2023-10

Neural Network based Unsupervised Face and Mask Detection in Surveillance Networks

Ani, Aninda Ruy

Independent University, Bangladesh

Draft Version

Neural Network based Unsupervised Face and Mask Detection in Surveillance Networks

Aninda Roy Ani, Sabir Saheel, Tarem Ahmed and Mohammad Faisal Uddin
RIoT Research Center
Department of Computer Science and Engineering
Independent University, Bangladesh
Dhaka, Bangladesh
Email: {1710551, 1710821, tarem, faisal}@iub.edu.bd

***Abstract*—In the post-pandemic world, surveillance cameras play a key aspect when it comes to detecting various kinds of security risks. These can range from burglars entering a premises to an individual wearing or not wearing a mask where convention dictates one way versus the other. We are proposing a system that would allow autonomously detecting these security risks with minimal human intervention. We propose using Multi-task Cascaded Convolutional Neural Networks as the face detector, a choice of a complete range of classic image feature extractors, and the Kernel-based Online Anomaly Detection algorithm to identify the potential risk in real time. We tested our proposed framework on three different datasets including real-world settings. Our proposed framework yielded high detection rates with low false alarm rates, in addition to being adaptive, portable, and requiring minimal infrastructure.**

I. Introduction

In the midst of the COVID-19 pandemic, wearing masks had become the new normal. Many places like malls, banks and educational institutions have made the wearing of face masks mandatory.

Software and autonomous tools for mask detection do exist, especially during the pandemic [1], but they are mostly proprietary, expensive, and not in wide deployment. Furthermore, because these systems are not open source, future developments on them are slow.

In this paper we are proposing a framework to help detect masked faces. Our proposed system is built upon widely available and open sourced image processing algorithms. Additionally, It can be used post-pandemic to detect intruders and burglars who have their faces covered suspiciously.

This paper also serves as the extension to the work done by Alvi et al. [2], where the Viola and Jones algorithm [3] was used for face detection. However, some studies [4]–[6] suggest that even with more complex features and classifiers, the Viola and Jones algorithm detector may deteriorate considerably in real-world applications. This is why we decided to move on to a contemporary, convolutional neural network algorithm.

Our framework consists of three key components: face detection, feature extraction, and unsupervised pattern matching technique, to create a system that can autonomously recognize anomalous events from an image sequence in real-time. We used the contemporary deep learning algorithm Multi-task Cascaded Convolutional Neural Network (MTCNN) [7] as our face detector. To conduct a comparative analysis of the basic feature extraction methods, we compared five classical methods: Haar wavelet decomposition [8], Gabor filter [9], and Canny [10], Laplacian [11] and Sobel [12] edge detectors. We then used the Kernel-based Online Anomaly Detection (KOAD) algorithm [13], [14] to detect anomalous images in real-time in an unsupervised manner.

A. Related work

A person identification system was developed by Aswal et al. [15] using two distinct methods a single-step pre-trained YOLOv3 model and a two-step process involving RetinaFace and VGGFace2. With a detection accuracy of 98.1%, they obtained encouraging findings. The datasets they utilized, however, only included 17 videos of seven individuals, all of which were filmed up close. They also stated that MTCNN did not achieve promising results in scaled, illuminated, and occluded faces. Contrarily, we found the MTCNN algorithm to perform sufficiently well even in the previously stated environments, and in unsupervised usage.

In the work on intruder detection by Alvi et al. [2], the Viola Jones algorithm was used for face detection followed by the KOAD algorithm to distinguish between masked and unmasked faces. The Viola Jones algorithm is much older than today's convolutional neural networks family, and could not detect multiple faces or faces from a distance which are more applicable for real world scenarios. Here we used the MTCNN algorithm and achieved a far greater accuracy in real-world applications.

KOAD was initially developed by Ahmed et al. [13], [14] to detect anomalies in backbone IP networks, and it was further extended in [16] where Kernel Density Estimates were proposed and for a fixed user tolerance level. KOAD was first used in automated surveillance in [17]. It was used to detect the presence of a rare intruder at a normally empty premises during the night. Further improvements were made by Anika et al. [18] where images were stitched from a moving camera.

B. Our Contribution

The primary objective of this paper was to use contemporary, open-sourced algorithms to identify whether a person is wearing a mask or not, in real world scenarios such as

Fig. 1: Sample of the images from Dataset 1 (online repository) with extracted faces using MTCNN.

tiny dictionary of approximately linearly independent items $\{\phi(\tilde{x_j})\}_{j=1}^{m}$. When compared to the approximation threshold ν, the projection error equation is:

$$\delta_t = min_a \; || \sum_{j=1}^{m} a_j \phi(\tilde{x_j}) - \phi(x_t) \; ||^2 \tag{2}$$

where $a = \{a_j\}_{j=1}^{m}$ is the optimal coefficient vector.

This projection error is calculated for each timestep in the KOAD algorithm [13], [14]. This error measure δ_t is then compared with two thresholds ν_1 and ν_2, where $\nu_1 < \nu_2$. If $\delta_t < \nu_1$, KOAD denotes that x_t is significantly linearly dependent on the dictionary and reflects standard behaviour. If $\delta_t > \nu_2$, KOAD determines that x_t is far outside the scope of normality and raises a "Red1" warning to notify an anomaly.

In the case if $\nu_1 < \delta_t < \nu_2$, KOAD deduces that x_t is significantly linearly independent from the dictionary to be evaluated as an unusual event. It could be an anomaly, or it could be the enlargement or movement of the space of normality itself. As a result, KOAD issues a "Orange" alarm, monitors the contribution of the relevant input vector x_t in understanding subsequent arrivals for l timesteps, and then makes a definite judgment. Details of the KOAD algorithm may be found in [22].

III. Experimental Setup

A. Dataset 1: Online Repository

This dataset [23] was collected from a public online repository. Example images are presented in Fig. 1. It consists of 7553 RGB images separated by portraits with and without masks. There were 3725 masked faces and 3828 non-masked faces. We have randomly chosen 90 images of non-masked faces and 10 images of masked faces. As we had to manually assign anomalies (masked faces), we have kept the number of data to a manageable quantity, as manually assigning anomalies to the full dataset would be unfeasible. We have done the same for Datasets 2 and 3 as well.

For Dataset 2 and 3, we have blurred the faces of each individual for privacy and also withheld the time stamps. Additionally, due to space limitations, we opted to present only one camera angle in this paper.

B. Dataset 2: Grocery Store

This dataset was collected from a grocery store. Sample images are presented in Fig. 2. CCTV footage was collected from two different cameras at different locations, covering two entrances to the grocery store. The continuous video resulted in 4274 frames, from which we selected 29. Our system extracted 75 cropped faces, which included 25 masked ones. Sample images with extracted faces are given in Fig. 3.

Fig. 2: Sample images from Dataset 2 (grocery store).

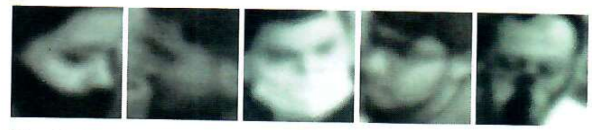

Fig. 3: Sample of the images from Dataset 2 (grocery store) with extracted faces using MTCNN.

C. Dataset 3: University

This dataset was collected from a university's CCTV surveillance system. Sample images are presented in Fig. 4. We collected footage from four different cameras. The continuous video feed here provided 11395 frames, from which we selected 33 images. Our system extracted 205 cropped faces, 43 of which were masked. Sample images from this dataset with extracted faces are presented in Fig. 5.

D. Research Methodology

For all three datasets, there were a variety of faces observed from different angles as well as in different poses. Datasets 2 and 3 contained faces at varied distances as they were taken from live CCTV footage. CNN initially uses boundary box regression and provides a boundary box when a face is detected. We have taken the coordinates of the boundary box to crop the face.

The next steps are applicable for all three datasets. The first step is preprocessing, which performs resizing and grayscaling. We have initially resized the images into 200×200 size since taking a similar size of images would be required for the final step where we use the KOAD algorithm to detect whether it is a face of a masked individual or not. After resizing the images, we converted the RGB images into grayscale.

We then move onto the feature extraction process. We used different filters to get different feature representations. Finally, the KOAD algorithm is run for pattern matching and masked face detection in an unsupervised manner.

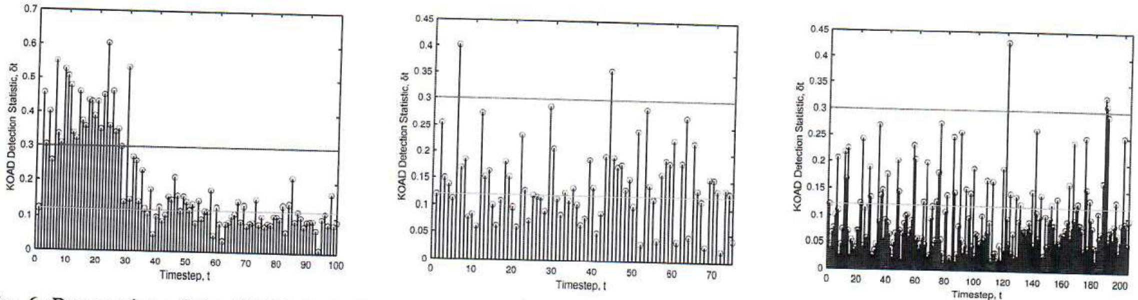

Fig. 6: Progression of the KOAD detection statistic δ_t with time when Gabor used as feature extractor for datasets 1, 2 and 3.

Method for feature extraction	**Threshold, v_1**	**Threshold, v_2**	**Dataset 1: Detection Rate (%)**	**Dataset 1: False Alarm Rate (%)**	**Dataset 2: Detection Rate (%)**	**Dataset 2: False Alarm Rate (%)**	**Dataset 3: Detection Rate (%)**	**Dataset 3: False Alarm Rate (%)**
Gabor	0.12	0.3	90	23.333	78	38	72.093	12.346
Wavelet Decomposition	0.1	0.2	0	0	0	0	0	0
Sobel	0.2	0.38	100	6.667	40	20	0	0.617
Laplacian	0.2	0.4	90	5.556	8	10	0	0.617
Canny	0.4	0.6	Invalid	100	Invalid	100	Invalid	100
Grayscaled	0.45	0.6	80	18.889	52	16	25.581	7.407
RGB	0.4	0.6	90	36.667	56	22	32.558	10.494

TABLE II: Detection and false alarm rate of different methods for all three datasets.

we used standard, readily-available and open-sourced building blocks to design this system. This makes it economical and easily implementable.

For future work, we would like to apply our framework on gait detection and night vision cameras.

References

[1] Datakalab, "Company webpage," available:. [Online]. Available: https://www.datakalab.com/

[2] A. Alvi, T. Ahmed, and M. F. Uddin, "Comparative study of traditional techniques for unsupervised autonomous intruder detection," in *International Conference on Advanced Information Networking and Applications*. Springer, 2021, pp. 519–530.

[3] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, vol. 1. IEEE, 2001, pp. I–I.

[4] B. Yang, J. Yan, Z. Lei, and S. Z. Li, "Aggregate channel features for multi-view face detection," in *IEEE international joint conference on biometrics*. IEEE, 2014, pp. 1–8.

[5] M.-T. Pham, Y. Gao, V.-D. D. Hoang, and T.-J. Cham, "Fast polygonal integration and its application in extending haar-like features to improve object detection," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2010, pp. 942–949.

[6] Q. Zhu, M.-C. Yeh, K.-T. Cheng, and S. Avidan, "Fast human detection using a cascade of histograms of oriented gradients," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2. IEEE, 2006, pp. 1491–1498.

[7] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016.

[8] D. Zhang, "Wavelet transform," *Fundamentals of Image Data Mining*, p. 35–44.

[9] L. Bui, D. Tran, X. Huang, and G. Chetty, "Face recognition based on gabor features," in *3rd European Workshop on Visual Information Processing*. IEEE, 2011, pp. 264–268.

[10] J. Canny, "A computation approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 670–700, 1986.

[11] X. Wang, "Laplacian operator-based edge detectors," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 5, pp. 886–890, 2007.

[12] N. Kanopoulos, N. Vasanthavada, and R. L. Baker, "Design of an image edge detection filter using the sobel operator," *IEEE Journal of solid-state circuits*, vol. 23, no. 2, pp. 358–367, 1988.

[13] T. Ahmed, M. Coates, and A. Lakhina, "Multivariate online anomaly detection using kernel recursive least squares," in *IEEE INFOCOM 2007-26th IEEE International Conference on Computer Communications*. IEEE, 2007, pp. 625–633.

[14] T. Ahmed, B. Oreshkin, and M. Coates, "Machine learning approaches to network anomaly detection," *Proc. SysML*, 2007.

[15] V. Aswal, O. Tupe, S. Shaikh, and N. N. Charniya, "Single camera masked face identification," in *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, Dec. 2020. [Online]. Available: https://doi.org/10.1109/icmla51294.2020.00018

[16] T. Ahmed, "Online anomaly detection using KDE," in *GLOBECOM 2009-2009 IEEE Global Telecommunications Conference*. IEEE, 2009, pp. 1–8.

[17] T. Ahmed, S. Ahmed, S. Ahmed, and M. Motiwala, "Real-time intruder detection in surveillance networks using adaptive kernel methods," in *2010 IEEE International Conference on Communications*. IEEE, 2010, pp. 1–5.

[18] A. Anika, K. L. Karim, R. Muntaha, F. Shahrear, S. Ahmed, and T. Ahmed, "Multi image retrieval for kernel-based automated intruder detection," in *2017 IEEE Region 10 Symposium (TENSYMP)*. IEEE, 2017, pp. 1–5.

[19] T. Ahmed, A.-S. K. Pathan, and S. Ahmed, "Adaptive algorithms for automated intruder detection in surveillance networks," in *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, 2014, pp. 2775–2780.

[20] C. Au, S. Skaff, and J. Clark, "Anomaly detection for video surveillance applications," in *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 4. IEEE, 2006, pp. 888–891.

[21] Y. Engel, S. Mannor, and R. Meir, "The kernel recursive least-squares algorithm," *IEEE Transactions on signal processing*, vol. 52, no. 8, pp. 2275–2285, 2004.

[22] S. Daneshgadeh, T. Ahmed, and A.-S. K. Pathan, "The kernel-based online anomaly detection algorithm: Detailed derivation and development," in *Security Analytics for the Internet of Everything*, A. U. M. Ahmed and A.-S. K. Pathan, Eds. USA: CRC Press, Feb 2020, pp. 155–192.

[23] "Face mask detection dataset." [Online]. Available: http://www.kaggle.com/omkargurav/face-mask-dataset