Real Time Multi Face Detection and Privacy Protection in Video Streams

Nawar Nihad Hasibe, Zied O. Ahmed and Anwar Hassan Al-Salh
Department of Computer Science, College of Science, Mustansiriyah University, Palestine Str., 10052 Baghdad, Iraq
Nono_nini38@uomustansiriyah.edu.iq, zied_othman@uomustansiriyah.edu.iq, Anwar.h.m@uomustansiriyah.edu.iq

Keywords: Face Detection, Recognition, Blurring, Deep Learning.

Abstract: This paper presents an intelligent and automated framework for multi-face detection, recognition, and real-time blurring in video sequences under realistic and complex conditions. The proposed system was developed and evaluated using a locally constructed dataset that includes significant variations in illumination, background complexity, facial pose, and crowd density. To effectively manage scene complexity, a deep learning-based approach was adopted for robust feature extraction and classification. The dataset was divided into 80% for training and 20% for testing, and the model was trained for five epochs to ensure computational efficiency while maintaining high performance. The proposed framework was evaluated using standard performance metrics to measure detection and recognition accuracy. Experimental results demonstrate that the system achieves a recognition accuracy exceeding 90% in complex real-time scenarios involving multiple individuals. Furthermore, the automatic face blurring module achieved a 100% success rate in protecting detected identities within the tested samples. The main contribution of this work lies in integrating detection, recognition, and privacy-preserving blurring into a unified intelligent system capable of operating in real-time environments. The proposed framework shows strong potential for practical applications in surveillance systems, security monitoring, and video data privacy protection.

1 INTRODUCTION

The widespread use of image and video sharing has raised serious concerns regarding privacy protection, particularly when facial information is involved. Facial data represents sensitive biometric information that can be exploited if not properly managed [1].

Recent progress in artificial intelligence and deep learning has led to substantial improvements in face detection and recognition systems [2], [3]. These technologies are widely applied in security monitoring, identity verification, and automated surveillance. However, processing multiple faces in dynamic video environments remains a challenging task, especially under varying lighting conditions, complex backgrounds, and different crowd densities, While many existing studies focus on detection or recognition independently, fewer research efforts address the integration of detection, classification, and privacy-preserving blurring within a unified framework. Ensuring accurate recognition while simultaneously protecting individual privacy through automated blurring introduces additional technical complexity [4]. Therefore, this research aims to develop and evaluate an intelligent and efficient system capable of detecting, recognizing, and automatically blurring multiple faces in both images and video sequences under realistic environmental conditions [5]. The proposed framework utilizes a customized dataset designed to simulate variations in lighting, background complexity, and crowd density. The originality of this study lies in the integration of multi-stage face processing within a single optimized system, contributing to enhanced reliability in privacy-aware visual analytics applications. The proposed methodology provides a practical solution for real-time face processing and privacy protection in surveillance, security, and social media environments.

2 RELATED WORKS

Eman Hato 2022 [6]. The study employed optical flow estimation to detect abrupt transitions between shots, utilizing a set of videos containing changes in lighting, camera movement, and objects within the scene. The results demonstrated high performance with an F-Score of 0.98, Recall of 0.97, and Precision of 1.00, along with good computational efficiency and an average execution time of approximately 7.88 seconds. Satya Prakash Yadav et al. 2023 [7] analyzed the performance of three object detection algorithms in difficult scenarios to determine the best option in terms of features and accuracy. Based on the findings, all competitors were outperformed by Faster R-CNN since it achieved the best accuracy, recall, precision, and loss.

The study establishes an IoT-guided facial recognition system utilizing the EfficientNet-B4 architecture. Considering the reliability of the model and the limited data, Transfer Learning with fine-tuning was employed. The outcomes showed an accuracy of almost 97% during different light conditions, locations [8].

Proposed a system aimed at protecting the privacy of faces in images. To train the system, used WIDER FACE dataset, one of the most popular datasets in the domain. Face detection was performed using the YOLOv6 algorithm. Privacy protection techniques applied were encryption, decryption, masking, and blurring. Its accuracy reached 0.98 in the training stage and A value of 0.96 was recorded in the testing stage [9].

Proposed a system for the assessment of privacy of digital images by analyzing how much information is leaked, evaluating the risk of identification, and analyzing the protection mechanisms in place [10]. It analyzes single and two stage detection algorithms focusing solely on the YOLO algorithm in all its versions. This study uses YOLOv8 and pose obfuscation techniques to protect identity and performs the test on a custom dataset. Results demonstrate that individual information is effectively concealed while the image maintained intact using YOLOv8 in combination with obfuscation.

3 FACE DETECTION

The classification of images and videos remains an open area for exploration for the computer vision research community, some of the application areas of image and video classification include face recognition, medical image diagnostics, security and surveillance, and remote sensing [11]. The focus of recent research work on image and video analysis has been on the application of deep learning techniques and the utilization of high-resolution images to develop effective decision support systems integrated with Internet of Things (IoT) devices, IoT refers to a network of devices that are physically embedded with software, a network connection, sensors, and electronic devices, which allows the devices to collect data, share information, and interact with each other. The primary objective of any face detection system is to determine whether a face exists in an image and if so, identify and accurately outline the borders of each face so that the detection algorithm can delineate boundaries around all faces [12].

4 VISUAL PRIVACY

Every minute, 72 hours of new video content is uploaded to YouTube. The visual content people capture and share may contain sensitive information. Privacy, in the traditional sense, consists of personal identity, social habits, preferences, inter-personal relationships, and social networks. The advance of technology, especially computer vision (CV), is able to deduce more data from images and videos [13]. This is a double-edged sword as it benefits humanity and increases the risk of a privacy invasion. The goal of visual privacy is to maximize the usability of visual content and minimize the risk posed by possible privacy breaches [14].

5 INTELLIGENT SYSTEM

The development of technology enabling machines to “think,” “learn,” and “reason” makes artificial intelligence, one of the most remarkable modern-day developments. This field generates groundbreaking solutions for multiple domains, including healthcare, education, and manufacturing [15]. With better algorithms and advanced technologies, AI algorithms are deeply rooted in our everyday lives, it is a revolutionary phenomenon in science and technology, reshaping the future and broadening, the scope of imagination and innovation [16].

5.1 Pre-Trained Models

Unlike traditional machine learning methods, deep learning involves multiple interconnected layers, allowing models to learn high-level representations of inputs. In computer vision, deep learning has achieved significant advances in tasks such as object detection, action recognition, motion tracking, semantic segmentation, pose estimation, and image classification [17].

Pre-trained models play a crucial role by transferring knowledge from large datasets, reducing the time and cost of model development, and improving overall accuracy. Most pre-trained models are trained on extensive datasets containing a variety of images, such as objects, tools, and animals [18]. In this research, AlexNet is employed due to its high effectiveness in feature extraction, which enhances face detection and recognition. AlexNet uses pre-trained weights, reducing training time and improving precision, particularly with varied datasets. The network consists of 23 layers, including five convolutional layers and three fully connected layers, and was trained on 1.2 million ImageNet images. It accepts inputs of 227×227×3 and is part of the Deep Learning Toolbox [19]. Figure 1 represent Alexa Net Architecture

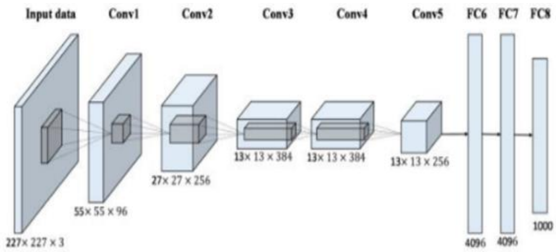

Figure 1: AlexNet architecture.

Figure 1 illustrates the architecture of AlexNet, showing the input layer, the convolutional layers (Conv1–Conv5), and the fully connected layers (FC6–FC8) responsible for classification. The first convolutional layers capture basic visual features, while deeper layers extract more complex patterns. Dimension reduction at certain layers improves processing speed without compromising accuracy.

6 METHODOLOGY

An intelligent system for face detection, recognition, and blurring was created for this research using deep learning techniques. The system was created using a number of successive steps, including data acquisition, data prep, and model selection. The model was trained, and then that model's performance was assessed using predefined criteria so that efficiency and accuracy of the results could be ensured.

The proposed system consists of three main parts which are executed in a linear and structured order. The first part is the acquisition of the necessary the data for the creation of the system, in order to build a dependable and well-defined dataset in multiple conditions. For the second part, a high-accuracy face recognition and identification model is used in a pre-trained neural network to detect and recognize faces. The third part then applies privacy- respecting algorithms to partition faces. Facial blurring is the third and most important part of the proposed system. The goal of this part is to blur the faces of the people who have been detected and identified in the videos in order to protect their privacy and maintain the integrity of the video. Figure 2 shows the proposed system.

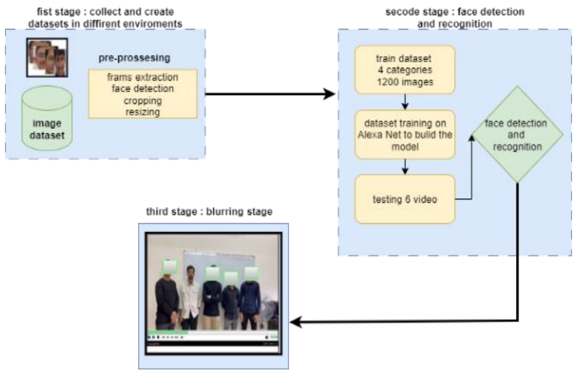

Figure 2: Proposed system.

During the first stage of the experiment, images were captured by videotaping students from the Al-Mustansiriya University. The videos were recorded in different locations, such as in the university's garden and in the corridor of the Computer Science Department, in order to capture different backgrounds and lighting conditions. The videos were processed to remove the audio and the images were extracted to prepare a dataset to be used for the research. The dataset consisted of four subjects, each of whom had 300 images, which amounts to 1200 images in total. The dataset was made available on GitHub [https://github.com/NawarNihad/mustansiriyah-cs-students-dataset-2025.git.] for access, structuring, and for the training and testing phases. The pre-processing stage consisted of important steps in order to enhance the quality of the data. One of the steps was the application of a facial recognition algorithm which enabled the determination of the relevant focus areas. The faces in the images were then cropped to remove the irrelevant parts. The images were adjusted as well in order to meet the requirements of the AlexNet model and to enhance the performance of the model in the following stages, these actions were performed in order to standardize the format of the data, reduce its complexity and enhance the model's performance in the subsequent stages. The samples of the dataset that were created and a code for the first person (001), the second (002), the third (003) and the fourth (004), The dataset was divided into 80% for training and 20% for testing.

In the second phase, the research focuses on face detection and recognition using the pre-trained AlexNet model. First, the video is loaded, and the VideoReader is used to read frames one by one up to a specified maximum or until the end of the video. The Cascade Object Detector is configured to locate faces in each frame, with counters set for the total number of frames processed, the number of detected and categorized faces, and the number of correct categorizations. For each frame of the video: faces are detected, and the total number of detected faces counter is updated. For each detected face: the face is cropped and resized to fit the AlexNet input size The face is labeled and given a confidence value. If the label is in the pre-defined category set (Cf), the face is added to the counter for labeled faces. If the confidence level th) is surpassed, the label is deemed right, and the counter for correct labels is increased. The labeled face and the confidence value are added to the processed frame for the recognition results to be shown while the video is played in the Video Player. This enables a detailed assessment of the detected and recognized faces system, using the pre-trained weights of AlexNet to shorten the time taken for learning and to enhance recognition even in the presence of multiple subjects and varying light conditions and angles. During the third stage of the research, face blurring is performed using the proposed detection-recognition-blurring pipeline. The complete operational procedure, including video frame processing, face detection, CNN-based recognition, confidence evaluation, Gaussian blurring, and result logging, is presented in Algorithm 1. The algorithm starts by importing the input MP4 video file, after which the VideoReader function is used to sequentially read individual frames for further processing.

The Cascade Object Detector is configured to locate faces in each frame, along with the Video Player and its counter settings: the number of processed frames, the number of detected faces, the number of categorized faces, the number of correct categorizations, and the number of blurred faces. Each frame is processed from 1 until the maximum number of frames or the total number of frames in the video. The frames are read, faces are detected (bboxes), and the total number of faces counter is updated.

For each face detected in a frame: the face is cropped and resized to match the AlexNet input size. The face is categorized to obtain a label and a confidence level.

If the label belongs to the set of categories (Cf) with 5 epochs, a learning rate of 0.001 and the confidence threshold (th = 10) is exceeded, Gaussian blur is applied to the face area, and the blurred face counter is updated. The processed frames have face labels and confidence values, which are used to capture the outputs and results to be displayed. The frame is then stored to the output directory.

```
Input:
       Pre-trained CNN model (AlexaNet)
       MP4 video file (Vid)
       Confidence threshold (th)
       Maximum number of frames
(frameLimit)
       Class set (Cf)
Output:
       Processed frames saved in output
folder
       Statistical log file including
recognition, detection, and blurring
accuracies
Steps:
       Load pre-trained CNN model (Alexa
Net).
       Load input video (Vid) using
VideoReader.
       Initialize:
       Face detector ←
vision.CascadeObjectDetector
       Video player ← vision.VideoPlayer
       Counters for processed frames,
detected faces, classified faces, correct
classifications, blurred faces.
       For each frame jf from 1 to
min(frameLimit, totalFrames) do
4.1 Read frame jf from video.
4.2 Detect faces → bboxes.
4.3 Update total face counter.
For each detected face in bboxes:
a. Crop and resize face to CNN input size.
b. Classify face → (label, confidence).
c. If (label ∈ Cf) then
     i. Increment classified face counter.
     ii. If (confidence > th) then
       - Increment correct classifications
counter.
       - Apply Gaussian blur to the face
region.
       - Update blurred face counter.
     End If
d. Annotate frame with label and confidence.
End For
4.4 Display processed frame.
4.5 Save processed frame in output folder.
       Release video player.
       Compute performance metrics:
       Classification Accuracy:
        Detection Accuracy
       blurring Accuracy
       Save results (statistics + logs)
into a text file in the output folder.
End Algorithm.
```

Algorithm 1: CNN-based detection-recognition-blurring pipeline for privacy-preserving video anonymization.

After frame-by-frame processing, the outputs in the logarithmic statistical file in the output directory are updated and capture the results. These include the total values of the classification accuracy, blurring accuracy and detection accuracy. The system evaluation in this stage is the most comprehensive, as it integrates the three distinct components of face detection, face recognition, and blurring. To determine its efficacy, the system has been deployed in multiple settings characterized by varying population sizes. The underlying guiding principles of the proposed algorithm are: i) face detection in video frames, ii) face recognition and classification through feature extraction (analysis) and the iii) privacy protection through blurring. Performing these distinct yet interrelated functions results in the ability to process multiple person situations and also offer improvements in detection and classification accuracy as well as preserve the quality of the images.

7 RESULTS AND DISCUSSION

The system tested on six different videos, each representing a different environment and a varying number of people. The system was applied to blur the faces of different numbers of people in each video to ensure its ability, to handle multiple scenarios. These tests yielded satisfactory results, with the system accurately detecting and recognizing faces and effectively applying blurring to identified individuals. These results demonstrate the system's, ability to operate efficiently in diverse environments, reflecting the effectiveness of combining AlexNet, Cascade Classifier, and Gaussian Blur technology, for privacy protection.

The proposed system was used on one person in order to evaluate the effectiveness of the system in detection, discrimination, and blurring in a given environment. Comprised of videos of this person, the videos were processed using facial detection and discrimination with a pre-trained AlexNet model, and Gaussian blur was applied to the face for privacy protection. The results are presented in the following Table 1.

The table shows the results of testing the proposed system on six different video clips with varying frame rates and a different number of faces in each video. The table shows how recognition accuracy differed among the different videos. In Vid6, the accuracy was the highest recording an accuracy of 94.34%, while Vid2 the accuracy was the lowest recording an accuracy of 30.16%. This difference is reflective of the the number of people present as well as the surroundings of the videos. With respect to the correct classifications and face detection, the system recognized and detected faces reasonably well, especially when the confidence level was more than 0.85. The results of the Vid face detection accuracy test show that Vid2 was the lower end of the spectrum at 3.29% and Vid4 was at the upper end of the spectrum at 25.94%. This shows that more complicated environments along with an increased number of people may lower the face detection accuracy. Additionally, the table shows that the accuracy of the blurring within all the videos was a consistent 100%. This is as a result of the blurring process that depends on the faces that are detected, and that are categorized with a high level of confidence, employing a Gaussian Blur to all of the detected faces. The system is built in a way that the face that was categorized is not the one that gets missed, thus ensuring that every face is blurred, thereby protecting the privacy of all the people that were detected, regardless of the number of people in the video and the atmospheres. This goes to show that the blurring step works in a way that is consistent and separate from the detection and discrimination process, thus achieving a level of result consistency that is 100% in all cases.

The proposed system was applied to two subjects to evaluate the effectiveness for detection, discrimination, and blurring, the results are shown in the next Table 2.

Recognition accuracy varied between the videos, with the highest accuracy recorded in Vid4 at 95.59% and the lowest in Vid2 at 35.82%. The system demonstrated good face detection and classification capabilities, particularly when the confidence level exceeded 0.85. Detection accuracy also varied between the videos, ranging from 17.47% in Vid2 to 55.56% in Vid1. Blurring accuracy, however, was 100% across all videos.

The proposed system was applied to three classes individuals to test the effectiveness of detection, discrimination, and blurring; the results are shown in the following Table 3.

Table 1: Results of detection-recognition and blurring for one class (001).

Video	Number of frames	Faces detection	Correct Classifications (th>0.85)	Recognition Accuracy	Detection Accuracy	Percentage of Detected Faces Correctly Blurred
Vid1	423	100	100	100.00%	23.64%	100.00%
Vid2	578	63	19	30.16%	3.29%	100.00%
Vid3	485	153	111	72.55%	22.89%	100.00%
Vid4	374	104	97	93.27%	25.94%	100.00%
Vid5	772	65	61	93.85%	7.90%	100.00%
Vid6	500	106	100	94.34%	20.00%	100.00%

Table 2: Results of detection-recognition, and blurring results for two classes of faces (001,002).

Video	Number of frames	Faces detection	Correct Classifications (th>0.85)	Recognition Accuracy	Detection Accuracy	Percentage of Detected Faces Correctly Blurred
Vid1	423	250	235	94.00%	55.56%	100.00%
Vid2	578	282	101	35.82%	17.47%	100.00%
Vid3	485	350	233	66.57%	48.04%	100.00%
Vid4	374	204	195	95.59%	52.14%	100.00%
Vid5	772	514	380	73.93%	49.22%	100.00%
Vid6	500	263	207	78.71%	41.40%	100.00%

Table 3: Results of detection, recognition and blurring for three classes (001, 002, 003).

Video	Number of frames	Faces detection	Correct Classifications (th>0.85)	Recognition Accuracy	Detection Accuracy	Percentage of Detected Faces Correctly Blurred
Vid1	423	351	335	95.44%	79.20%	100.00%
Vid2 (Does not contain C)	578	282	42	14.89%	7.27%	100.00%
Vid3	485	450	305	67.78%	62.89%	100.00%
Vid4	374	274	261	95.26%	69.79%	100.00%
Vid5	772	616	478	77.60%	61.92%	100.00%
Vid6 (Does not contain C)	500	285	199	69.82%	39.80%	100.00%

Table 4: Results of detection, recognition and blurring for four classes (001, 002, 003and 004).

Video	Number of frames	Faces detection	Correct Classifications (th>0.85)	recognition Accuracy	Detection Accuracy	Percentage of Detected Faces Correctly Blurred
Vid1	423	423	407	96.22%	96.22%	100.00%
Vid2(Does not contain C)	578	578	94	16.26%	16.26%	100.00%
Vid3	485	485	344	70.93%	70.93%	100.00%
Vid4	374	374	361	96.52%	96.52%	100.00%
Vid5	772	772	429	55.57%	55.57%	100.00%
Vid6(Does not contain C)	500	500	281	56.20%	56.20%	100.00%

The results show that recognition accuracy varied across the videos, with the highest accuracy recorded in Vid1 at 95.44% and Vid4 at 95.26%, while the lowest accuracy was in Vid2 (which does not contain category C) at 14.89%. Regarding face detection and correct classifications, the system demonstrated good face detection and recognition capabilities, although performance decreased in videos lacking category C. Detection accuracy also varied, ranging from 7.27% in Vid2 to 79.20% in Vid1, reflecting the impact of the number of people and the presence of the target category on the system's accuracy. Blurring accuracy remained constant at 100% across all videos.

The proposed system was applied to three classes individuals to test the effectiveness of detection, discrimination, and blurring; the results are shown in the following Table 4.

The results show that the system was able to detect all faces in some videos, such as Vid1, Vid3, Vid4, Vid5, and Vid6, with the number of detected faces nearly equaling the number of frames in the video. Recognition accuracy was highest in Vid4 at 96.52% and Vid1 at 96.22%, while it was lowest in Vid5 at 55.57% and Vid2 (which does not contain category C) at 16.26%. Detection accuracy showed a similar pattern, ranging from 16.26% in Vid2 to 96.22% in Vid1, reflecting the impact of the target audience and the number of people in the video on the system's performance. Blurring accuracy remained constant at 100% in all videos, demonstrating the system's effectiveness, in protecting the privacy of detected individuals.

It was noted that detection accuracy was low if the videos focused on one person, even if the videos contained many people. Focusing on a single face lead to inaccuracy because detection, as well as discrimination, are performed on the selected viewed faces only. That, in consequence, leads to less accurate processing than if all faces are regarded as a whole. With the increase in the number of faces in focus, the system is provided with higher data volume, thus significantly improving detection and classification accuracy. With regard to the blurring stage, it achieved 100% accuracy for all videos. Gaussian blur, irrespective of the number of people in the video, is applied to all faces with high confidence detection and classification. This proves that the blurring stage works independently and reliably after the detection and discrimination stages, which provides high efficiency in the safeguarding the privacy of people detected.

8 CONCLUSIONS

This study presented an intelligent and automated AI-based system for multi-face detection, recognition, and privacy-preserving blurring in real-time video environments. The proposed framework successfully integrates multiple face processing stages into a unified architecture capable of handling complex and dynamic scenarios. Experimental results demonstrated high recognition accuracy and consistent classification performance across videos containing varying numbers of individuals and diverse environmental conditions.

The system maintained stable performance under challenges such as lighting variations, different camera angles, and changes in crowd density. Notably, recognition accuracy exceeded 90% in complex scenes, while the automated blurring module achieved a 100% success rate in ensuring privacy protection across all tested samples. These results confirm the robustness and reliability of the proposed framework in real-world applications.

The main contribution of this work lies in combining detection, recognition, and privacy-aware blurring within a single intelligent system designed for real-time deployment. This approach provides an effective solution for video surveillance, security monitoring, and privacy-sensitive analysis tasks. Future work may focus on improving recognition algorithms, expanding the dataset to include more diverse and challenging scenarios, and incorporating advanced classification strategies to enhance model generalization, computational efficiency, and overall system performance.

REFERENCES

[1] S. Ahern, D. Eckles, N. S. Good, S. King, M. Naaman, and R. Nair, “Over-exposed?: Privacy patterns and considerations in online and mobile photo sharing,” in Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI ’07), New York, NY, USA, 2007, pp. 357-366.

[2] J. M. Such, J. Porter, S. Preibusch, and A. Joinson, “Photo privacy conflicts in social media: A large-scale empirical study,” in Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI ’17), New York, NY, USA, 2017, pp. 3821-3832.

[3] S. R. Sahoo and B. B. Gupta, “Fake profile detection in multimedia big data on online social networks,” Int. J. Inf. and Computer Security, vol. 12, no. 2-3, pp. 303-331, 2020.

[4] K. Jain, S. R. Sahoo, and J. Kaubiyal, “Online social networks security and privacy: Comprehensive review and analysis,” Complex & Intelligent Systems, vol. 7, no. 5, pp. 2157-2177, 2021, [Online]. Available: https://doi.org/10.1007/s40747-021-00409-7.

[5] E. Meunier and P. Bouthemy, “Unsupervised Space-Time Network for Temporally-Consistent Segmentation of Multiple Motions,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), June 2023, pp. 22139-22148, doi: 10.1109/CVPR52729.2023.02120.

[6] S. P. Yadav, M. Jindal, P. Rani, V. H. C. de Albuquerque, C. d. S. Nascimento, and M. Kumar, “An improved deep learning-based optimal object detection system from images,” Multimedia Tools and Applications, 2023, [Online]. Available: https://doi.org/10.1007/s11042-023-16736-5.

[7] H. H. Ali, J. R. Naif, and W. R. Humood, “A new smart home intruder detection system based on deep learning,” Al-Mustansiriyah Journal of Science, vol. 34, no. 2, pp. 60-69, 2023, [Online]. Available: https://doi.org/10.23851/mjs.v34i2.1267.

[8] M. H. AbdAlkreem, R. S. Salman, and F. K. Al-Jibory, “Detect people’s faces and protect them by providing high privacy based on deep learning,” Tehnički Glasnik [Technical Gazette], vol. 18, no. 1, pp. 92-99, 2024, [Online]. Available: https://doi.org/10.31803/tg-20231210183347.

[9] S. Raees and M. S. H. Al-Tamimi, “The role of artificial intelligence in providing people with privacy: Survey,” Journal of Applied Engineering and Technological Science, vol. 5, no. 2, pp. 813-829, 2024, [Online]. Available: https://doi.org/10.37385/jaets.v5i2.4013.

[10] R. Plaud and J.-L. Lisani, “Two Deep Learning Solutions for Automatic Blurring of Faces in Videos,” arXiv preprint arXiv:2409.14828, 2024, [Online]. Available: https://doi.org/10.48550/arXiv.2409.14828.

[11] M. Zamir et al., “Face detection and recognition from images and videos based on CNN and Raspberry Pi,” Computation, vol. 10, no. 9, art. no. 148, 2022, [Online]. Available: https://doi.org/10.3390/computation10090148.

[12] M. Broz, “Number of photos (2023): Statistics, facts, and predictions,” 2023, [Online]. Available: https://www.phototutorial.com/number-of-photos/.

[13] T. Aichner, M. Grünfelder, O. Maurer, and D. Jegeni, “Twenty-five years of social media: A review of social media applications and definitions from 1994 to 2019,” Cyberpsychology, Behavior, and Social Networking, vol. 24, no. 4, pp. 215-222, 2021, [Online]. Available: https://doi.org/10.1089/cyber.2020.0134.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2012, pp. 1097-1105.

[15] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[16] Sharma and S. Kaushik, “A review of convolutional neural networks in computer vision,” Artificial Intelligence Review, Springer, 2024.

[17] X. Han, Y. Zhong, L. Cao, and L. Zhang, “Pre-trained AlexNet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification,” Remote Sensing, vol. 9, no. 8, 2017, [Online]. Available: https://doi.org/10.3390/rs9080848.

[18] MathWorks, “AlexNet Convolutional Neural Network,” Deep Learning Toolbox, MathWorks, 2024.

[19] P. Purwono, A. Ma’arif, W. Rahmaniar, H. I. K. Fathurrahman, A. Z. K. Frisky, and Q. M. ul Haq, “Understanding of Convolutional Neural Network (CNN): A Review,” International Journal of Robotics and Control Systems, vol. 2, no. 4, pp. 739-748, 2022, [Online]. Available: https://doi.org/10.31763/ijrcs.v2i4.888.