Journal of Information Technology Management

Print ISSN: 2008-5893

Online ISSN: 2423-5059

Deep-Learning-CNN for Detecting Covered Faces with Niqab

Abdulaziz A. Alashbi*

* Corresponding Author, Ph.D. Candidate, Media and Game Innovation Centre of Excellence, Institute of Human Centered Engineering, University Technology Malaysia, 81310 Skudai, Johor, Malaysia. 2School of Computing, Faculty of Engineering, University Technology Malaysia, 81310, Skudai, Johor, Malaysia. E-mail: asaabdulaziz2@live.utm.my

Mohd Shahrizal Sunar

Professor, Media and Game Innovation Centre of Excellence, Institute of Human Centered Engineering, University Technology Malaysia, 81310 Skudai, Johor, Malaysia. 2School of Computing, Faculty of Engineering, University Technology Malaysia, 81310, Skudai, Johor, Malaysia. E-mail: shahrizal@utm.my

Zieb Alqahtani

Ph.D. Candidate, Media and Game Innovation Centre of Excellence, Institute of Human Centered Engineering, University Technology Malaysia, 81310 Skudai, Johor, Malaysia. E-mail: zralqahtani@graduate.utm.my

Abstract

Detecting occluded faces is a non-trivial problem for face detection in computer vision. This challenge becomes more difficult when the occlusion covers majority of the face. Despite the high performance of current state-of-the-art face detection algorithms, the detection of occluded and covered faces is an unsolved problem and is still worthy of study. In this paper, a deep-learning-face-detection model Niqab-Face-Detector is proposed along with context-based labeling technique for detecting unconstrained veiled faces such as faces covered with niqab. An experimental test was conducted to evaluate the performances of the proposed model using the Niqab-Face dataset. The experiment showed encouraging results and improved accuracy compared with state-of-the-art face detection algorithms.

Keywords: Face-detection, Object-detection, Computer Vison, Deep learning, Artificial Intelligence, Convolutional Neural Network

DOI: https://orcid.org/10.22059/jitm.2022.84888 Manuscript Type: Research Paper University of Tehran, Faculty of Management

Received July 23, 2021 Accepted March 25, 2021

Introduction

Face detection; a sub-domain of object detection is a well-known area in computer vision which has been investigated for two decades and remains a trending research area. It is considered the first step for all face application systems related, such as face verification, face recognition, video surveillance, human computer interaction and emotion recognition (Bai, Zhang, Ding, & Ghanem, 2018; Masi, Wu, Hassner, & Natarajan, 2018; Zhang, Wu, Hoi, & Zhu, 2020).

The successful implementation of machine learning from the stunning work of Viola and Jones has triggered a lot of improvements in face detection (Viola & Jones, 2001).

Impressive progress has been made in detecting human faces from digital images where an average performance of 98% is achieved by Hu and Ramanan (2017) in the unconstrained face-detection benchmark (FDDB) dataset (Jain & Learned-Miller, 2010). However, the detection of faces in certain scenarios such as occluded and partially occluded face detection remains a challenge in computer vision and worth of investigation. The challenge in occluded faces usually arises when the covering of face is either partially or fully as in Figure (1)

The successful implementation of machine learning from the stunning work of Viola and Jones has triggered a lot of improvements in face detection (Viola & Jones, 2001). The covering may be due to work requirement such as masks as in hospitals or due to religious beliefs as in some Muslim societies where ladies are required to cover their faces while being outdoors or in the presence of non-relatives.

Figure 1. Examples of heavily occluded faces covered with Niqab

There is an increasing demand to improve face detectors for occlusion because the task is now becoming more complex and too hard (Ge, Li, Ye, & Luo, 2017). The detection of heavily occluded faces is critical for several applications such as video-surveillance and video-event-analysis. It is highly demanded for security monitoring and people-counting applications (Hou & Pang, 2010). However, very few researches have been done in that direction and directly addressing the detection under occlusion. For example, Hotta (2007) used local features with Support Vector Machine (SVM) to detect faces under partial occlusion, Chen, Song, and He (2018) proposed occlusion aware framework based on

convolutional neural network model to addresses the occlusion problem in face detection, Alafif, Hailat, Aslan, and Chen (2017) trained a single CNN model on large partial occluded faces images to detect unconstrained multi-view partially occluded and non-partially occluded faces.

Although faces under partial occlusion have been addressed in general by the aforementioned works, the challenge of highly occluded faces was not considered. Nowadays, security check and surveillance using human face detection is widely popular and highly demanded in almost everywhere in our daily life places such as public buildings, residential houses, airports and so on. Face detection algorithms have to be accurate so that no security breach due to undetected faces being occluded or covered. However, the current algorithms are not performing well when detecting heavily occluded faces.

This research significantly contributes by improving the performance of face detection for occluded and covered faces. We proposed in this paper Niqab-Face-Detection model, which is a Deep-learning Face Detection model to detect heavily occluded and covered faces in digital images. This research helped to improve the detection and localization of heavily occluded faces such as faces covered with niqab as in Figure (1). An experimental test conducted to examine its performance in Niqab-face dataset benchmark (Alashbi & Sunar, 2019). The obtained results indicated a competitive performance compared to the current state-of-the-art face-detection algorithms.

Literature Review

The detection of human face in images has been heavily studied as a sub-area of computer vision researches since 1960s. one of the earliest works that tried to classify human face in digital photos was proposed in 1972 by(Sakai, Nagao, & Kanade, 1972). Several hundreds of methods, techniques and approaches had been proposed for face detection. A well-organized review on early works of face detection researches up until 1999 can be found in (M.-H. Yang, Kriegman, & Ahuja, 2002) and (Hjelmås & Low, 2001).

However, most of these early works were inapplicable in the real-world situations and unable to obtain better performance in unconstrained conditions such as occluded faces until the existence of the brilliant work of viola-Jones face detection (Zafeiriou, Zhang, & Zhang, 2015). It enabled the application and implementation of face detection in real world applications such as in digital cameras. In 2007 a framework based on boosting algorithms and cascade structures proposed by Y.-Y. Lin, Liu, and Fuh (2007) to efficiently detect faces with occlusion by applying reinforcement training in order to reduce false positive alarms and cascading with evidence for detecting occluded faces.

Deformable Part–based Model which uses a single filter on histogram of oriented gradients (HOG) features to represent an object category had potentially obtained better accuracy over cascade and boosting features. Wang, Han, and Yan (2009) proposed an approach to integrate the advantage of part-based detectors to the sliding window by

aggregating Histogram of Gradient Local Binary Pattern HOG-LBP to handle the detection of human occlusion. A hierarchical deformable part model was described by Ghiasi and Fowlkes (2015) to captures face appearance, shape and occlusion. J. Kim, Sung, Yoon, and Park (2005) proposed a classification algorithm based on SVM to distinguish the partially occluded face from the normal face or non-facial images for surveillance ATM system. B. Kim, Ban, and Lee (2009) utilized the depth information of an eye obtained by binocular saliency map model to detect multiple occlusions of human faces in a stereo image. In Liao, Jain, and Li (2016) they proposed an effective feature for unconstrained scenes as in occluded faces called Normalized Pixel Differences (NPD). To deal with partially occluded faces as in heads covered with scarf for instance, Qezavati, Majidi, and Manzuri (2019) combined haar-cascade and LBPH to extract features and SVM classifier for the classifications. However due to its complexity and high computational cost this made it difficult to be used in real-world application. In general, machine learning based face detection approaches design features manually and optimize classifiers separately. It is not an end-to-end architecture.

The new era of computer vision utilizing the deep-learning-CNN-based for image classification and object detection algorithms that can learn features automatically and do optimization simultaneously in an end-to-end architecture. It started in 2012 when AlexNet, a Deep-Learning-CNN classification model achieved top best performance in ImageNet classification challenge (Krizhevsky, Sutskever, & Hinton, 2012).

Deep-learning-CNN methods for face detection can be categorized into two approaches; Region-Proposal-based (RPN) approach and Single Shot Detection (SSD). Region-proposal architecture works as two-stage detection, in the first stage Region Proposal network (RPN) generates pool of 2000 proposals per image followed by a second stage Deep-Learning-CNN classifier to classify the proposed proposals, the majority of Region-based approaches are based on (Ranjan, Patel, & Chellappa, 2017) and (Jiang & Learned-Miller, 2016). The drawback of this approach is the generation task increases the overall computation time that makes it difficult to be adopted as real-time face detection due to its slow speed. Sliding window-based approaches is the other category in which a face score is computed along with its bounding box coordinates at every location in the feature map in different scales, therefore it is faster than region proposal. Single shot face detection (SSD) introduced by Najibi, Samangouei, Chellappa, and Davis (2017) and Redmon, Divvala, Girshick, and Farhadi (2016) algorithms are most common face detection algorithms belongs to this category. SSD deals with face as a single-shot problem and straight predicting and regressing face bounding box using deep-CNN. In YOLO the detection is a regression problem that takes an input image and learns the face probabilities with bounding box coordinates at once. In (Li, Lin, Shen, Brandt, & Hua, 2015) CNN-based cascade face detector proposed and outperformed all state-of-the-art detectors in the public benchmark AWF with 15fps in 2015. Multi-Task-CNN framework (MTCNN) proposed by (Qin, Yan, Li, & Hu, 2016), a famous framework for predicting simultaneously face detection and five landmark localizations. In the first stage CNN-layer P-Net a region proposal net-work which propose regions with bounding boxes, the

obtained regions are refined by Non-Maximum Suppression (NMS) technique to eliminate overlapped bounding boxes. The output of previous stage P-net is fed to the second stage network R-net which perform more filtering on false positive candidates and also apply NMS on bounding boxes for more calibration. The final stage O-net takes the output of R-net and output face bounding box along with five facial landmark points.

In the context of occluded face detection, grid Loss for detecting occluded faces was introduced by Opitz, Waltner, Poier, Possegger, and Bischof (2016) to address how to train CNN to detect occluded objects. They proposed a loss layer for CNN called grid loss which divides the conv layer to spatial blocks and optimize hinge loss for each block separately. This was directly applied to Face detection which is a subdomain from object detection. Adversarial Occlusion-aware Face Detection framework which directly addresses the occlusion problem in face detection was presented in (Chen et al., 2018). An impressive performance was achieved by Bai et al. (2018) in detecting very small faces by using Generative adversarial network (GAN) to learn the small faces and generate a high-resolution version after which the generated faces are passed to another classifier to classify faces from background.

In general, the current CNN-based detection algorithms suffer when dealing with the occlusion detection. They almost were trained on public datasets such as Widerface (S. Yang, Luo, Loy, & Tang, 2016) and FDDB (Jain & Learned-Miller, 2010). Although these datasets are very rich of unconstrained faces with different pose, lighting and illumination variation and with some degree of occlusions, heavily occluded faces as in Figure (1) are not included except of some photos of masked faces. These detectors were not explicitly trained on this kind of occlusion; therefore, their performance decreases dramatically in the presence of high degree of occlusion. This research addresses the problem of detecting faces in heavily occluded scenario as in faces veiled with niqab for instance.

Methodology

A) Deep Learning Method

In this paper Niqab-Face-detection model is proposed, which is a deep-learning-CNN-based model to detect the highly occluded faces, it can detect faces where most of the face parts are not visible such as faces covered with niqab. The model was trained on Niqab dataset with a pretrained model of Mobilenet-SSD face detection algorithm and used as the base of Niqab-face detection model. Mobilenet-SSD is a one stage face detector based on Single Shot Multibox Detector (Liu et al., 2016) which use Mobilenet-CNN architecture as the base feature extractor due to its computation efficiency(Howard et al., 2017). Given an input image SSD outputs multiple boxes of faces in one single forward pass of the network and apply non-max superstition (NMS) technique to eliminate most of the bounding boxes of low prediction scores keeping only the top with highest score, the architecture of Mobilenet-SSD is illustrated in figure (2).

Transfer learning, which is a common deep-learning practice technique is used to train the proposed Niqab-face to speed up the training. Inspired by a work of T.-Y. Lin, Goyal, Girshick, He, and Dollár (2017); (Wan, Chen, Zhang, Zhang, & Wong, 2016) the online hard examples mining and focal loss techniques were implemented during training to minimize false alarms and improve the training. Images were preprocessed, normalized and rescaled to 300x300 to fit the recommended input size of SSD-Mobilenet.

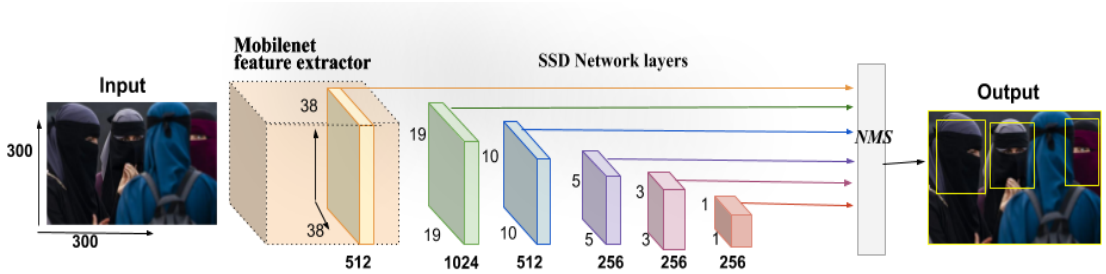

Figure 2. Mobilenet-SSD architecture composed of two parts: feature extraction and SDD

B) Dataset

The model was trained on Niqab-face, a dataset proposed specifically to address the challenge of heavily occluded and covered face detection. It consists of 10k of images with 12k faces, more than 50% of faces are highly occluded and covered with niqab or other face veil such as medical mask that block most of the face parts except of the two eyes. A total of 1200 images were used for training the model with an average of 4.3% faces per image. A dedicated of 30% of images were used for validation and testing.

Images were labelled manually; a bounding box was marked on each face with respect to the face contexts that may help adding additional information to avoid the conventional method as it is not suitable for occluded faces, it takes into account the visible parts of the face and ignoring the surrounding context, as illustrated in Figure (3).

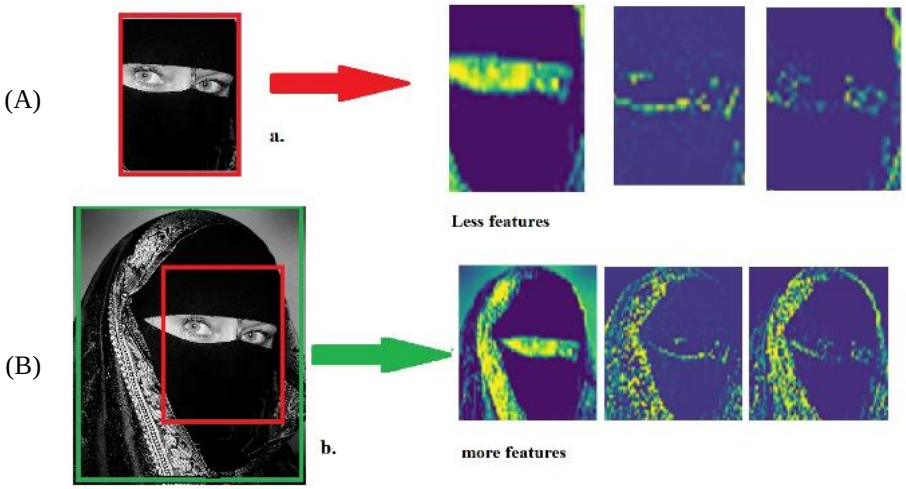

Figure 3. The importance of the labeling with context;
(A) Conventional Labeling, red rectangle labeling; face only with no context resulted in limited features learned by CNN.
(B) Contextual-labeling technique, green rectangle labeling with respect to context that gives more features to be learned be CNN.

C) Evaluation metrics

The task of face detection is to find if there is a face in an image and provide its location, the first step of evaluation metrics is comparing the localization output bounding box to the ground truth using Intersection Over Union (IOU) as in PASCAL VOC challenge (Everingham et al., 2015). For measuring class prediction scores the following metrics are used: TP, FP, FN, Precision and Recall.

True positive (TP) is used to indicate faces that have been predicted with prediction score > threshold and IOU > IOU threshold. False positive (FP) is defined as predicted faces with prediction score > threshold but has IOU < all ground truth IOU bounding boxes. False negative (FN) which is the number of ground truth faces not detected by the face detection algorithm. Precision is the number of TP divided by the sum of TP and FP. Recall is the number of TP divided by the total number of ground truth faces

Findings and Discussion

The training process was performed for approximate of 8 hours until an average precision of 90% and a Recall of 73.5% had obtained. During the testing phase 30% of images not used during training were used to evaluate and measure the model's performance. The major aim of our findings is to classify occluded face within an image and providing the face location within image as set of bounding box coordinates if there is face. Niqab-Face detector has successfully identified 261 as TP, one FP, and 179 as FN, with 99% and 59% for precision and recall respectively.

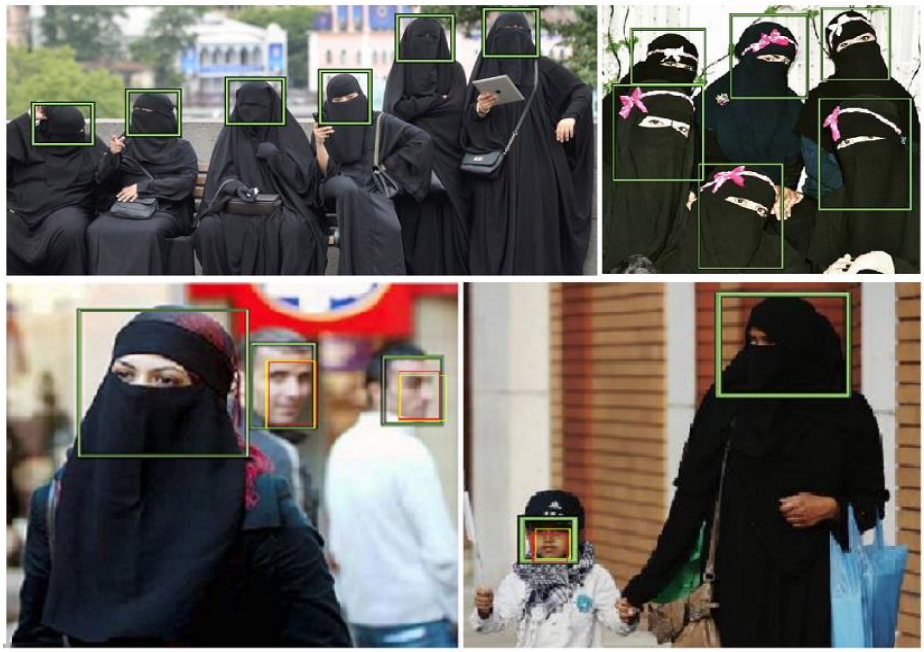

Figure 3. Example of detected faces by Niqab-Face detector indicated by green rectangle. MTCNN yellow rectangle and SSD red rectangle

Table (1) below shows the comparison results of our Niqab-Face-detection with previous studies of well-known face detection algorithms. MTCNN and Mobilenet, Niqab-Face Precision is 11% more than, Mobilenet and 4.8% more than MTCNN respectively. Table (1) summarize the results of our proposed model along with MTCNN and Mobilenet results.

Table 1. Niqab-Face performance comparison with MTCN and Mobilenet face detectors

Face detection algorithms	TP	FP	FN	Accuracy	Precision	Recall
MTCNN	82	5	379	18.5%	94.2%	17.7%
Mobilenet	97	12	355	21.9%	88.9%	21.2%
Niqab-Face	261	1	179	59.1%	99.6%	59.9%

It is assumed that the absence of heavily occluded faces images samples in the training dataset of MTCNN and Mobilenet algorithms were behind the low performance of these models. The use of other techniques such as hard sample mining and focal loss during training helped for increasing the good performance of Niqab-Face detector. Figure (4) shows examples photos detected by Niqab-Face and other two face detectors

Conclusion

In this paper, Niqab-Face detector was proposed to detect heavily occluded faces. The experiment showed a very encouraging result and improved performance compared with state-of-the-art face detection algorithms MTCNN and Mobilene. Although the score of 59.9% for recall obtained by Niqab-Face was the highest among the others, it still needs more improvement. The main reason of getting less recall score is the limited number of photos for training. Deep-learning-CNN face detectors requires a lot of images in order to improve its performance. Future work will be to retrain Niqab-face detector with more images along with other useful techniques such as data augmentation. Other deep-learning framework such as YOLO and Retina-Net will be investigated.

Conflict of interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article

References

Alafif, T., Hailat, Z., Aslan, M., & Chen, X. (2017). *On detecting partially occluded faces with pose variations.* Paper presented at the 2017 14th International Symposium on Pervasive Systems, Algorithms and Networks & 2017 11th International Conference on Frontier of Computer Science and Technology & 2017 Third International Symposium of Creative Computing (ISPAN-FCST-ISCC).

Alashbi, A. A. S., & Sunar, M. S. (2019). *Occluded Face Detection, Face in Niqab Dataset.* Paper presented at the International Conference of Reliable Information and Communication Technology.

Bai, Y., Zhang, Y., Ding, M., & Ghanem, B. (2018). *Finding tiny faces in the wild with generative adversarial network.* Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Chen, Y., Song, L., & He, R. (2018). Adversarial Occlusion-aware Face Detection. *arXiv preprint arXiv:1709.05188V6*.

Everingham, M., Eslami, S. A., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2015). The pascal visual object classes challenge: A retrospective. *International journal of computer vision, 111*(1), 98-136.

Ge, S., Li, J., Ye, Q., & Luo, Z. (2017). *Detecting masked faces in the wild with lle-cnns.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

Ghiasi, G., & Fowlkes, C. C. (2015). Occlusion coherence: Detecting and localizing occluded faces. *arXiv preprint arXiv:1506.08347*.

Hjelmås, E., & Low, B. K. (2001). Face detection: A survey. *Computer Vision and Image Understanding, 83*(3), 236-274.

Hotta, K. (2007). Robust face detection under partial occlusion. *Systems and Computers in Japan, 38*(13), 39-48.

Hou, Y.-L., & Pang, G. K. (2010). People counting and human detection in a challenging situation. *IEEE transactions on systems, man, and cybernetics-part a: systems and humans, 41*(1), 24-33.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., . . . Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.

Hu, P., & Ramanan, D. (2017). *Finding tiny faces.* Paper presented at the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Jain, V., & Learned-Miller, E. G. (2010). Fddb: A benchmark for face detection in unconstrained settings. *UMass Amherst Technical Report*.

Jiang, H., & Learned-Miller, E. (2016). *Face detection with the faster R-CNN.* Paper presented at the 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017).

Kim, B., Ban, S.-W., & Lee, M. (2009). *Multiple Occluded Face Detection Based on Binocular Saliency Map.* Paper presented at the International Conference on Neural Information Processing.

Kim, J., Sung, Y., Yoon, S. M., & Park, B. G. (2005). *A new video surveillance system employing occluded face detection.* Paper presented at the International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *Imagenet classification with deep convolutional neural networks.* Paper presented at the Advances in neural information processing systems.

Li, H., Lin, Z., Shen, X., Brandt, J., & Hua, G. (2015). *A convolutional neural network cascade for face detection.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

Liao, S., Jain, A. K., & Li, S. Z. (2016). A fast and accurate unconstrained face detector. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 38*(2), 211-223.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). *Focal loss for dense object detection.* Paper presented at the Proceedings of the IEEE international conference on computer vision.

Lin, Y.-Y., Liu, T.-L., & Fuh, C.-S. (2007). Face Detection with Occlusions. *Images & Recognition, 13*(1), 4-21.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). *Ssd: Single shot multibox detector.* Paper presented at the European conference on computer vision.

Masi, I., Wu, Y., Hassner, T., & Natarajan, P. (2018). *Deep face recognition: A survey.* Paper presented at the 2018 31st SIBGRAPI conference on graphics, patterns and images (SIBGRAPI).

Najibi, M., Samangouei, P., Chellappa, R., & Davis, L. S. (2017). *Ssh: Single stage headless face detector.* Paper presented at the Proceedings of the IEEE International Conference on Computer Vision.

Opitz, M., Waltner, G., Poier, G., Possegger, H., & Bischof, H. (2016). *Grid loss: Detecting occluded faces.* Paper presented at the European conference on computer vision.

Qezavati, H., Majidi, B., & Manzuri, M. T. (2019). *Partially Covered Face Detection in Presence of Headscarf for Surveillance Applications.* Paper presented at the 2019 4th International Conference on Pattern Recognition and Image Analysis (IPRIA).

Qin, H., Yan, J., Li, X., & Hu, X. (2016). *Joint training of cascaded CNN for face detection.* Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Ranjan, R., Patel, V. M., & Chellappa, R. (2017). Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 41*(1), 121-135.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). *You only look once: Unified, real-time object detection.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

Sakai, T., Nagao, M., & Kanade, T. (1972). *Computer analysis and classification of photographs of human faces*: Kyoto University.

Viola, P., & Jones, M. (2001). *Rapid object detection using a boosted cascade of simple features.* Paper presented at the Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on.

Wan, S., Chen, Z., Zhang, T., Zhang, B., & Wong, K.-k. (2016). Bootstrapping face detection with hard negative examples. *arXiv preprint arXiv:1608.02236*.

Wang, X., Han, T. X., & Yan, S. (2009). *An HOG-LBP human detector with partial occlusion handling.* Paper presented at the 2009 IEEE 12th international conference on computer vision.

Yang, M.-H., Kriegman, D. J., & Ahuja, N. (2002). Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24*(1), 34-58.

Yang, S., Luo, P., Loy, C.-C., & Tang, X. (2016). *Wider face: A face detection benchmark.* Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Zafeiriou, S., Zhang, C., & Zhang, Z. (2015). A survey on face detection in the wild: past, present and future. *Computer Vision and Image Understanding, 138*, 1-24.

Zhang, J., Wu, X., Hoi, S. C., & Zhu, J. (2020). Feature agglomeration networks for single stage face detection. *Neurocomputing, 380*, 180-189.

Bibliographic information of this paper for citing:

Alashbi, Abdulaziz A.; Shahrizal Sunar, Mohd & Alqahtani, Zieb (2022). Deep-Learning-CNN for Detecting Covered Faces with Niqab. *Journal of Information Technology Management*, Special Issue, 114-123.