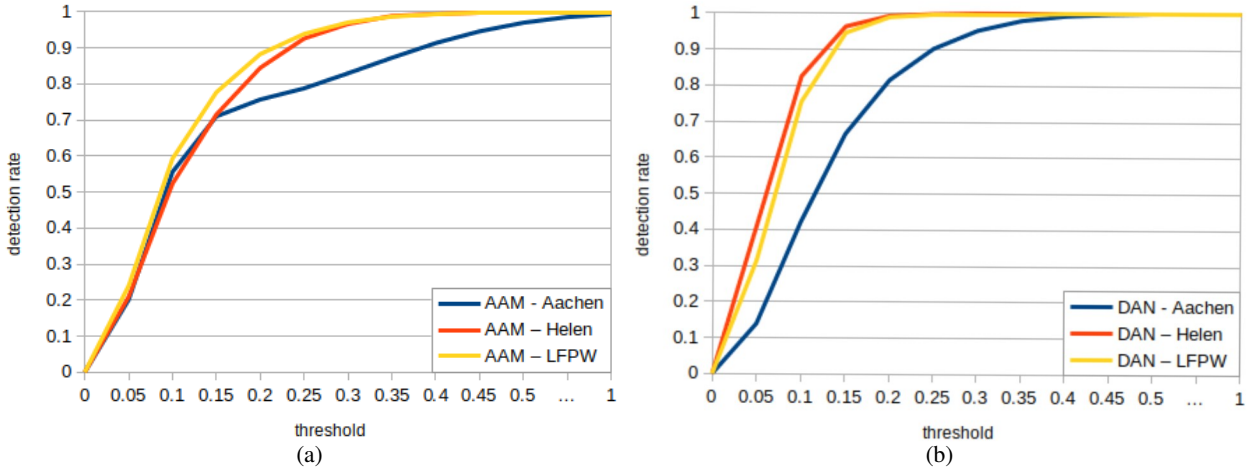
Results, in table 1, show the average and the standard deviation of the localization error in terms of NRMSE obtained by evaluating the different FLD models on the CSMAD database. The first column of the table corresponds to AAM approach trained on different databases: where *'TH'*, *'SynTH'* and *'VIS'* refer to thermal data, synthesized thermal data and visible data, respectively. The second column reports the same results for DAN approach. The localization errors reported by the FLD models trained and tested on thermal face data is relatively higher than the errors reported by the model trained and tested on the original visible images. This is mainly due to the conversion of the face images from highly informative domain, the visible spectrum, to low informative domain as the thermal spectrum, resulting in loss of information relevant for accurate FLD. We also observe that the detection models trained on synthesized thermal data exhibit considerably lower errors than the models trained on Aachen database, which demonstrates the efficiency of our proposed solution.

| | **AAM** | **DAN** |
|---|---|---|
| **Aachen (TH)** | 0.14349 (±0.105) | 0.14595 (±0.052) |
| **LFPW (SynTH)** | 0.11779 (±0.062) | 0.08265 (±0.026) |
| **Helen (SynTH)** | 0.13200 (±0.057) | 0.07309 (±0.022) |
| **LFPW (VIS)** | 0.04020 (±0.015) | 0.04299 (±0.012) |
| **Helen (VIS)** | 0.04568 (±0.031) | 0.03146 (±0.011) |

Table 1: Average NRMSE (± standard deviation) reported on CSMAD database.

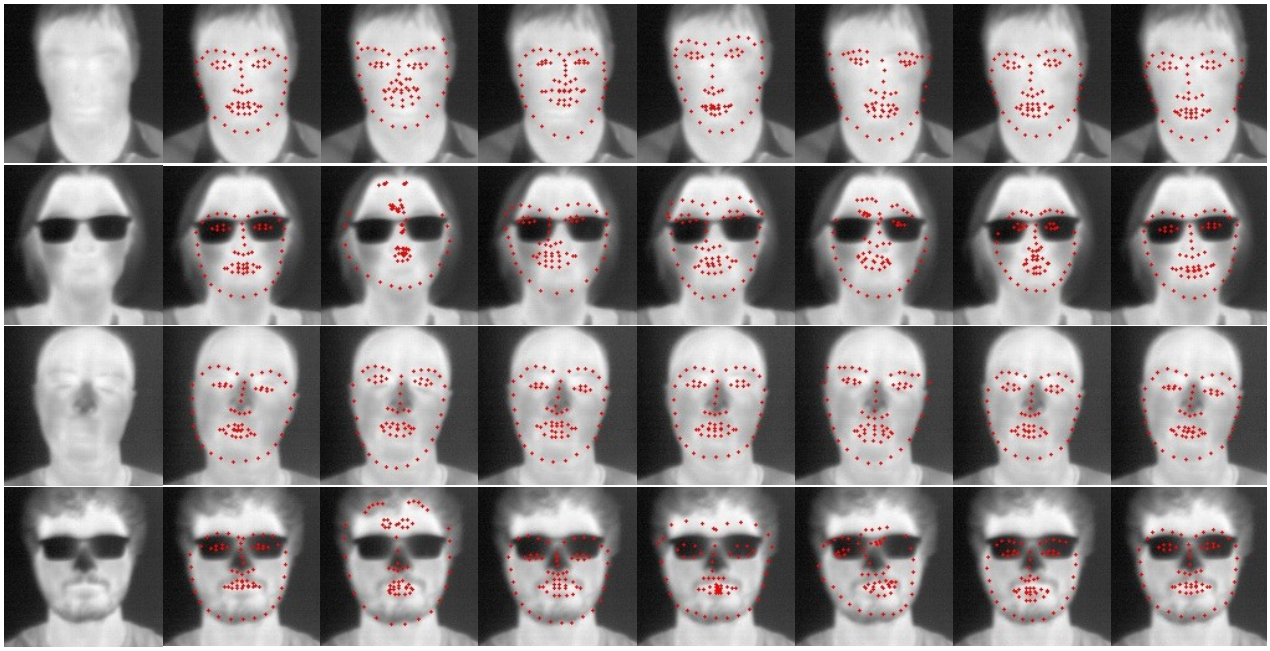
The plots, presented in Figure 2, illustrate the detection rate that corresponds to a defined threshold value for FLD models trained on different databases. We swiped the detection threshold from 0.0 to 1.0 with a step of 0.05. We observe that the two facial landmark detectors trained on Aachen database, represented by the blue curve, yield to significantly low detection rates compared to the detectors trained on the synthesized thermal data. This can be justified by the fact that Aachen models have been trained on very high resolution, high contrast images captured with very high thermal sensitivity. These images are very different from the images provided by the existing thermal face databases, as it is the case for CSMAD database. In addition, the detection rates obtained using DAN approach are considerably higher than the detection rates obtained using AAM. This confirms the efficacy of deep learning solutions in FLD task.

Additional qualitative results, presented in figure 3, depict the performance of each model of FLD on thermal face images with some facial variations. We note that the facial landmark detectors trained on Aachen database [14], shown in column (c) and (f), fail to accurately localize most of the facial traits even under the least challenging variation. However, all the four models trained on the synthesized thermal data provide more accurate landmark localization. Furthermore, we observe that deep learning based detectors (columns (f), (g) and (h)) yield to a more meticulous facial landmark localization compared to statistical modelling based detector. Besides, deep learning models seem to be very robust against challenging facial variation such as occlusion by glasses (lines 2 and 4), as they managed to predict the facial landmark coordinates that are closer to the ground truth whereas the AAM based models tend to fail once it is tested on challenging face variations.




(a) (b)

Figure 2: Detection rate variation of facial landmark detection models evaluated on CSMAD database: (a) Active Appearance Model (b) Deep Alignment Network.



(a) (b) (c) (d) (e) (f) (g) (h)

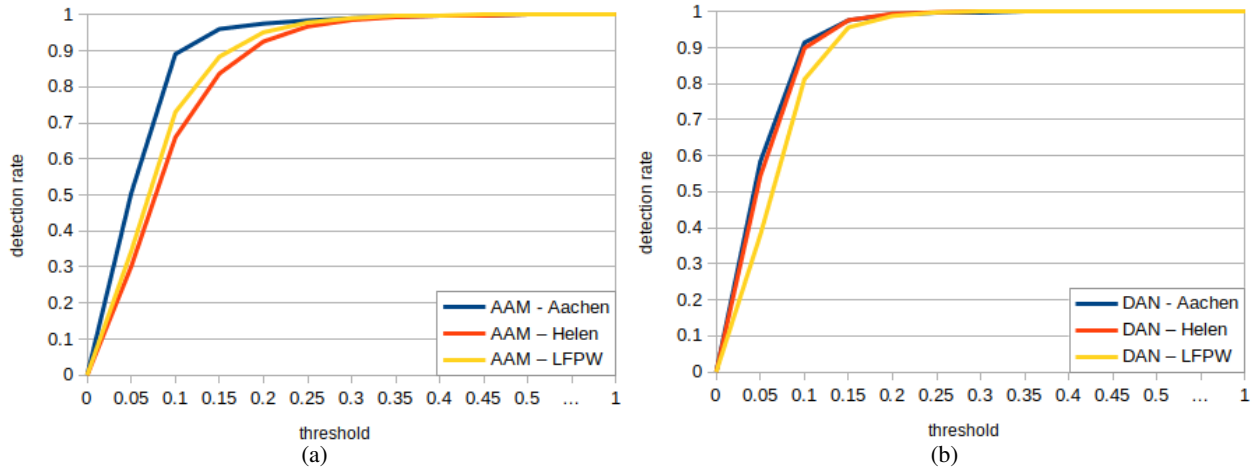
Figure 3: Qualitative results of the different facial landmark detection models on samples of CSMAD database.**(a)**: thermal reference, **(b)**: ground truth, **(c)**:AAM-Aachen, **(d)**: AAM-LFPW, **(e)**: AAM-Helen, **(f)**: DAN-Aachen, **(g)** DAN-LFPW, **(g)**: DAN-Helen.

#### 5.3.2 Evaluation on high quality thermal face data

For fair comparison, the FLD models are also evaluated on high quality thermal data. Aachen database [16] was extended to include thermal face images depicting facial expression variations providing 68 points landmark annotation as well. The expression variation subset of Aachen database is used for our evaluation. We recall that the data provided by Aachen database is characterized by spatial resolution of $1024 \times 768$ pixels and NETD$<$30mK.

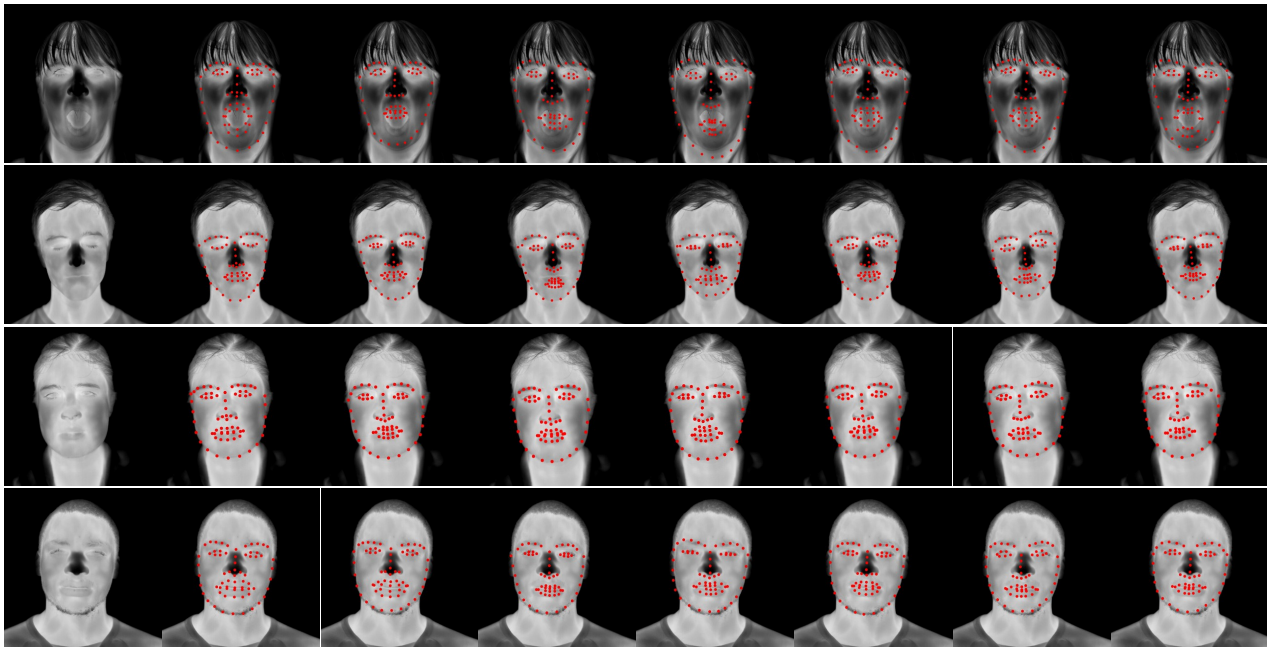
Table 2 presents the average and the standard deviation of the localization error of different FLD models when tested on the expression subset of Aachen database. The detection models trained on Aachen database report lower, but with slight difference, localization errors than the detection models trained on synthesized thermal data. These results are somehow expected as the detection models trained on Aachen database are evaluated on data of same thermal quality acquired with the same thermal sensor.

Detection rates of the different FLD models are illustrated in figure 4. For AAM approach, the detection rate reported by the model trained on Aachen data is considerably




(a) (b)

Figure 4: Detection rate variation of facial landmark detection models evaluated on the expression subset of Aachen database: (a) Active Appearance Model (b) Deep Alignment Network.



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 5: Qualitative results of the different facial landmark detection models on samples of the expression subset of Aachen database. **(a)**: thermal reference, **(b)**: ground truth , **(c)**:AAM-Aachen, **(d)**: AAM-LFPW, **(e)**: AAM-Helen, **(f)**: DAN-Aachen, **(g)** DAN-LFPW, **(g)**: DAN-Helen.

| | **AAM** | **DAN** |
|---|---|---|
| **Aachen (TH)** | 0.07267 (±0.031) | 0.06061 (±0.020) |
| **LFPW (SynTH)** | 0.09534 (±0.034) | 0.07827 (±0.015) |
| **Helen (SynTH)** | 0.10700 (±0.039) | 0.06409 (±0.014) |

Table 2: Average NRMSE (± standard deviation) reported on the expression subset of Aachen database.

higher compared to the models trained on synthesized thermal data. However for DAN, we notice that the curve corresponding to the model trained on Aachen database overlaps with the curve obtained using the model trained on synthesized thermal data from Helen, attesting that the two models perform similarly.

Figure 5 presents some samples of the expression subset of Aachen database portraying the performance of each FLD model. Overall, FLD was less challenging when applied on high quality than on low quality thermal data, as revealed when we compare figure 3 and figure 5. For AAM approach, facial landmark detectors trained on synthesized data perform slightly poorer than the detector trained on

Aachen database. Nevertheless, when using DAN, the three different facial landmark detectors achieve similar performances as they all succeeded to meticulously locate the facial landmarks. For some face variations, we can observe that the model trained on synthesize thermal Helen database (column (h)) detected adequately some challenging landmarks, as the bottom lip (row 1) and closed eyes (row 2), whereas the facial landmark detector trained on Aachen did not manage to correctly predict the localization of these landmarks (column (f)).

#### 5.3.3 Qualitative evaluation on thermal samples of different quality

Given that there are no public thermal face databases, other than Aachen's [16], provided with full facial landmark annotation, further quantitative performance assessment cannot be performed on more data. Therefore, some qualitative results are illustrated in figure 6 to demonstrate that the facial landmark detector trained on synthesized thermal data can operate accurately on thermal data of different quality. Results obtained using the DAN approach trained on Aachen database *'DAN-Aachen'* are shown in row 2 of the figure 6. We have presented, in row 3, results obtained using the DAN model trained on the synthesized thermal data from Helen database *'DAN-Helen'*, as it is the best performing model.

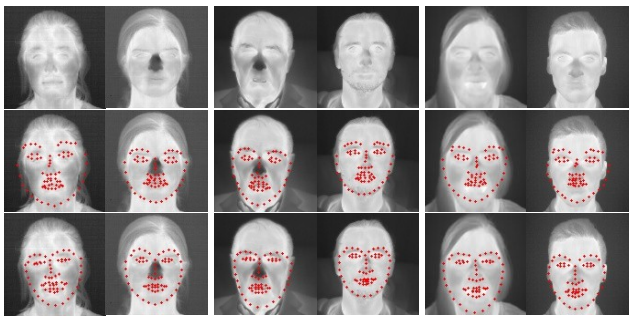



Figure 6: Qualitative results of facial landmark detection on samples of different thermal face databases, using DAN-Aachen in row 2 and DAN-Helen in row 3. **(a)**: UND-X1 database [12], **(b)**: thermal database of Military University of Technology in Warsaw [17] **(c)**: samples collected in our lab.

The presented samples are randomly selected from 3 different databases: UND-X1 database [12] of spatial resolution of 312×239 pixels and NETD<100mK, thermal face database provided by the Military University of Technology in Warsaw [17] of spatial resolution of 640×480 and NETD<50mK, and some samples we acquired in our lab using a thermal sensor of spatial resolution of 620×512 and NETD<50mK. We can observe that for all the samples presented, the model trained on the synthesized thermal data *'DAN-Helen'* has succeeded to correctly localize the facial landmarks, outperforming the model trained on Aachen database *'DAN-Aachen'*.

Given all the results and observations presented above, one may conclude that our proposed concept has yielded to obtain a facial landmark detector that can be suitable to a wide range of thermal data quality.

## 6. Conclusion

In this paper, we addressed the lack of public thermal face databases provided with full annotation for face analysis applications. We introduced an unexplored concept consisting of converting data from one domain to another to tackle this shortage of annotated data. Particularly, we proposed to synthesize artificially a thermal face database with full landmark annotation by converting existing face databases in visible spectrum that have been designed for facial landmark detection task to thermal spectrum. Two different facial landmark approaches were trained on the synthesized thermal face data and tested on low quality and then on high quality thermal data, proving the robustness of the trained models. Our approach was evaluated and compared with two facial landmark detection baseline models provided by Kopazcka et al. [14, 15]. These models were trained on high quality thermal data that yielded to a considerable decrease in performance when tested on thermal face databases that are publicly available. Conclusively, the facial landmark detection models trained on synthesized thermal data had significantly outperformed the baseline models trained on Aachen database when evaluated on lower quality thermal data. Whereas, when tested on high quality thermal data, our proposed models perform similarly to the baseline models that is more adapted for thermal images of such quality.

The best performing model trained on the synthesized thermal face data has achieved an average localization error NRMSE of 0.07 and 94.59% of detection rate at threshold value of 0.15×IOD when evaluated on low quality thermal data. This facial landmark detection model will be shortly made available, as facial landmark detection is an essential step for many face analysis tasks and that to our knowledge there are no public facial landmark detection tools that are available for thermal spectrum. Inter-spectral data synthesis is also reproducible to tackle any lack of available data for tasks that requires extensive annotation.

## 7. Acknowledgment

Our research activities in paired visible and thermal imagery are partly supported through funding from FR FUI COOPOL and ANR-19-FLJO-0004 OKLOS.

## References


[1] A. H. Alkali, R. Saatchi, H. Elphick, and D. Burke. Eyes' corners detection in infrared images for real-time noncontact respiration rate monitoring. In *2014 World Congress on Computer Applications and Information Systems (WCCAIS)*, pages 1–5. IEEE, 2014.

[2] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar. Localizing parts of faces using a consensus of exemplars. volume 35, pages 2930–2940. IEEE, 2013.

[3] S. Bhattacharjee, A. Mohammadi, and S. Marcel. Spoofing deep face recognition with custom silicone masks. In *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–7. IEEE, 2018.

[4] X. Cao, Y. Wei, F. Wen, and J. Sun. Face alignment by explicit shape regression. *Int. Journal of Computer Vision*, 2014.

[5] J. S. Chung, A. Senior, O. Vinyals, and A. Zisserman. Lip reading sentences in the wild. pages 3444–3453, 2017.

[6] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. volume 23, pages 681–685. IEEE, 2001.

[7] D. Cristinacce and T. F. Cootes. Feature detection and tracking with constrained local models. In *Bmvc*, volume 1, page 3. Citeseer, 2006.

[8] X. Dong, Y. Yan, W. Ouyang, and Y. Yang. Style aggregated network for facial landmark detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 379–388, 2018.

[9] S. Farokhi, J. Flusser, and U. U. Sheikh]. Near infrared face recognition: A literature survey. *Computer Science Review*, 21:1 – 17, 2016.

[10] L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2414–2423, 2016.

[11] B. Johnston and P. de Chazal. A review of image-based automatic facial landmark identification techniques. volume 2018, page 86. Springer, 2018.

[12] X. C. P. J. F. Kevin and W. Bowyer. Visible-light and infrared face recognition. In *Workshop on Multimodal User Authentication*, page 48. Citeseer, 2003.

[13] D. E. King. Dlib toolkit. https://github.com/davisking/dlib.

[14] M. Kopaczka, K. Acar, and D. Merhof. Robust facial landmark detection and face tracking in thermal infrared images using active appearance models. In *VISIGRAPP (4: VISAPP)*, pages 150–158, 2016.

[15] M. Kopaczka, L. Breuer, J. Schock, and D. Merhof. A modular system for detection, tracking and analysis of human faces in thermal infrared recordings. *Sensors*, 19(19):4135, 2019.

[16] M. Kopaczka, R. Kolk, and D. Merhof. A fully annotated thermal face database and its application for thermal facial expression recognition. In *2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pages 1–6. IEEE, 2018.

[17] M. Kowalski and A. Grudzień. High-resolution thermal face dataset for face and expression recognition. *Metrology and Measurement Systems*, 25(2), 2018.

[18] M. Kowalski, J. Naruniec, and T. Trzcinski. Deep alignment network: A convolutional neural network for robust face alignment. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 88–97, 2017.

[19] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang. Interactive facial feature localization. In *European conference on computer vision*, pages 679–692. Springer, 2012.

[20] F. Liu, Q. Zhao, D. Zeng, et al. Joint face alignment and 3d face reconstruction with application to face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 2018.

[21] P. Liu, Y. Yu, Y. Zhou, and S. Du. Single view 3d face reconstruction with landmark updating. In *2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, pages 403–408. IEEE, 2019.

[22] K. Mallat, N. Damer, F. Boutros, and J.-L. Dugelay. Robust face authentication based on dynamic quality-weighted comparison of visible and thermal-to-visible images to visible enrollments. In *2019 22th International Conference on Information Fusion (FUSION)*, pages 1–8. IEEE, 2019.

[23] K. Mallat, N. Damer, F. Boutros, A. Kuijper, and J.-L. Dugelay. In *2019 International Conference on Biometrics (ICB)*, pages 1–8. IEEE, 2019.

[24] K. Mallat and J.-L. Dugelay. A benchmark database of visible and thermal paired face images across multiple variations. In *2018 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–5. IEEE, 2018.

[25] M. I. N. P. Munasinghe. Facial expression recognition using facial landmarks and random forest classifier. In *2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS)*, pages 423–427, 2018.

[26] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic. 300 faces in-the-wild challenge: The first facial landmark localization challenge. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 397–403, 2013.

[27] P. Shen, S. Wang, and Z. Liu. Facial expression recognition from infrared thermal videos. In S. Lee, H. Cho, K.-J. Yoon, and J. Lee, editors, *Intelligent Autonomous Systems 12*, pages 323–333, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[28] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, 2014.

[29] S. Sonkusare, D. Ahmedt-Aristizabal, M. J. Aburn, V. T. Nguyen, T. Pang, S. Frydman, S. Denman, C. Fookes, M. Breakspear, and C. C. Guo. Detecting changes in facial temperature induced by a sudden auditory stimulus based on deep learning-assisted face tracking. In *Scientific Reports*, 2019.

[30] H.-W. Tzeng, H.-C. Lee, and M.-Y. Chen. The design of isotherm face recognition technique based on nostril localization. In *Proceedings 2011 International Conference on System Science and Engineering*, pages 82–86. IEEE, 2011.

[31] S. Wang, Z. Liu, P. Shen, and Q. Ji. Eye localization from thermal infrared images. *Pattern Recognition*, 46(10):2613–2621, 2013.

[32] Y. Wu and Q. Ji. Facial landmark detection: A literature survey. *International Journal of Computer Vision*, 127(2):115–142, 2019.

[33] X. Xiong and F. De la Torre. Supervised descent method and its applications to face alignment. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 532–539, 2013.

[34] X. Yin, X. Yu, K. Sohn, X. Liu, and M. Chandraker. Towards large-pose face frontalization in the wild. pages 3990–3999, 2017.

[35] G. Zhao, X. Huang, M. Taini, S. Z. Li, and M. Pietikäinen. Facial expression recognition from near-infrared videos. *Image and Vision Computing*, 29(9):607 – 619, 2011.