# Effects of Pre-processing on the Performance of Transfer Learning Based Person Detection in Thermal Images

Huda, Noor Ul; Gade, Rikke; Moeslund, Thomas B.

*Published in:*
Proceedings of IEEE 2nd International Conference on Pattern Recognition and Machine Learning

*Publication date:*
2021

*Document Version*
Accepted author manuscript, peer reviewed version

# Effects of Pre-processing on the Performance of Transfer Learning Based Person Detection in Thermal Images


1st Noor Ul Huda
*Dept. of Architecture, Design and Media Technology*
*Aalborg University*
Aalborg, Denmark
0000-0003-2096-6200

2nd Rikke Gade
*Dept. of Architecture, Design and Media Technology*
*Aalborg University*
Aalborg, Denmark
0000-0002-8016-2426

3rd Thomas B. Moeslund
*Dept. of Architecture, Design and Media Technology*
*Aalborg University*
Aalborg, Denmark
0000-0001-7584-5209



***Abstract*—Thermal images have the property of identifying objects even in low light conditions. However, person detection in thermal is tricky, due to varying person representations depending upon the surrounding temperature. Three major polarities are commonly observed in these representations i.e., 1. person warmer than the background, 2. person colder than the background and 3. person's body temperature is similar to background. In this work, we have studied and analyzed the performance of the detection network by using the data in its original form and by harmonizing the person representation in two ways i.e., dark persons in the light background and light persons in a darker background. The data passed to each testing scenario was first pre-processed using histogram stretching to enhance the contrast. The work also presents the method to separate the three kinds of images from thermal data. The analysis is performed on publicly available outdoor AAU-PD-T and OSU-T datasets. Precision, recall, and F1 score is used to evaluate network performance. The results have shown that network performance is not enhanced by performing the mentioned pre-processing. Best results are obtained by using the data in its original form.**



***Index Terms*—Thermal images, Pre-processing, Transfer learning, Thermal polarities, Thermal data, Heat effects**


## I. Introduction

Person detection is the fundamental problem in any human-related computer vision based approach. Different camera setups, including RGB and thermal, have been proposed for person detection.

In RGB domain, well-defined solutions based on machine learning and deep learning have been reported. However, RGB cameras have their limitation in low light and total blackout conditions. Whereas thermal cameras have the advantage to perform very well in such situations. On the contrary, the outcome from the thermal cameras gets affected by the temperature of the surrounding objects as well as the temperature of the environment.

Most of the reported solutions in thermal domain are based on finest videos and image-based data experiments. The data used had the characteristics of higher person temperature vs lower background temperature, along with small occlusion, less reflection and weather effects. Therefore, It becomes less effective to apply the available research to outdoor environments.

The heat effect on thermal images has the worst outcome. It normally results in three different polarities of person representation in an image, as shown in Fig. 1: 1. person appearing darker as compared to the background. It occurs when the environmental temperature gets higher than the person temperature. 2. person appearing lighter than the background, which is when the person gets much warmer compared to the environment. 3. person appearance similar to the background. This happens when the person temperature as well as the environment temperature rise to the same level.

In this era of deep learning and big data solving, the approach to deal with the aforementioned or any other challenges is to increase the dataset either by producing synthetic data or by recording more videos. Increase the training data helps the classifier to learn every possible effect. On the other hand, some studies have proposed pre-processing techniques for thermal data to improve the performance of the convolutional neural network (CNN) [1], [2].

In this work, we have investigated the effects of pre-processing the thermal data for person detection using deep learning network. The implied pre-processing approach is to homogenize the data by converting the images to the same polarities/representation. Each representation is tested by using similar CNN and train settings to analyze which data type helps the network to perform better. We hypothesise that proposed pre-processing techniques should improve the performance of the subsequent algorithms. The pre-processing should help in model adaptation if they have an impact. Based on the above hypothesis following are the two main contributions in this paper, 1. Evaluation of polarity homogenization based pre-processing techniques for person detection using a deep neural network in thermal images, 2. A new method for polarity detection and polarity

Fig. 1. Person representation in thermal images (a) Person appeared darker w.r.t background, (b) Person appeared lighter w.r.t background, (c) Person appeared similar to background

homogenization in thermal data.

The rest of the paper is arranged as: Section 2 presents the overview of thermal person detection techniques as well as pre-processing methods explored both in thermal and RGB data. Section 3 describes the proposed experiments, and the results are presented in Section 4. Section 5 concludes this work.

## II. Related work

In this section, deep learning solutions for thermal person detection and pre-processing techniques for enhancing deep network performance are introduced.

Many deep learning techniques have also been proposed for person detection. In [3] maximally stable extremal regions (MSERs) and CNN based methods were proposed for person detection using thermal pedestrian [4], OSU color thermal [5] and terravic motion IR [6] datasets. Tumas et. al [7] combined HOG and CNN for pedestrian detection for FIR domain. Heo et. al [8] combined YOLO and adaptive boolean-map-based saliency methods for pedestrian detection using CVC-09 [9] dataset. In [1] Huda et. al investigated the effects of environment and weather conditions for players detection in the outdoor soccer field. They used the transfer-learning approach using YOLO3 for analysis and evaluation. In [10], authors implemented a cascade object detector to detect human silhouette in thermal images. They aimed to develop a pedestrian detection system in poor light conditions. Zhang et. al [11] addressed the lack of availability of thermal dataset for implementing CNN networks. To deal with this challenge authors proposed RGB to thermal translation models to generate synthetic thermal data for training deep networks.

In the domain of deep learning several pre-processing techniques have been reported for improving the detection performance. In [12] zero component analysis is reported to have the most significant effect on the performance of image classification using CNN. The noise removing techniques i.e., non-local filtering, bilateral filtering and total variation denoising methods were studied to improve the image quality before it is passed to deep neural network [13]. Diah et. al [14] studied the influence of resizing, face detection, cropping, adding noise and normalization on CNN performance for emotion detection. Francisco et. al [15] found intensity normalization to have the most effect on the diagnosis of Parkinson's disease using CNN based models. In [16] logarithmic and square root transformation methods were proposed for enhancing mammogram to detect breast cancer. Square root was found to have more influence on the performance of CNN based detection network. Image inversion, blurring, histogram stretching, and equalization methods were used to pre-process thermal images for person detection using CNN [2]. Homomorphic filtering and OTSU thresholding methods are proposed in [17] for improving image quality for concrete cracks detection using CNN. In [18] it is suggested that it is better to not pre-process the data. Their results showed that most CNN networks perform better if trained from scratch and only using data augmentation. Whereas, in [2] it is shown that inversion and histogram stretching techniques perform better while training a CNN based model using transfer learning for thermal person detection. Huda et. al [1] also proposed that inversion of thermal images to same person representation w.r.t background may help in the improvement of performance.

In the reported literature, we have perceived that the impact of pre-processing is mostly positive. In a few cases, the pre-processing does not improve the performance of the detection network. Our aim in this paper is to investigate and evaluate the impact of data homogenization based pre-processing technique using a CNN network performance on a thermal person detection dataset.

## III. Methodology

In this work, the key investigation is to evaluate the role of image homogenization and image enhancement based pre-processing techniques used for transfer learning from a pretrained CNN network. We have used Yolov3, which is pretrained on RGB Imagenet data. The network is utilized due to its higher detection accuracy and the ability to detect smaller objects in an image.

As discussed earlier, heat effects alter the representation of a person in a thermal image in three ways.

- Person appeared lighter w.r.t background
- Person appeared darker w.r.t background

- Person appeared similar to background

To add more clarity in the images of similar background, we are using histogram stretching to enhance the person intensities in the background. After histogram stretching, we have proposed and tested the following possible ways of passing the data to the learning network. The possibilities are graphically explained in Fig. 2 and the details are as follows.

- Normal data: Both train and test datasets are kept in their original form.
- Enhanced data: Histogram stretching is applied to both train and test datasets to enhance the image contrast.
- Light person: Train dataset is homogenized and enhanced. The homogenization is performed by inverting and making the person representation lighter on dark background for all the images. Image enhancement is performed by histogram stretching. The test dataset is altered by detecting and inverting the events of dark person representation on the lighter background (the procedure of detection is explained later in the section). The test data is also enhanced by using histogram stretching.
- Dark person: Train dataset is homogenized and enhanced. The homogenization is performed by inverting and making the person representation darker w.r.t the lighter background. Image enhancement is performed by histogram stretching. In the test dataset, the events of light person representation with a darker background are detected and inverted. The test data is also enhanced by using histogram stretching.
- Light person test on normal data: Train dataset is homogenized and enhanced. The homogenization is performed by inverting and making the person representation lighter on dark background for all the images. Image enhancement is performed by histogram stretching. Test data is used as it is.
- Dark person test on normal data: The homogenization is performed by inverting and making the person representation darker on light background for all the images. Image enhancement is performed by histogram stretching. Test data is used as it is.

For the testing data, the first step is the detection of the polarity of the image. Detection of events is performed in two steps, i.e., 1- sunlight detection and 2- human temperature detection.

**Sun light detection:** In thermal imagery, the images captured in high sunlight are the images that are brighter than the other images. For the detection of brighter images, image segmentation is performed. Sum of entropy-based thresholding is used for segmentation of images to get the brighter spots separated from the darker spots in the images. Afterwards, the accumulated pixel value is calculated by summing up all the pixels. The sum represents the number of bright pixels. A threshold is applied to the number of bright pixels to identify the images with high sunlight.

**Low human body temperature detection:** After the detection of brighter images, the next step is to find the lower person intensities in those brighter images. The data is separated based on histogram maps of both polarities of images i.e., light person with a dark background and dark person with a light background, separately. Two main features are used to detect low body temperature i.e., 1. intensity peaks vs the number of intensity bin in the histogram map of the image and 2. value of peak intensity in the histogram map.

## IV. Evaluation protocol

### A. Training

*1) Dataset:* For training, we have used the training data presented in [1], [19]. [19] is an indoor recorded soccer thermal dataset and consists of 3000 images. Every image consists of 8 persons playing soccer and occludes at some point while playing. The data is used as pre-training thermal data [1]. AAU-PD-T [1] training and testing data consist of 1941 and 1000 images, respectively, from the outdoor soccer field. In these images persons are running, playing, and doing exercises. The data is characterized as far players (far), images with snow (snow), images with the wind (wind), good outdoor (NR), occlusion (Oc), opposite temperature (OT), shadow (Sh), and similar temperature (ST).

*2) Network setting for training:* Re-training of Yolov3 is performed using training data. In order to maintain consistency in results, training setup parameters, except number of iterations, are kept constant and exactly same as reported in [1] (i.e., learning rate = 0.001, momentum = 0.9, and decay = 0.0005). The iterations are set to 10000 to have long term analysis and observe deviations if any.

### B. Testing

*1) Sun light detection:* AAU-PD-T dataset is used for separating high-temperature images based on thresholding as defined in section III. In the AAU-PD-T dataset, OT and ST category contain the images with high sunlight. Fig. 3 clearly shows that sum of pixel values is directly related to image intensity values. NR category images also show some spikes for the sum of pixel values due to some images captured on sunny days.

*2) Low human body temperature detection:* AAU-PD-T is used to observe the threshold values to separate the two polarities. It can be seen in fig. 4 that images with the person having lower body temperature than background seem to have lower intensities and a dip between bins 100 and 200. Also, these images mostly have two peaks. Whereas the images with persons having higher body temperature than background have high-intensity values especially in middle bins i.e., between bin 100 and bin 200.

*3) Dataset for evaluating the testing scenerios:* AAU-PD-T [1] test data and OSU-T [4] data are used for the evaluation of testing scenarios presented in Fig. 2. AAU-PD-T contains 1000 image with variable polarities. OSU-T dataset consists of 284 images. It is recorded in the Ohio State University campus at a pedestrian intersection and has images with both polarities.

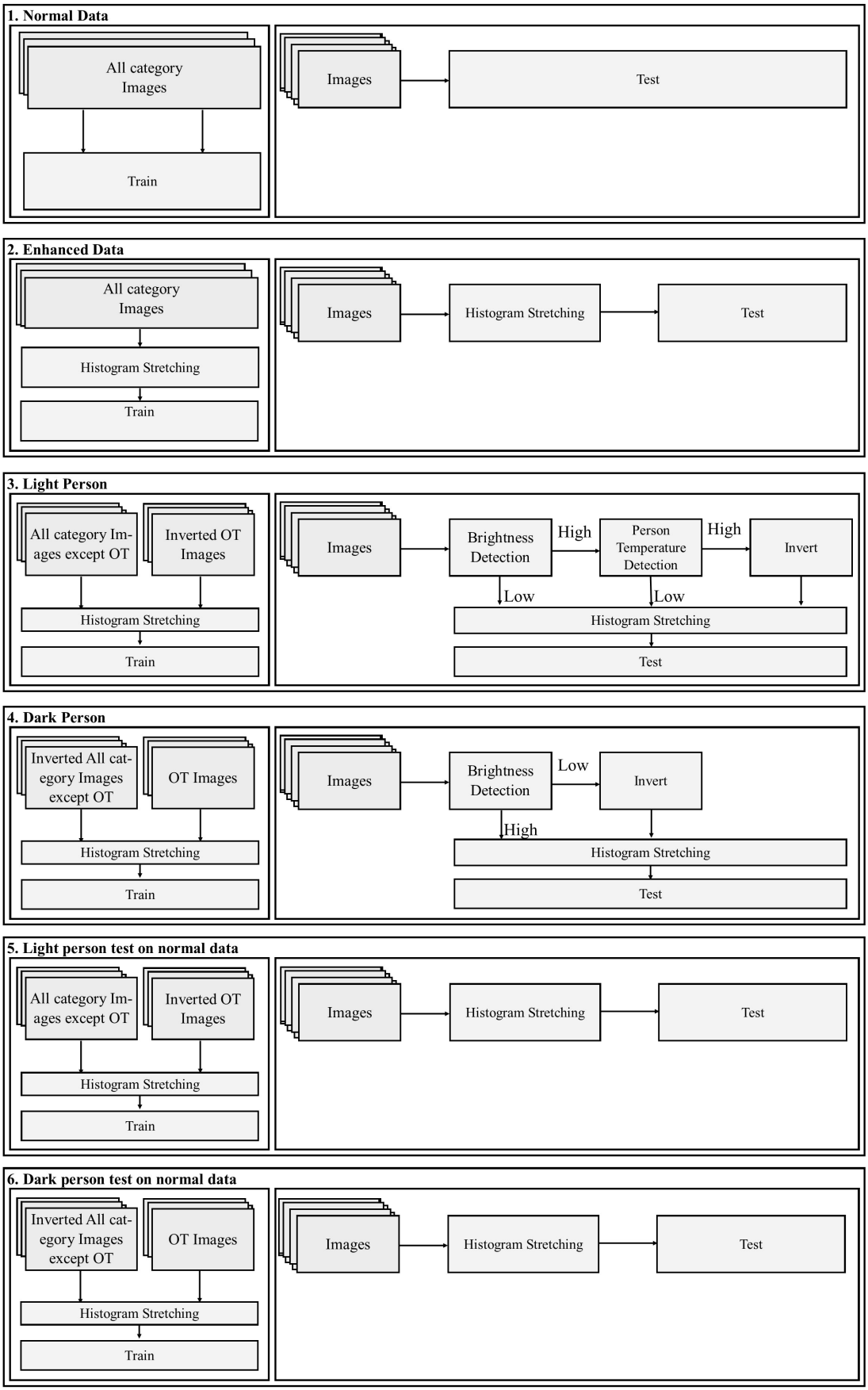


Fig. 2. Test setups

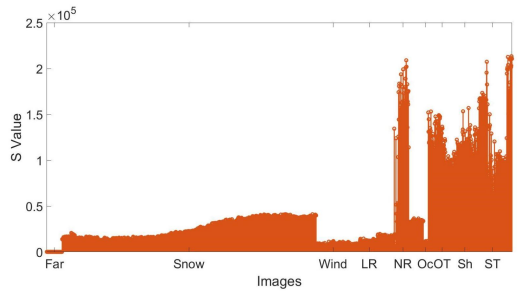


Fig. 3. Sum of pixel values vs categorical data

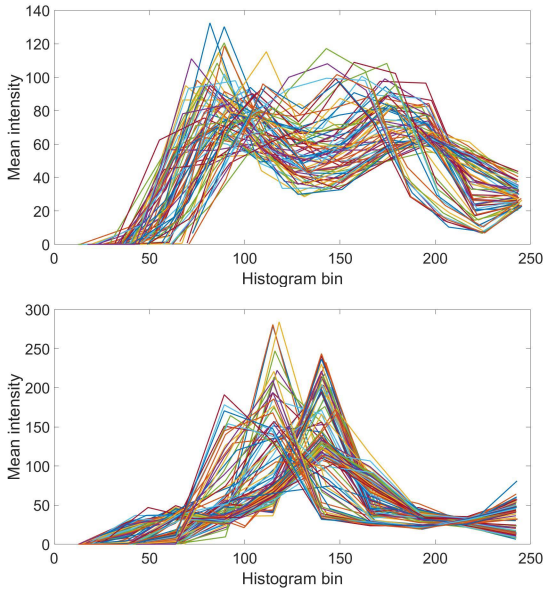


Fig. 4. Mean intensity values for respective histogram bins. (a) Images with lower body temperature than background. (b) Images with higher body temperature than background.

### C. Evaluation parameters

Precision and recall are used as evaluation parameters.

$$\text{F1 score} = \frac{\text{2xPrecisionxRecall}}{\text{Precision+Recall}} \tag{1}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP + FN}}, \quad \text{Precision} = \frac{\text{TP}}{\text{TP + FP}} \tag{2}$$

Here TP (True Positives) are the number of persons correctly detected, FP (False Positives) are the number of persons that are falsely detected. TN (True Negatives) are the correctly identified negative frames (frames with no persons) and FN (False Negatives) are the persons that are not detected. The training and testing of all combinations are performed using a graphical processing unit GTX 1080 with Linux Ubuntu 16.04.

## V. Results

The results for each iteration are plotted in Fig. 5, and Fig. 6 and the results of max F1 Score for all iterations is presented in Table I.

It can be seen from Fig. 5 and Fig. 6 that there is no significant difference in performance, i.e., precision and recall, between each testing setup except for dark person test on normal data. The performance of this setup is the lowest, which is understandable because most of the images in the datasets has a person representation of lighter intensity w.r.t the background. The results of the AAU-PD-T dataset shows that precision is similar for each test setup, whereas recall is higher for normal data test setup. The lower performance of recall for pre-processed datasets can be explained by the fact that training CNN network with high contrast images increases the chances of miss-detection for slightly lower contrast images.

The results of the OSU dataset shows that recall is higher than precision, which shows that FP detection is higher than FN detections. This is because the OSU dataset has better contrast images, and the person is labelled when appeared more than 50%. In the training, dataset person visibility varies and it is not fixed to any percentage for labelling. Therefore, any person that is detected in the OSU dataset with visibility lower than 50% will appear as FP. This effect got enhanced when the images are pre-processed. The overall results in

TABLE I
F1 Score for test setups

| Data | AAU-PD-T | OSU-T |
|---|---|---|
| Normal data | **0.70** | **0.65** |
| Enhanced data | 0.69 | 0.58 |
| Light person | 0.67 | 0.63 |
| Dark person | 0.65 | 0.61 |
| Light person test on normal data | 0.66 | 0.63 |
| Dark person test on normal data | 0.18 | 0.12 |

Table I show that the best performance is achieved when CNN is trained with normal data in its original form. When the network is trained with pre-processed data, it gets sensitive to a particular type of data and can react to any small change in testing data. In real-life applications, this kind of processed data can have more drastic effects as data is unpredictable.

## VI. Conclusion

In this work, the performance of a deep learning-based person detection network is analyzed for thermal data. The impact of inversion for creating homogeneity in person representation and histogram stretching for image enhancement were evaluated by proposing six testing setups. Results showed that the performance of the CNN network does not improve by pre-processing. Techniques improving the contrast of the images have a negative impact as the data in real-life scenarios is not restricted to better contrast images and increasing contrast may also enhance the noise. As different studies [2] have shown, pre-processing to improve data diversity and amount of data to process may help to improve detection performance. On the other hand, if the dataset is diverse and pre-processing is used to restrict the data in one form or to induces homogeneity, it causes degradation in the performance of the deep neural network.

## References

[1] N. U. Huda, B. D. Hansen, R. Gade, and T. B. Moeslund, "The effect of a diverse dataset for transfer learning in thermal person detection," *Sensors*, vol. 20, no. 7, p. 1982, 2020.

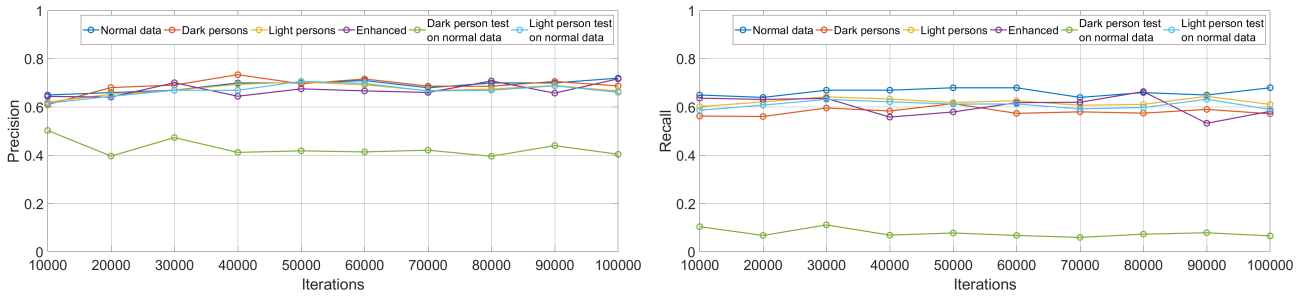


Fig. 5. Results of AAU-PD-T [1] dataset.

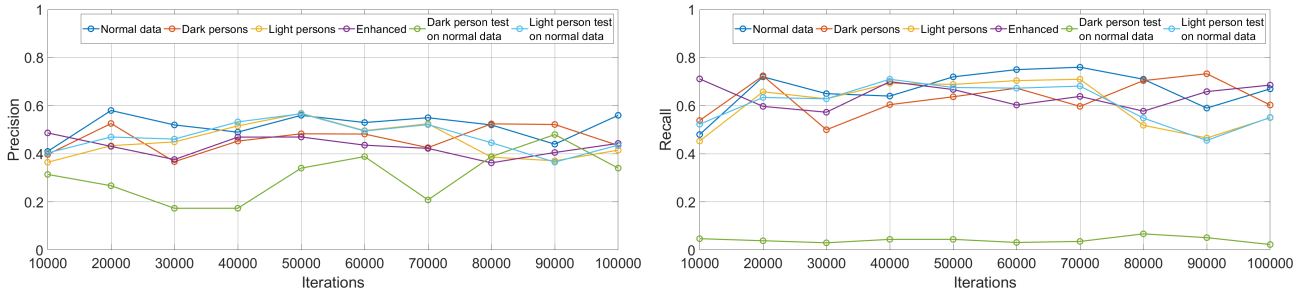


Fig. 6. Results of OSU-T [4] dataset


[2] C. Herrmann, M. Ruf, and J. Beyerer, "Cnn-based thermal infrared person detection by domain adaptation," in *Autonomous Systems: Sensors, Vehicles, Security, and the Internet of Everything*, vol. 10643. International Society for Optics and Photonics, 2018, p. 1064308.

[3] C. Herrmann, T. Müller, D. Willersinn, and J. Beyerer, "Real-time person detection in low-resolution thermal infrared imagery with mser and cnns," vol. 9987, 2016.

[4] J. W. Davis and M. A. Keck, "A two-stage template approach to person detection in thermal imagery," in *IEEE Workshops on Applications of Computer Vision*, vol. 1, Jan. 2005, pp. 364–369.

[5] J. W. Davis and V. Sharma, "Background-subtraction using contour-based fusion of thermal and visible imagery," *Computer Vision and Image Understanding*, vol. 106, no. 2, pp. 162–182, 2007.

[6] R. Miezianko and D. Pokrajac, "People detection in low resolution infrared videos," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, US, Jun. 2008, pp. 1–6.

[7] P. Tumas, A. Jonkus, and A. Serackis, "Acceleration of hog based pedestrian detection in fir camera video stream," in *2018 Open Conference of Electrical, Electronic and Information Sciences (eStream)*. IEEE, 2018, pp. 1–4.

[8] D. Heo, E. Lee, and B. C. Ko, "Pedestrian detection at night using deep neural networks and saliency maps," *Electronic Imaging*, vol. 2018, no. 17, pp. 060 403–1–060 403–9, 2018.

[9] Y. Socarras, S. Ramos, D. Vázquez, A. López, and T. Gevers, "Adapting pedestrian detection from synthetic to far infrared images," in *International Conference on Computer Vision (ICCV) Workshop*, jan. 2013.

[10] A. Nowosielski, K. Małecki, P. Forczmański, A. Smoliński, and K. Krzywicki, "Embedded night-vision system for pedestrian detection," *IEEE Sensors Journal*, 2020.

[11] L. Zhang, A. Gonzalez-Garcia, J. van de Weijer, M. Danelljan, and F. S. Khan, "Synthetic data generation for end-to-end thermal infrared tracking," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 1837–1850, 2018.

[12] K. K. Pal and K. Sudeep, "Preprocessing for image classification by convolutional neural networks," in *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*. IEEE, 2016, pp. 1778–1781.

[13] J. Yim and K.-A. Sohn, "Enhancing the performance of convolutional neural networks on quality degraded datasets," in *2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, 2017, pp. 1–8.

[14] D. A. Pitaloka, A. Wulandari, T. Basaruddin, and D. Y. Liliana, "Enhancing cnn with preprocessing stage in automatic emotion recognition," *Procedia computer science*, vol. 116, pp. 523–529, 2017.

[15] F. J. Martinez-Murcia, J. M. Górriz, J. Ramírez, and A. Ortiz, "Convolutional neural networks for neuroimaging in parkinson's disease: is preprocessing needed?" *International journal of neural systems*, vol. 28, no. 10, p. 1850035, 2018.

[16] A. Marchesi, A. Bria, C. Marrocco, M. Molinara, J.-J. Mordang, F. Tortorella, and N. Karssemeijer, "The effect of mammogram preprocessing on microcalcification detection with convolutional neural networks," in *2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, 2017, pp. 207–212.

[17] R. Fu, H. Xu, Z. Wang, L. Shen, M. Cao, T. Liu, and D. Novák, "Enhanced intelligent identification of concrete cracks using multi-layered image preprocessing-aided convolutional neural networks," *Sensors*, vol. 20, no. 7, p. 2021, 2020.

[18] L. F. Rodrigues, M. C. Naldi, and J. F. Mari, "Comparing convolutional neural networks and preprocessing techniques for hep-2 cell classification in immunofluorescence images," *Computers in Biology and Medicine*, vol. 116, p. 103542, 2020.

[19] R. Gade and T. B. Moeslund, "Constrained multi-target tracking for team sports activities," *IPSJ Transactions on Computer Vision and Applications*, 2018.


## VII. Authors Information

| Name | Title | Research field |
|---|---|---|
| Noor Ul Huda | PhD Candidate | Video Data Analysis, Image processing https://vbn.aau.dk/da/persons/140446/projects/ |
| Rikke Gade | Associate Professor | Computer vision, Sports Analysis, http://rgade.blog.aau.dk/ |
| Thomas B. Moeslund | Full Professor | Computer vision and AI, http://thbm.blog.aau.dk/ |