

## PREVENTING CHILD KIDNAPING AT HOME USING CCTV THAT UTILIZES FACE RECOGNITION WITH YOU ONLY LOOK ONCE (YOLO) ALGORITHM

Joniwan<sup>1\*</sup>, Gerry Firmasyah<sup>2\*\*</sup>, Agung Mulyo Widodo<sup>3\*\*\*</sup>, Budi Tjahjono<sup>4\*\*\*\*</sup>

<sup>1,2,3,4</sup>Universitas Esa Unggul

\* [Joniwan21041989@gmail.com](mailto:Joniwan21041989@gmail.com) \*\* [gerry@esaunggul.ac.id](mailto:gerry@esaunggul.ac.id) \*\*\* [agung.mulyo@esaunggul.ac.id](mailto:agung.mulyo@esaunggul.ac.id)

\*\*\*\* [budi.tjahjono@esaunggul.ac.id](mailto:budi.tjahjono@esaunggul.ac.id)

### ABSTRACT

The rise of child abductions often occurs because of weak parental supervision and weak vigilance of strangers who intentionally or unintentionally approach children. Early warning when a child is approached by an unknown person can cancel the intention of the kidnapper. Indeed, CCTV can be used as a tool to thwart crime. Face Recognition is one of the best biometric recognition systems for identifying someone whose body or DNA has certain characteristics that distinguish one person from another. This research resulted in the right collaboration between CCTV which is equipped with the YOLOv8 algorithm which can detect a human object in real-time and also utilizes face recognition which can prevent child abduction by providing an early warning to kidnappers when approaching children. The results of the first test are the optimal combination using 3 cameras and the CNN method and the shooting time in the afternoon. This combination succeeded in providing an early warning in the form of a siren whose sound is issued with the help of a loudspeaker when a child is left alone and approached by an unknown person with the distance between the camera and the object to be detected as far as 1 meter to 3 meters which takes 15 seconds to 22 seconds to complete the detection process in an image. Then the results of further testing with 6 cameras using the LBPH method take between 17 seconds to 18 seconds to complete the detection process in an image, SVM takes 20 seconds to 42 seconds and CNN takes 17 seconds to 24 seconds.

**Keywords:** *child abduction, CCTV, face recognition, YOLOv8, CNN*

This article is licensed under [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/) 

### INTRODUCTION

Children are a gift from God Almighty which affects the dignity of the human being as a whole. Every child born does not escape the duties and responsibilities of their parents. Children have human dignity that must be upheld so that every child born has rights that the child does not have to ask for. Parents must protect their children. Children are not able to protect themselves from various actions that can harm children. Children are very vulnerable to physical and mental trauma that continues throughout the child's life. (Rika Saraswati, 2015). Abduction cases have increased recently, citing data from the Indonesian police, in 2022 there were 233 cases of kidnapping and as many as 28 cases were children who were victims (Polri, 2023). This kidnapping case was also based on several motives such as the desire to have children, hostage-taking, sexuality (Siregar, 2022), harassment, sale of organs, human trafficking (Aidil, 2023), and so on.

Generally, at an early age children are very vulnerable to this. The risk of kidnapping exists for several reasons, such as a lack of parental protection and a lack of supervision, and a lack of vigilance for strangers approaching children. The crime of kidnapping a child is an act that is intentional or unintentional, planned or not planned by the perpetrator, which is carried out not only because there is the intention of the perpetrator, but also because there is an opportunity. Many technologies are deployed to deal with cases of child abductions, but most

of these technologies can only help search for children when kidnappings have occurred, for example by utilizing GPS (Global Positioning System) technology attached to children (Zulfikar Sembiring, 2019), using cellphone cameras to recording the incident and sending it to the closest person and also utilizing CCTV (Closed Circuit Television) that is crossed by children to detect the current whereabouts of the kidnappers and where the child is being held, hostage. In a study conducted by Akira Miyahara and Itaru Nagayama entitled "An Intelligent Security Camera System for Kidnapping Detection" (Nagayama, 2013) new security cameras can detect kidnapping when someone enters the car not of their own free will, which indicates an abduction is taking place occurs and the next step is only taken when the victim is in the hands of the perpetrator. Utilization of GPS is also carried out by creating a smartwatch that is connected to the Firebase real-time database by sending latitude and longitude positions to see the child's position in the map as has been researched by Bali State Polytechnic Students (I N E Indrayana, 2019) whose application can be carried out when this occurs. kidnapping of children. And CCTV is often used as accurate evidence when cases of theft or kidnapping occur.

Based on the available data, it can be concluded that the current technology has not been used to prevent kidnappings, but rather the steps that must be taken in the event of a kidnapping. Of course, the nature of preventing kidnapping is better than having to experience abduction first, of course by increasing supervision of children's activities when the child is left alone. (Hasna, 2023) The purpose of this research is to find a combination of CCTV technology and artificial intelligence as a solution that can assist parents in supervising children who are left alone, continuous supervision is not carried out by parents who cannot leave their activities. Today's CCTV has implemented a more modern system, namely the Internet Protocol Camera which is a further development of analog CCTV technology which can automatically transfer data to be spread over the internet using a predetermined IP address (Ajie Pambudhi, 2017). And if the CCTV is implanted with face recognition technology which is the result of the deep learning branch of artificial intelligence, then the CCTV will become a smart CCTV that can recognize faces and objects caught by the camera. This research will combine CCTV with face recognition so that it can recognize the face of a child who wants to be supervised when he is alone in a room. And in this research, several algorithms will be added so that this CCTV can recognize familiar people who are near children and unknown people who try to approach children and can give warnings to unknown people where the researcher hopes that this warning can deter evil intentions from these unknown persons so that child abductions can be prevented.

## **METHODS**

### **Definition and History of CCTV**

CCTV (Closed Circuit Television) is a recording device that uses one or more video cameras and produces video or audio data (Deka Rakhmadhani, 2018). CCTV is the use of video cameras that transmit signals or broadcasts aimed at a certain scope of equipment, namely a specific limited set of monitors. CCTV technology has a function to monitor a place from another place. The CCTV system was invented by Walter Bruch and was first used in 1942 in Peenemünde, Germany. Until the 1990s, ATMs were usually equipped with CCTV systems. CCTV camera installation is also widely used in streets, parks, crime areas, and private homes.

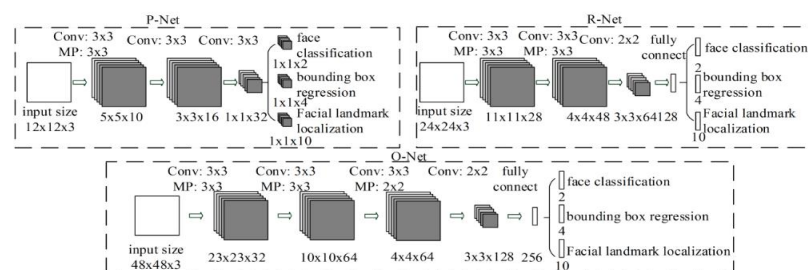
The development of CCTV systems is becoming more modern every year and the market offers a variety of technologies. Now video surveillance can be used as evidence of a crime, vehicle traffic monitoring, and more. (Eric L. Pizza, 2019).

**Face Recognition**

Face Recognition is an individual identification process based on unique facial characteristics. This technology has developed rapidly in recent years and has been applied in areas such as security, surveillance, and identity recognition. Face Recognition is one of the best biometric recognition systems for identifying someone whose body or DNA has certain characteristics that distinguish one person from another. Face recognition can be done using digital image processing techniques and pattern recognition algorithms (Mei Wang, 2020). Facial recognition techniques with feature-based methods will look for unique facial features such as eyes, nose, and lips, and then compare these features with known facial databases (Dong Chen, 2012).

**Deep learning**

Deep learning is a machine learning technique that uses artificial neural networks to learn complex and abstract features from data. The deep learning technique most commonly used in facial recognition is the Convolutional Neural Network (CNN) (Noviana Dewi, 2021).



**Fig. 1.** MTCNN (Multi-Task Cascaded Convolutional Neural Networks)

MTCNN or Multi-Task Cascaded Convolutional Neural Networks are neural networks used to detect faces and tag faces in images. MTCNN was published in 2016 by Zhang et al (Kaipeng Zhang, 2016). MTCNN is a multi-step framework that exploits the inherent correlation between face detection and alignment to improve its performance, each neural network improves recognition results by passing its input through CNN, followed by Non-Maximum Smoothing (NMS), a method that reduces the number of bounding boxes. (Mühler, 2018). Consisting of multi-task cascaded CNNs (MTCNNs) that function to find mouths and eyes, as well as partial facial images to extract static features, then CNN will complexly separate groups that are not part of the face and refine the results and display the position of facial landmarks that can significantly improve the learning process. The "face recognition" library was developed by Adam Geitgey. The "face recognition" library uses deep learning-based facial recognition methods. More specifically, the method used is the Convolutional Neural Network (CNN) which allows for high-accuracy face detection, feature extraction, and face comparison.

**You Only Look Once (YOLO)**

YOLO is currently the most popular real-time object detector according to (Haitong Lou, 2023) the reason that makes it polar is its network architecture which is very lightweight, and efficiently offers aggregation methods and more accurate observation results. The latest version

of Yolo according to (Glenn Jocher (14), 2023) is YOLOv8 which builds on advances in deep learning and computer vision, offering unparalleled performance in terms of speed and accuracy. Its sleek design makes it suitable for a wide range of applications and easily adaptable to a variety of hardware platforms, from edge devices to cloud APIs. YOLOv8 is the latest version of YOLO from Ultralytics. As a SOTA (state-of-the-art, state-of-the-art) model, YOLOv8 builds on the successes of previous versions and introduces new features and enhancements to improve performance, flexibility, and efficiency. YOLOv8 supports all visual intelligence tasks including recognition, segmentation, pose estimation, tracking, and classification.

### **OpenCV**

Open Source Computer Vision Library, also known as OpenCV, is an open-source computer vision and machine learning software library that provides a common infrastructure for computer vision applications and accelerates the use of machine perception in commercial products. OpenCV can run on the C ++, Python, Java, and MATLAB platforms (team, 2023) and supports Windows, Linux, Android, and Mac OS Operating Systems (Jose Sigut, 2020). OpenCV is real-time computer vision that utilizes GPU acceleration to operate programming functions in real-time that are used for collaborative development, simulation, testing, and implementation of image processing and computer vision-based systems.

### **Local Binary Patterns Histogram (LBPH)**

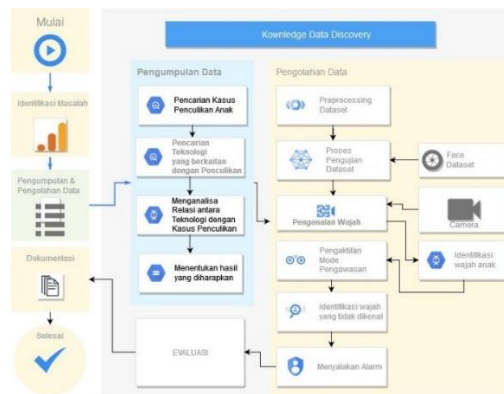
LBPH is a method for feature extraction from digital images used in face recognition and pattern recognition. This method combines the concepts of Local Binary Patterns (LBP) and histograms to produce recognizable facial descriptions. In LBPH, LBP is used to extract features from each pixel in the face image, then the histogram is calculated for each pixel block (Muhammad Ilal Nabsi, 2021). This histogram is then used to represent the face as a feature vector.

### **Support vector machine (SVM)**

SVMs are especially useful for face recognition because they can accurately classify and regress complex data sets, such as feature faces, using non-linear boundaries. (Mishra, Dash, & Tiwari, 2023) The main principle of SVM is to find the best hyperplane that can separate two classes of data with the maximum distance. SVM works by creating a separator between data in the feature space by maximizing the margin between the two classes. SVM can also use kernel functions to transform the feature space so that data that is not linearly separated can be hyperplane separated. SVM works by finding the best hyperplane that separates two classes of data with the maximum distance. The hyperplane will be found by maximizing the margin, namely the distance between the hyperplane and the closest training points of each class. If the data cannot be separated linearly, SVM can use kernel transformations to project the data into a different feature space. After the model is trained, SVM can be used to predict classes from new data. The new data will be projected into the same feature space (if using the kernel) and then analyzed by the model to predict its class.

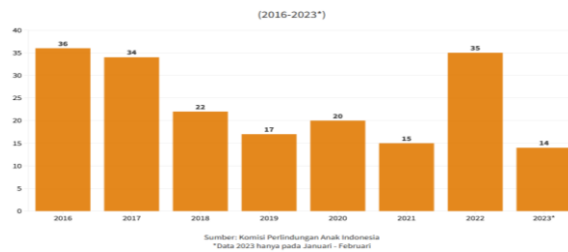
## **RESULTS AND DISCUSSION**

**Steps Study**



**Fig. 2.** Research Design

Identification of problems regarding how artificial intelligence technology can be useful for society in real implementation to reduce the level of criminality. Collect data by searching for child abduction cases. The following is information on the number of child abduction cases that have occurred in Indonesia by year (Pratiwi, 2023).



**Fig. 3.** Kidnapping case children that happened in Indonesia

After searching for technology related to kidnapping at the previous stage, researchers found that technology was utilized when the kidnapping event had already occurred. So the researchers hope to be able to find a combination of existing technologies to be used as a tool to prevent kidnappings, especially kidnappings of children that occur in the yard. This is where the merger of this technology is expected to cancel the intention of the kidnapping itself.

**Preprocessing and Dataset Training Testing**

Facial datasets with method collect the face dataset to be made as source training.

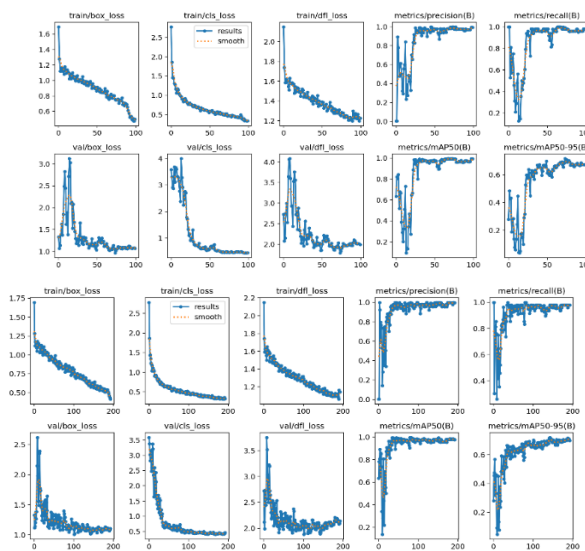
Age	On version	Number Of Datasets
0 – 18 years	No Mask	149 Images
19 – 65 years		149 Images
66 years and over		149 Images
0 – 18 years	With Mask	172 Images
19 – 65 years		151 Images
66 years and over		147 Images

Table 1. The faces that will be used as training are taken from the Kaggle dataset which consists of 918 faces of various ages and a combination of faces wearing masks and faces without masks. With the composition of the dataset division for training 70% of as many as 642 faces and for testing 30% of as many as 276 faces

Epoch	Hours	Results			
		<i>mAP0.5</i>	<i>mAP0.5:0.95</i>	<i>P</i>	<i>R</i>
100	0.118	0.995	0.714	1	0993
200	0.248	0.976	0.717	0.959	0979

Table 2. Process of testing the dataset train using the Python-3.10.12 programming language which runs on the Google Colab platform by utilizing the Ultralytics YOLOv8.0.135 library and torch-2.0.1+cu118 and activating hardware accelerators with the Tesla Graphics Processing Unit (GPU) type T4, 15102MiB and tested it with 2 different epochs, namely 100 epochs and 200 epochs to see the difference in the results of the two epoch parameters with the most efficient time. The number of epochs cannot be determined with certainty, the number of epochs will always be different for each dataset, therefore tests must be carried out to find out how many epochs are suitable for the face dataset.

The first test with 100 epochs was completed in 0.118 hours with a P of 1.000, R of 0.993, mAP50 of 0.995, and mAP50-95 of 0.714. The second test with 200 epochs was completed in 0.248 hours with a P of 0.959, R of 0.979, mAP50 of 0.976, and mAP50-95 of 0.717.



**Fig. 4.** Trains 100 & 200 epochs

### Testing Real-Time Datasets

Object detection experiments were carried out 5 times using an IP camera and only 3 which were between 1 to 3 meters away could be encoded. The results of the confidence detected as a human object within 1 meter are 0.9347 with an inference of 275.3ms, a speed of 6.0ms preprocess, and 5.0ms post-process. Meanwhile, for a distance of 2 meters, a confidence of 0.9366 was found with a speed of 4.0ms preprocess, 283.2ms inference, and 5.0ms post-



However, the tolerance value of 0.55 which has not become the basis is reviewed by conducting several tests on the face class taken by sampling by increasing the tolerance value to the highest number at 0.99 and analyzing the best tolerance level of the lowest and highest values in each dataset class. Testing of face sampling as many as 46 faces displayed in front of the camera for 5 seconds to record the lowest tolerance value and the highest tolerance value which will later be used as the basis for the tolerance value in the next coding.

Class	Faces	Result Face Distance		
		Min	Average	Max
Child supervised	2	0.2988	0.3633	0.3843
Known people	12	0.3784	0.4566	0.5828
Stranger	32	0.4124	0.6401	0.8922

Table 3. The results of the distance from the sampling of the 46 faces above for the data set for the class of unknown people are quite significant between the min and the max because the sampling of unknown people is taken from various faces of various ages from small children to the elderly group, and the lowest distance results obtained from the comparison between the faces of the small children and the faces of the children being watched, while the highest distance is obtained from the comparison between the faces of the parents and the faces of the children being watched. From the table above, it can be concluded that the highest average comparison of distances between classes in the dataset (supervised class of children and known class of people) is 0.4566 with an error rate of 0.0442 which is the result of a comparison of the average dataset class of 0.4566 minus the smallest value of the class of people unknown at 0.4566. However, it is undeniable that the face comparison distance tolerance figure of 0.4566 is not the most optimal number because this figure will continue to change according to supporting components such as light intensity, sun, and light conditions, and the level of distance between the face and the camera, therefore the number 0.4566 will only be used as the default number given by the system when face\_distance tolerance setting is not done. There is a function in the system that will take a tolerance number according to the actual conditions when the child is within a radius to be supervised. The system will activate facial recognition and take the largest calculated distance between the faces in the current condition and the faces in the dataset and store them in the database for reuse. The tolerance figure is the number that is most likely to be used in the current conditions.

The researcher also compared the number of cameras used. The first comparison was made using the object detection process using the paralytics YoloV8 library, the You Only Look Once method, and the facial recognition process using the face\_recognition library, the Convolutional Neural Network method. comparison using camera 1, camera 2, and camera 3 was carried out during the day with high light intensity compared to the afternoon with low light intensity with comparisons of present and unsupervised children who were approached by unknown persons, as well as a comparison of the distance of the camera to the object of the child being supervised with a variation of 1-meter distance, 2-meter distance, and 3-meter distance. The use of 3 cameras shows a true positive trend in lighting conditions that are not

**Preventing Child Kidnaping at Home Using CCTV that Utilizes Face Recognition with You Only Look Once (YOLO) Algorithm**

too high for the detection of children's faces to be supervised both within 1 meter of the camera and up to 3 meters.

Confusion Matrix				
Day Light	1 Meter	2 Meter	3 Meter	Time
<b>Unsupervised Child</b>				
1 Camera	True Positive (TP)	True Positive (TP)	False Negative (FN)	3-5 seconds
2 Camera	True Positive (TP)	True Positive (TP)	False Negative (FN)	13-18 seconds
3 Camera	True Positive (TP)	True Positive (TP)	False Negative (FN)	15-22 seconds
<b>Child With Supervision</b>				
1 Camera	False Positive (FP)	False Positive (FP)	False Positive (FP)	3-6 seconds
2 Camera	True Negative (TN)	False Positive (FP)	False Positive (FP)	15-24 seconds
3 Camera	True Negative (TN)	True Negative (TN)	False Positive (FP)	14 - 40 seconds
Night Light	1 Meter	2 Meter	3 Meter	Time
<b>Unsupervised Child</b>				
1 Camera	True Positive (TP)	True Positive (TP)	False Negative (FN)	4-7 seconds
2 Camera	True Positive (TP)	True Positive (TP)	False Negative (FN)	8-11 seconds
3 Camera	True Positive (TP)	True Positive (TP)	True Positive (TP)	15-22 seconds
<b>Child With Supervision</b>				
1 Camera	False Positive (FP)	False Positive (FP)	False Positive (FP)	4-6 seconds
2 Camera	True Negative (TN)	True Negative (TN)	False Positive (FP)	8-11 seconds
3 Camera	True Negative (TN)	True Negative (TN)	False Positive (FP)	14 - 40 seconds

**Fig. 7. Confusion Matrix**

To see how many cameras can capture the faces of children who will be supervised at the same time, a second comparison is made which is carried out in a closed room measuring 3 meters x 3 meters with lighting using a 12-watt LED lamp and using camera 1, camera 2, camera 3, camera 4, camera 5 and camera 6 for taking pictures of children to be supervised. Comparisons were made with the object detection process using the paralytics YoloV8 library, the You Only Look Once method, and the facial recognition process using 3 methods to be compared, namely the face\_recognition library, the Convolutional Neural Network method, the CV2 library, the Local Binary Patterns Histogram (LBPH) method, and also the Support vector machine (SVM).

Shoot	CNN			SVM			LBPH		
	Detection	TimeStamp	Distance	Detection	TimeStamp	Distance	Detection	TimeStamp	Distance
1	TRUE	1691885602		TRUE	1691883884		TRUE	1691905279	
2	TRUE	1691885624	22	TRUE	1691883904	20	TRUE	1691905295	16
3	TRUE	1691885640	16	TRUE	1691883878	14	TRUE	1691905308	13
4	FALSE	1691885656	16	FALSE	1691883892	14	TRUE	1691905321	13
5	TRUE	1691885672	16	TRUE	1691883905	12	TRUE	1691905333	12
6	TRUE	1691885688	16	TRUE	1691883921	16	FALSE	1691905346	13
7	TRUE	1691885706	18	TRUE	1691883935	14	FALSE	1691905358	12
8	TRUE	1691885723	17	TRUE	1691883949	14	FALSE	1691905370	12
9	FALSE	1691885739	16	FALSE	1691883963	14	FALSE	1691905383	13
10	TRUE	1691885754	15	TRUE	1691883976	13	FALSE	1691905395	12
11	TRUE	1691885771	17	TRUE	1691883991	15	FALSE	1691905408	13
12	FALSE	1691885787	16	FALSE	1691884005	14	TRUE	1691905420	12
13	TRUE	1691885802	15	TRUE	1691884018	14	TRUE	1691905433	13
14	TRUE	1691885818	17	TRUE	1691884033	14	TRUE	1691905446	13
15	TRUE	1691885836	17	TRUE	1691884047	14	TRUE	1691905461	15
16	TRUE	1691885853	17	TRUE	1691884062	15	TRUE	1691905475	14
17	TRUE	1691885868	15	TRUE	1691884076	14	TRUE	1691905487	12
18	FALSE	1691885884	16	FALSE	1691884090	14	TRUE	1691905500	13
19	FALSE	1691885900	16	TRUE	1691884104	14	FALSE	1691905513	13
20	FALSE	1691885915	15	FALSE	1691884118	14	FALSE	1691905528	15
21	TRUE	1691885930	15	TRUE	1691884132	14	TRUE	1691905540	12
22	TRUE	1691885947	17	TRUE	1691884147	15	TRUE	1691905553	13
23	FALSE	1691885962	15	FALSE	1691884161	14	FALSE	1691905567	14
24	TRUE	1691885977	15	TRUE	1691884174	13	TRUE	1691905579	12
25	FALSE	1691885992	15	FALSE	1691884188	14	TRUE	1691905592	13
26	TRUE	1691886007	15	TRUE	1691884201	13	TRUE	1691905605	13
27	TRUE	1691886023	16	TRUE	1691884216	15	TRUE	1691905618	13
28	TRUE	1691886039	16	TRUE	1691884230	14	FALSE	1691905632	14
29	TRUE	1691886055	16	TRUE	1691884246	16	TRUE	1691905646	14
30	TRUE	1691886072	17	TRUE	1691884260	14	TRUE	1691905661	15
31	FALSE	1691886088	16	FALSE	1691884274	14	FALSE	1691905674	13
32	FALSE	1691886103	15	FALSE	1691884288	14	FALSE	1691905688	14
33	FALSE	1691886117	14	TRUE	1691884301	13	FALSE	1691905701	13
34	FALSE	1691886133	16	FALSE	1691884315	14	FALSE	1691905713	12
35	FALSE	1691886147	14	TRUE	1691884329	14	TRUE	1691905726	13
36	TRUE	1691886162	15	TRUE	1691884343	14	FALSE	1691905739	13
37	FALSE	1691886177	15	FALSE	1691884357	14	FALSE	1691905751	12
38	FALSE	1691886191	14	FALSE	1691884371	14	FALSE	1691905765	14
39	FALSE	1691886205	14	FALSE	1691884385	14	TRUE	1691905777	12

**Fig. 8. Analyze using 4 Cameras and 3 different Methods**

Shoot	CNN			SVM			LBPH		
	Detection	TimeStamp	Distance	Detection	TimeStamp	Distance	Detection	TimeStamp	Distance
1	TRUE	1691889090		TRUE	1691889009		TRUE	1691906062	
2	TRUE	1691889113	23	TRUE	1691889033	24	TRUE	1691906078	16
3	TRUE	1691889130	17	TRUE	1691889055	22	TRUE	1691906093	15
4	FALSE	1691889146	16	FALSE	1691889075	20	TRUE	1691906109	16
5	TRUE	1691889162	16	TRUE	1691889094	19	TRUE	1691906125	16
6	TRUE	1691889178	16	TRUE	1691889115	21	FALSE	1691906139	14
7	TRUE	1691889194	16	TRUE	1691889136	21	TRUE	1691906154	15
8	TRUE	1691889210	16	TRUE	1691889156	20	FALSE	1691906169	15
9	FALSE	1691889227	17	FALSE	1691889177	21	TRUE	1691906183	14
10	TRUE	1691889241	14	TRUE	1691889197	20	FALSE	1691906198	15
11	TRUE	1691889258	17						

**Fig. 9. Analyze using 5 Cameras and 3 different Methods**

Shoot	CNN			SVM			LBPH		
	Detection	TimeStamp	Distance	Detection	TimeStamp	Distance	Detection	TimeStamp	Distance
1	TRUE	1691889335		TRUE	1691889307		TRUE	1691906324	
2	TRUE	1691889350	24	TRUE	1691889340	42	TRUE	1691906341	17
3	TRUE	1691889378	19	TRUE	1691889370	21	TRUE	1691906358	17
4	TRUE	1691889397	19	FALSE	1691889398	28	TRUE	1691906375	17
5	TRUE	1691889414	17	TRUE	1691889419	21	TRUE	1691906392	17
6	TRUE	1691889433	19	TRUE	1691889443	22	FALSE	1691906410	18
7	TRUE	1691889452	19	TRUE	1691889461	20	TRUE	1691906427	17
8	TRUE	1691889470	18	TRUE	1691889482	21	TRUE	1691906445	18
9	TRUE	1691889489	19						

**Fig. 10.** Analyze using 6 Cameras and 3 different Methods

Method	Shoot	RESULT				Time	
		True	False	True %	False %		
LBPH	4 Camera	39	22	17	56.41%	43.59%	12 - 16 seconds
	5 Camera	10	7	3	70.00%	30.00%	14 - 16 seconds
	6 Camera	8	7	1	87.50%	12.50%	17 - 18 seconds
SVM	4 Camera	39	26	13	66.67%	33.33%	13 - 20 seconds
	5 Camera	10	8	2	80.00%	20.00%	19 - 24 seconds
	6 Camera	8	7	1	87.50%	12.50%	20 - 42 seconds
CNN	4 Camera	39	26	13	66.67%	33.33%	14 - 22 seconds
	5 Camera	10	8	2	80.00%	20.00%	14 - 23 seconds
	6 Camera	9	9	0	100.00%	0.00%	17 - 24 second

**Fig. 11.** Summary Comparison using 6 Cameras and 3 different Methods

The results of the comparison show that at the limit of using up to 6 cameras, there are still system errors in detecting children's faces that must be supervised using the Local Binary Patterns Histogram (LBPH) method and also the Support vector machine (SVM). In the details above, it can be seen that the system took pictures simultaneously with 6 cameras 8 times and 7 of them the faces of the children being watched could be detected and there was 1 time where the two methods failed to detect the child's face in the captured image. However, the Convolutional Neural Network (CNN) method is capable of perfectly detecting children's faces using 6 cameras that take 9 images simultaneously and perform child face detection in these images. Unfortunately, this method is ranked second in terms of processing speed. Local Binary Patterns Histogram (LBPH) has the best processing speed whereas in capturing images using 6 cameras it only takes between 17 seconds to 18 seconds to complete the detection process in an image. Meanwhile, the Convolutional Neural Network (CNN) takes 17 seconds to 24 seconds to complete the detection process on an image taken with 6 cameras. And the Support vector machine (SVM) method is the method that requires the most time to complete, which is around 20 seconds to 42 seconds.

### Face Recognition Process

After the training process and dataset testing are complete, then the result file from the above training will be taken to carry out the facial recognition process. The facial dataset that will be stored as data to identify is divided into 2 classifications. Classification of faces Besides the child's face class to be supervised, there are also other face classes which are people who are recognized as supervisors. It is this supervisor face class that will trigger later whether the surveillance mode will be activated or remain in an inactive state. The supervisor's face may be more than one person. The data in this class is data on people who are the child's family or acquaintances who are identified as someone who can supervise the child while in the supervision location. This dataset is also the data of people whose faces are detected in the camera, the surveillance mode will not be activated because it will enter the known person data class.

### The Process of Identifying a Child's Face

When all the data on the faces of the children to be supervised and the faces of familiar people are entered into the dataset, now is the time for the system to read the environmental

situation that is within the camera's surveillance radius. If the camera finds a child's face being supervised and then marks it into the system, the system will look at the other parameters in the camera's capture and look for whether there is a familiar face that will be marked as a child supervisor.

### **Process of Activating Surveillance Mode**

If a known person is not found in the camera search, then the surveillance mode will be activated, but if a familiar face appears in the camera, then the surveillance mode will be deactivated immediately, this process will continue to repeat until the child's face being watched is no longer visible to the eyes. camera.

### **Identify Unfamiliar Faces**

When surveillance mode is active, the camera will replace human work as a child supervisor who is considered to be alone on the home page, when the camera finds an unfamiliar face interacting or is in the camera capture for more than 5 seconds, the system will mark that person's face as a face of an unknown person and immediately activates the alarm mode to give a warning to those around you that an unknown person has been detected interacting with the child for more than 5 seconds. The 5-second interval is affected by the spontaneous smile which is a characteristic signal capable of providing the initial and most striking changes in facial appearance which tends to respond quickly to facial expressions in the context of spontaneous association. (Karen L. Schmidt, 2003) This rapid reaction is most likely in response to the initiation of expression rather than any other display phase, since spontaneous smiles usually last for at least 3–4 seconds on average (Frank et al., 1993)

### **The Process of Turning on the Alarm**

It is in this process that the results expected by the researcher occur when the alarm is activated, so it can invite the attention of the people around and make strangers discourage their evil intentions. Turning on the sound of a siren which will be assisted by a loudspeaker that has been connected using Bluetooth.

## **CONCLUSION**

In the tests carried out, the researchers found that the performance of the You Only Look Once (Yolo) algorithm used in the system to prevent the kidnapping of children at home using CCTV that utilizes face recognition is highly dependent on the components used during the training process. The use of the Tesla T4 Graphics Processing Unit (GPU), 15102MiB which is in the Google Colab feature greatly cuts the time needed to conduct training for both epoch 100 and epoch 200. However, in running Python code, testers do not use an additional Graphics Processing Unit (GPU) so the delay processing time is felt in the appearance of video captured by the camera. In this case of child abduction, the algorithm still runs in the background and gives satisfactory results. What needs attention is the position of the camera and the reflection of light as well as the glare level of the sun which makes the light source behind the child's face object to be processed. With this condition, it will seem that it does not produce a decent image, because the object looks dark, defeated by the light source which is directly hit by the object and the camera. The number of cameras used also affects the accuracy of facial recognition and the algorithms that are run. The more cameras used to fill in the blanks from all sides, the easier it will be to detect and identify the child's face. The type of camera used will also affect the brightness and sharpness of the image when the system has to split an image into several

facial images while maintaining the resolution level of the images. Using 6 cameras is the right choice to look after children who are left alone. Using more cameras in the process of supervising children is the best choice, but the research above shows that the use of 6 cameras processed by facial recognition using the Convolutional Neural Network (CNN) method is sufficient to maximize the system in recognizing children's faces. In testing using 3 cameras and using the Convolutional Neural Network (CNN) face recognition method, it gives good results in the afternoon, it can identify at a distance of up to 3 meters but takes quite a long time 15-22 seconds. Whereas for distances up to 2 meters, it can be identified with just 2 cameras at night in a fairly short time of around 8-11 seconds. Meanwhile, for identification using only 1 camera is not recommended, because mistakes can occur in identifying people who should be recognized or people who become supervisors are detected as unknown people, even though the time needed is relatively short, namely around 3-6 seconds when the experiment is carried out during the day. Using more cameras on each side and avoiding light coming from behind can help recognize objects and recognize children's faces, faces of familiar people and identify faces of strangers, offset by the use of the Graphics Processing Unit (GPU) component so that computing time becomes faster.

## REFERENCES

- Aidil, M. (2023, January 13). *BBC New Indonesia*. Retrieved from <https://www.bbc.com:https://www.bbc.com/indonesia/articles/cg13z52mln4o>
- Ajie Pambudhi, F. I. (2017). *Monitoring dan Analisis IP Camera Pada Jaringan Internet*. 2.
- Andryanto, S. D. (2023, February 7 ). *Kasus Penculikan Anak Meningkat Awal 2023, Apa Lagi Selain Penculikan Malika?* (Tempo) Retrieved 06 18, 2023, from <https://nasional.tempo.co/read/1688519/kasus-penculikan-anak-meningkat-awal-2023-apa-lagi-selain-penculikan-malika>
- Deka Rakhmadhani, A. W. (2018). *Alat Pengaman Ruangan Dengan Closed Circuit*. *WIDYA TEKNIK*, Vol. 7, No. 1, 68.
- Digmi, I. (2018, Jan 25). *www.digmi.id*. Retrieved from *memahami epoch batch size dan iteration*: <https://imam.digmi.id/post/memahami-epoch-batch-size-dan-iteration/>
- Dong Chen, X. C. (2012). *Bayesian Face Revisited: A Joint Formulation*. *ECCV*, 567.
- Eric L. Piza, B. C. (2019). *CCTV surveillance for crime prevention. A 40-year systematic review with meta-analysis*. *Criminology & Public Policy*, 136.
- Fitri Damayanti, A. S. (2010). *Absensi Berbasis Pengenalan Wajah Dengan Pendekatan ( 2 D P C A )*. *jurnal simantec*, 159.
- Fitri Damayanti, A. S. (2010). *Absensi Berbasis Pengenalan Wajah Dengan Pendekatan ( 2 D P C A )*. *journal simantec*, 160.
- Glenn Jocher (14), A. C. (2023, 07 17). *Ultralytics YOLOv8 Docs*. Retrieved from <https://ultralytics.com/:https://docs.ultralytics.com/>
- Glenn Jocher, L. S. (2023, july 17). *Ultralytics YOLOv8 Docs*. Retrieved from [ultralytics:https://docs.ultralytics.com/datasets/detect/coco/](https://docs.ultralytics.com/datasets/detect/coco/)
- Gloriawan, J. I. (n.d.). *Implementasi Convolutional Neural Network (Cnn) Untuk Illumination-Invariant Face Recognition Menggunakan Dataset Extended Yale Face Database B*.
- Haitong Lou, X. D. (2023). *DC-YOLOv8: Small size Object detection algorithm based on*. *Preprints (www.preprints.org)*, 2.

- Hasna. (2023, January 16). *Marak Penculikan Anak, Dosen UNESA Ungkap Faktor dan Pencegahannya*. (UNESA) Retrieved January 16, 2023, from <https://www.unesa.ac.id/marak-penculikan-anak-dosen-unesa-ungkap-faktor-dan-pencegahannya>
- Heboh Anak di Bekasi Dibius dan Diculik dalam Karung, Polisi: Hoaks!* (2023, January 30). (Detik) Retrieved June 18, 2023, from <https://www.detik.com/jabar/berita/d-6541752/heboh-anak-di-bekasi-dibius-dan-diculik-dalam-karung-polisi-hoaks>
- Hikvision Launches Smart IP Solution 2.0*. (2018). (Hikvision) Retrieved from <https://www.hikvision.com/en/newsroom/news/press-releases/hikvision-launches-smart-ip-solution-20>
- I N E Indrayana, I. P. (2019). Aplikasi Pemantauan Posisi Anak-Anak Menggunakan Smart Watch dan Database Firebase. *Jurnal Ilmu Komputer VOL. XII No. 2*, 61.
- Iwasokun, G. B. (2021). Geo-Fence Technique for Prevention of Human Kidnapping. *International Journal of Smart Security Technologies, Volume 8* (Issue 2), 36.
- Jaiswal, A. (2022, October 20). *Face Detection using Haar-Cascade using Python*. (analyticsvidhya) Retrieved 06 2023, from <https://www.analyticsvidhya.com/blog/2022/10/face-detection-using-haar-cascade-using-python/>
- Jose Sigut, M. C. (2020). OpenCV Basics: A Mobile Application to Support the Teaching of Computer Vision Concepts. *IEEE TRANSACTIONS ON EDUCATION*.
- Kaipeng Zhang, Z. Z. (2016). Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks. *IEEE Signal Processing Letters, VOL. 23*(NO. 10), 1501.
- Karen L. Schmidt, J. F. (2003). Signal characteristics of spontaneous facial expressions: automatic movement in solitary and social smiles. *Biological Psychology* 65, 50.
- MAHDI, I. (2018). Rancang Bangun Sistem Pendeteksi Penyusup Menggunakan Sensor Passive Infrared Dengan Bunyi Alarm Dan Mengirim Pesan Singkat. *JURNAL Tekno SAINS FST, UTY*, 1.
- Md. Elias Hossain, M. R. (2019). Efficient Anti-Kidnapping and Anti-Harassment (Avoidance-Detection-Notification) Mobile Application for Unwanted Incidents. *2019 IEEE Student Conference on Research and Development (SCoReD)*, 116.
- Mei Wang, W. D. (2020). Deep Face Recognition: A Survey. *journal pre-proofs*, 2.
- Muhammad Ilal Nabsi, A. F. (2021). Pengenalan Citra Wajah Guru Menggunakan Local Binary Pattern Histogram. *The 4th Conference on Innovation and Application of Science and Technology (CIASTECH 2021)*, 430.
- Mühler, V. (2018, July 16). *Realtime JavaScript Face Tracking and Face Recognition using face-api.js' MTCNN Face Detector*. (itnext.io) Retrieved 06 2023, from <https://itnext.io/realtime-javascript-face-tracking-and-face-recognition-using-face-api-js-mtcnn-face-detector-d924dd8b5740>
- Munawir, L. F. (2020). Implementasi Face Recognition pada Absensi Kehadiran Mahasiswa Menggunakan Metode Haar Cascader. *Jurnal Nasional Informatika dan*.
- Nagayama, A. M. (2013). An Intelligent Security Camera System for Kidnapping Detection. *Journal of Advanced Computational Intelligence Vol.17 No.5, 2013*, 752.
- Noviana Dewi, F. I. (2021). Implementasi Deep Learning Menggunakan Convolutional. *factor exata*, 35.

- Penerapan Algoritma Deteksi Wajah dalam Sistem Digital.* (2023, June 13). Retrieved from verihubs: <https://verihubs.com/blog/algoritma-deteksi-wajah/>
- Perdana, A. H. (2019). Implementasi Sistem Deteksi Mata Kantuk Berdasarkan Facial Landmarks Detection Menggunakan Metode Regression Trees.
- Polri, P. B. (2023, Januari 13). *Waspada, Jumlah Anak Korban Penculikan Makin Banyak.* (Polisi Republik Indonesia) Retrieved Januari 13, 2023, from [https://pusiknas.polri.go.id/detail\\_artikel/waspada,\\_jumlah\\_anak\\_korban\\_penculikan\\_makin\\_banyak](https://pusiknas.polri.go.id/detail_artikel/waspada,_jumlah_anak_korban_penculikan_makin_banyak)
- Pratiwi, F. S. (2023, March 6). *Tahun 2023 Baru Dua Bulan, KPAI Catat 14 Kasus Penculikan Anak.* (dataindonesia) Retrieved from <https://dataindonesia.id/varia/detail/tahun-2023-baru-dua-bulan-kpai-catat-14-kasus-penculikan-anak>
- Rika Saraswati, S. C. (2015). *Hukum Perlindungan Anak di Indonesia.* Bandung: PT. Citra Aditya Bakti.
- Ritchie, C. (2021, September 27). *AiMS —Meet the New VMS to Reduce Kidnapping with AI Video Analytics.* Retrieved from <https://skyrecofficial.medium.com:https://skyrecofficial.medium.com/aims-meet-the-new-vms-to-reduce-kidnapping-with-ai-video-analytics-8ecba7e831e2>
- Sayed Al-Aidid, D. S. (2018). Sistem Pengenalan Wajah dengan Algoritma Haar. *Jurnal Rekayasa Elektrik, 14*, 66.
- Siregar, A. R. (2022, January 14). *Kompas.* Retrieved from <https://megapolitan.kompas.com:https://megapolitan.kompas.com/read/2022/01/14/19121051/bawa-kabur-anak-yang-bermain-di-depan-rumah-penculik-di-tangsel-ditangkap>
- team, O. (2023). *About OpenCV.* (opencv) Retrieved from <https://opencv.org/about/>
- Wang, C.-F. (2018, July 27). *How Does A Face Detection Program Work? (Using Neural Networks).* (towardsdatascience) Retrieved 06 2023, from <https://towardsdatascience.com/how-does-a-face-detection-program-work-using-neural-networks-17896df8e6ff>
- Wang, Y. &. (2020). Smart CCTV for Public Safety: A Review. . *IEEE Access, 8*, 150320-150333.
- Zulfikar Sembiring, R. M. (2019). Perancangan Alat Pelacak Lokasi Dalam Mengantisipasi Penculikan Anak. *Techno.COM, Vol. 18, No. 1*, 13.