Design and Research on the Use of AI-Based Image Recognition Technology for Collaborative Campus Safety Management

Chia-Yu Chang Chien*

Department of Computer Science and Information Engineering, National Taitung University;
e-mail: zx415523@gmail.com
*Corresponding Author: zx415523@gmail.com
DOI: https://DOI.org/10.30211/JIC.202402.010
Submitted: Oct. 18, 2024 Accepted: Dec. 15, 2024

ABSTRACT

Several significant issues that campuses have long faced include school bullying and unauthorized individuals intruding on school premises. Similar incidents frequently appear in the news, and bullying events can cause severe physical and psychological harm to both students and teachers. The unauthorized entry of unknown persons into schools poses potential dangers to everyone on campus, as no one can predict their intentions. Despite schools' efforts in promoting anti-bullying measures and safety awareness, the effectiveness of preventing such incidents remains limited without concrete actions or measures. Therefore, I believe that substantive protective measures are necessary to further strengthen campus safety and create a secure learning environment.

This study explores the integration of the Internet of Behaviors (IoB) and AI technologies for behavioral recognition within the campus. The research focuses on two key aspects. First, we collect real-time image data from sensors, cameras, or surveillance systems to address school bullying. Using IoT for data transmission, this information is analyzed, and image recognition combined with deep learning techniques is employed to determine if physical bullying is occurring. Once an incident is detected, an immediate alert is sent to the school administration for prompt action, thereby preventing bullying in real time. Second, the prevention of unauthorized individuals entering the campus is achieved through similar IoT and image recognition methods. While ensuring campus safety, it is also crucial to consider the privacy of staff and students and ensure that the application of these technologies remains within legal boundaries.

Keywords: Image Recognition, Deep Learning, Internet of Things, Campus Safety

1. Introduction

[1] It highlights that the current regulations and mechanisms of the Ministry of Education regarding school bullying remain passive, relying on reports before initiating an investigation process. He strongly urges the ministry to establish more proactive mechanisms for preventing and investigating school bullying. He particularly emphasizes that such mechanisms must protect the whistleblower’s privacy, allowing public authority to effectively prevent and combat

school bullying. Additionally, due to the closed nature of schools and the pressure from numerous review committees, there is often significant interpersonal pressure within schools. Senior teachers or principals often downplay cases, lacking follow-up measures or appropriate assistance. Therefore, we strongly recommend involving external third-party personnel in the investigation of school bullying cases.

[2] states that, according to statistics from Europe, the United States, and even Taiwan, the prevalence of school bullying is as high as 20–30%. The key to addressing this issue is prevention and early identification and active intervention when bullying occurs. She advocates for creating a positive environment and effective channels to encourage children to speak up against bullying. She also stresses that both teachers and students should engage in open discussions about bullying incidents and their solutions, raising awareness and reinforcing that bullying is an unacceptable behavior.

In recent years, the government has prioritized campus safety, implementing stronger measures. Currently, campus security primarily relies on security personnel, which, while somewhat effective, faces challenges. For example, security personnel often cannot detect dangerous situations immediately, leaving some risk. This highlights the limitations of traditional human oversight in campus safety management, underscoring the need for a system that can recognize dangerous behaviors in real-time.

According to statistics from the Ministry of Education's Campus Security Information Network [4-5], Taiwan received 14,707 reports of violent incidents, deviant behaviors, and conflicts with discipline in the first half of 2023, and 17,386 reports in the second half. High schools, middle schools, and elementary schools accounted for 93.4% of these reports. However, these are only the reported incidents. There are likely many unreported cases where victims did not dare to come forward, and many witnesses of bullying incidents may want to report them but are either unsure of how to do so or fear becoming the next victim.

On the other hand, unknown individuals entering school premises also pose a threat to staff and students. Even though surveillance cameras are widespread, the limited number of security personnel makes it difficult to monitor and respond to every part of the campus. Modern AI technology can combine with traditional security measures in such understaffed environments to safeguard campus safety. This motivates my research into developing a system that utilizes the Internet of Things (IoT) and image recognition technologies to establish a third-party system capable of detecting real-time images, analyzing events, and issuing alerts.

2. Literature Review

2.1 About Facial Recognition

In [6], it is mentioned that automatic emotion recognition based on facial expressions is an intriguing research area with practical applications emerging in various domains, such as security, healthcare, and human-computer interfaces. Researchers in this field continually develop technologies to interpret and extract features of facial expressions through methods such

as decoding and encoding, enabling computers to make better predictions and judgments. Because deep learning has been so successful, different architectures have been used to make it work better. Two examples are Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). Researchers have begun employing this technology to identify human emotions and have achieved significant results. Studies in [7-9] used the SE-ResNet architecture as the neural network model for training and testing. Comparisons and cross-database validations were conducted between RAF-DB and AffectNet. To attain better performance, transfer learning techniques were employed, with models trained on AffectNet serving as pre-trained models for training on RAF-DB. Both RAF-DB and AffectNet are datasets used in emotion recognition research.

Various mechanisms can authenticate the validity of identity for authentication purposes. According to [10], Radio Frequency Identification (RFID) has rapidly developed and is widely used in dormitory access control systems, library borrowing systems, attendance systems, and payment systems due to its convenience. However, RFID can only verify the legitimacy of the card and cannot ascertain whether the user is the cardholder. In [11], it is stated that the core of smart campuses is safety, with many campus security units primarily relying on surveillance equipment for campus monitoring to enhance safety. Weather, on the other hand, can easily affect surveillance devices, making images less clear. This is why optical flow or other extra devices are needed to make surveillance footage clearer.

2.2 About Suspicious Individuals

Identifying suspicious human activities from surveillance footage is an important research area in image processing and computer vision, as mentioned in [12]. Through visual monitoring, human activities in sensitive public areas such as bus stops, train stations, airports, banks, shopping malls, schools, parking lots, and roadways can be observed. This can help prevent terrorism, theft, illegal parking, vandalism, fights, robberies, and other suspicious activities.

Since continuous observation of public places is quite challenging, intelligent surveillance systems are needed to monitor human activities in real-time, classify them as normal or abnormal, and issue alerts. In recent years, a wealth of research in the field of visual surveillance has emerged, focusing on identifying and detecting abnormal activities. In [13], we typically categorize pedestrian detection, tracking, and activity recognition into two types of methods: traditional methods and deep learning-based methods. Traditional methods include the V. Jones method for facial recognition and pedestrian detection, as well as HOG and DPM methods. However, these methods are computationally intensive, time-consuming, and require human involvement.

Recently, CNN-based deep learning techniques have garnered attention due to their accuracy in pedestrian identification. R-CNN was the first deep learning model used for object detection, and other multilayer convolutional networks, such as Mask R-CNN and its variants, have also been widely applied. Single-stage CNN models, such as You Only Look Once (YOLO) and SSD, have also achieved success in pedestrian detection. However, traditional real-time pedestrian detection methods have become less applicable.

Therefore, the YOLO network was introduced as an object regression architecture to improve detection speed and accuracy. We have shown that the improved YOLOv5 method effectively detects small and constant pedestrians.

2.3 About Handheld Weapons

According to [14], in the modern technological era, people use surveillance cameras in various regions to prevent crime. Many locations have installed a large number of cameras, requiring security personnel to monitor them all simultaneously. Generally, if a crime occurs, security personnel rush to the scene, check the recordings, analyze the footage, and collect necessary evidence; therefore, establishing an active system at crime scenes is essential. If software can immediately alert security personnel upon detecting a threatening object, they can take prompt action to prevent potential criminals from committing crimes. Thus, it is crucial to establish a system that can learn to detect threatening objects. Many studies have shown that handheld weapons are among the most critical elements in various criminal activities, including theft, illegal hunting, and terrorism. One of the solutions to address such criminal activities is to install monitoring systems or control cameras to enable security departments to take appropriate measures at an early stage.

In this research, due to the lack of a standard dataset for weapon detection and identification, images of weapons downloaded from the internet are used as the dataset. The downloaded weapon images must be high-quality pictures taken from different angles to successfully detect and identify real-world weapons. Furthermore, we need to remove irrelevant objects from each weapon image to improve the accuracy of the neural network model. Therefore, the downloaded weapon images are individually examined and modified using different computer applications as necessary, such as filling, masking, background cleaning, resizing, and rotating. Once the prepared images for each weapon category are converted into a dataset, the ResNet architecture is used for model training.

2.4 Internet of Behaviors

In [15], Dr. Göte Nyman mentioned in his 2012 blog that the Internet of Behaviors (IoB) was created to allow technology companies to predict user needs through various online tools. The Internet of Things (IoT) has matured today, and IoB can penetrate households, transportation, healthcare, and businesses, thereby improving the quality of human life. In [16], it is proposed that emotional robots based on IoT and artificial intelligence can provide emotional service technologies for humans, enabling emotional robots to perform related services according to human needs. Furthermore, [17] notes that online surfing has become a popular activity for many consumers, who not only shop online but also search for relevant information about products and services before making purchases. As a result, many recommendation systems are currently actively providing users with information about relevant product needs.

3. Research Methods and Procedures

The Internet of Behaviors (IoB) leverages big data analysis and machine learning to assess and identify potential campus safety incidents. Upon detection of such incidents, the system transmits alerts to the IoB, which then reports the information to the campus safety management platform for appropriate action. The architecture diagram is shown in Figure 1.

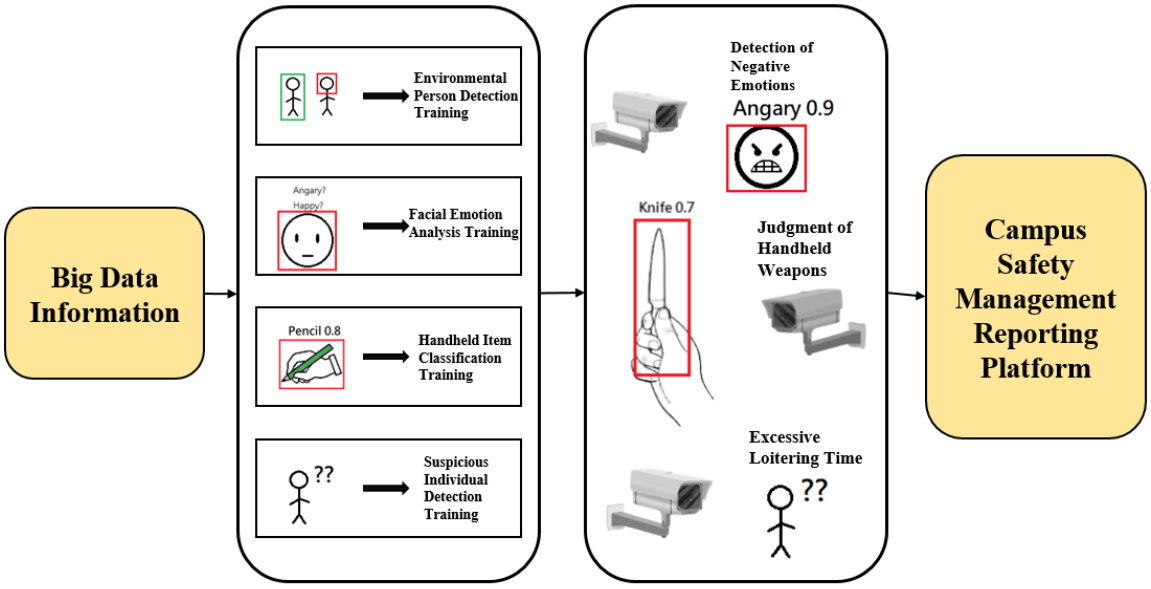

Figure 1. System Architecture Diagram

The research methodology is mainly divided into the following four steps: "Image Data Collection," "IoT Technology," "Object Detection," and "Event Determination," as shown in Figure 2.

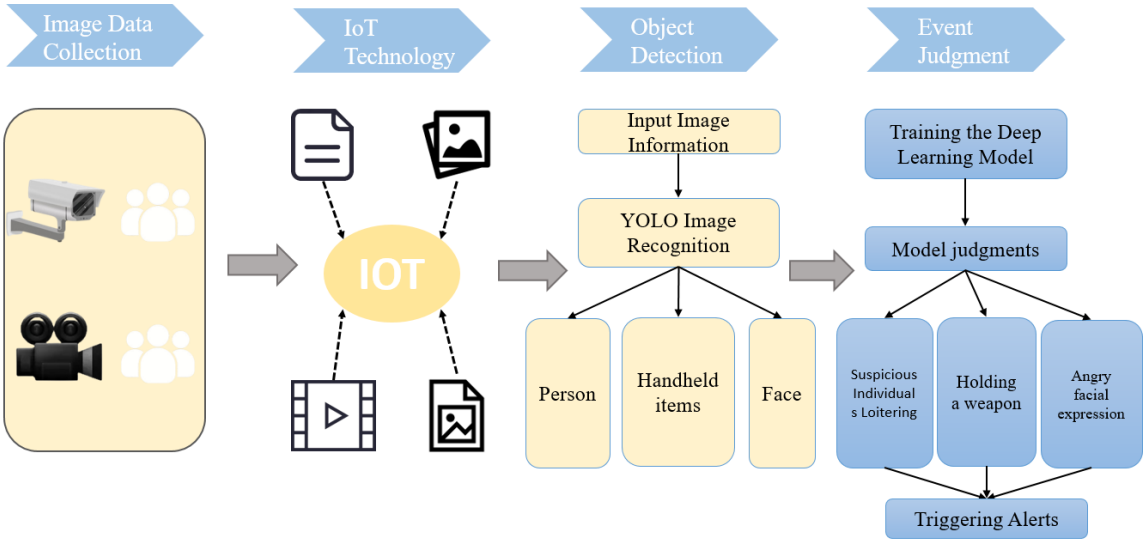

Figure 2. Research Process Flowchart

3.1 Image Data Collection

In the first phase, image data will be collected within the campus. Cameras, sensors, and surveillance systems will be utilized to capture real-time imagery of the campus area. The collected data will be used for subsequent object detection and event judgment.

3.2 Internet of Things (IoT) Technology

The surveillance systems and cameras will be connected to the IoT gateway, enabling the observation devices to transmit data to the school's main server, remote servers, or cloud platforms. The IoT gateway serves as an intermediary device, typically used to transfer data from various sensors and devices to the network.

3.3 Object Detection

In the third phase, the study will employ the YOLO (You Only Look Once) object detection model to identify "individuals," "faces," and "held items" within campus scenes. YOLO is a fast and accurate object detection model capable of detecting multiple object types and their locations in a single computation. Training a YOLO model requires a substantial number of images and corresponding annotated data, as shown in Figure 3.

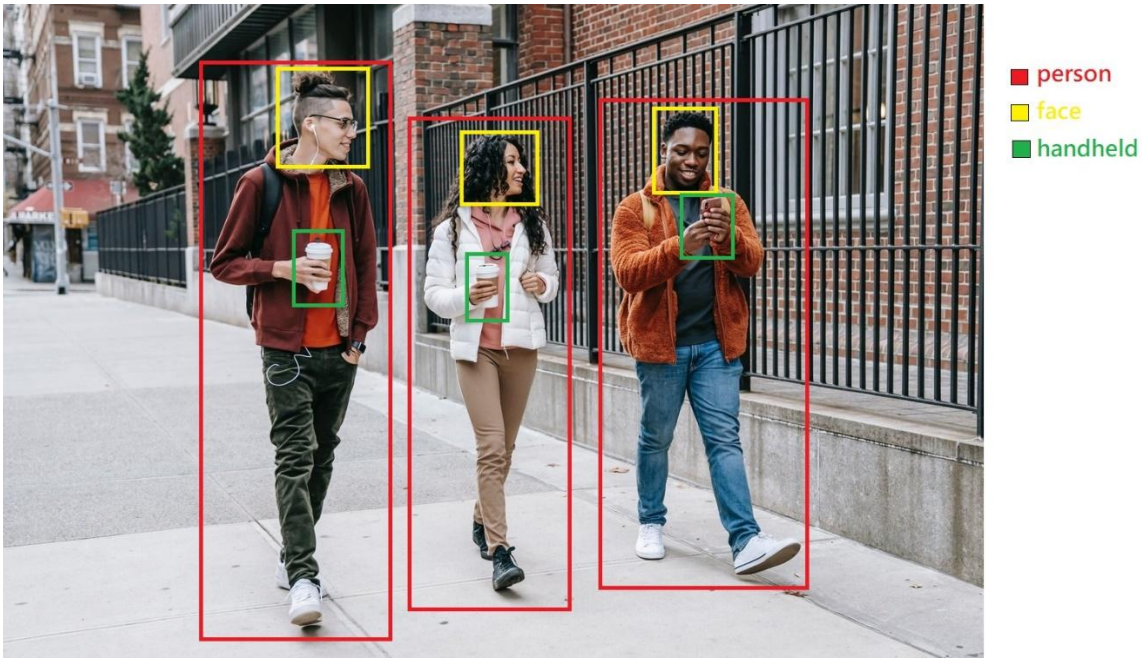

Figure 3. Object Detection Annotation Data

3.4 Event Judgment

After the objects have been annotated, it is essential to proceed with event judgment. The threats to campus safety are categorized into three main types: “Suspicious Individuals Loitering,” “Facial Emotion Recognition,” and “Judgment of Handheld Weapons.”

4. Results and Discussion

4.1 Suspicious Individuals Loitering

In this study, the YOLOv5 model is utilized to capture individuals and assess loitering time for judgment. Once a target enters the monitoring range, a safety classification and timing system is initiated, as illustrated in Figure 4. The duration of activity and loitering within the same surveillance camera’s field of view will be monitored and categorized according to a

safety classification table for alerts or notifications. Additionally, the location of the surveillance camera will influence the notification threshold values. For instance, longer dwell times in areas such as classrooms, hallways, and playgrounds are considered normal, thus increasing the safety threshold. Conversely, in more secluded areas like external walls, restrooms, or remote locations on campus, extended loitering times significantly raise the likelihood of dangerous behavior. Therefore, the notification thresholds T1 and T2 will be adjusted according to time and location to accommodate various scenarios, shown in Figure 4.

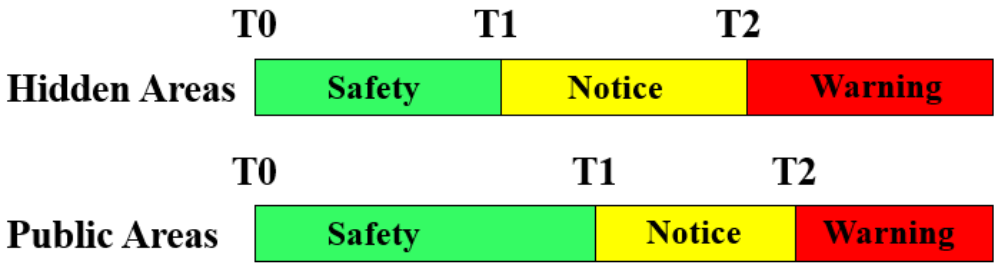

Figure 4. Loitering Threshold Diagram

The time is primarily divided into three segments. If the loitering time falls within the safe segment, it is considered within an acceptable range of safety. In the attention segment, there should be vigilance for any unusual intentions, but no specific reporting is required. Once the warning segment is reached, it indicates that the individual has lingered or loitered for too long, necessitating relevant guidance or reporting actions.

4.2 Facial Emotion Recognition

In this study, the SE-ResNet architecture is employed for model training. The SE (Squeeze and Excitation) Net architecture is illustrated in Figure 5.

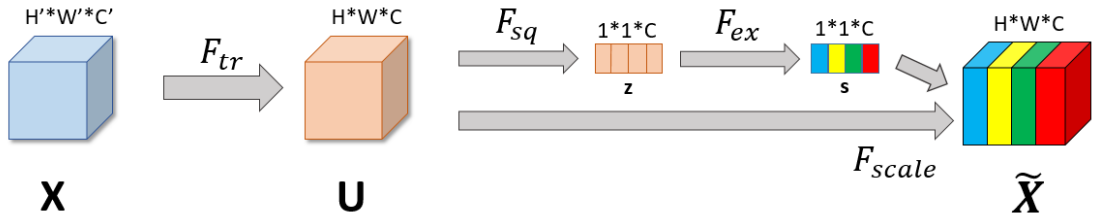

Figure 5. SE-Net Architecture Diagram

The notation F_{tr} in the figure represents the transformation from the original image X to the convolved image U. Here, $X \in \mathbb{R}^{W' \times H' \times C'}$ and $U \in \mathbb{R}^{W \times H \times C}$, where W and H denote the width and height of the image, respectively, and C represents the number of channels in the image. The formula for F_{tr} is as follows:

$$u_c = v_c \times X = \sum_{s=1}^{C'} v_c^s \times x^s$$

……[Formular 1]

In Formular (1) v_c represents the c-th convolutional kernel, and x^s denotes the s-th input. After completing the transformation, we can obtain C feature maps, each of size H*W . The primary goal of this step is to explicitly model the interdependencies between channels, thereby enhancing the learning of convolutional features. This allows the network to increase its sensitivity to the characteristics of the information.

4.3 In Squeeze:

In the squeeze stage, F_{sq} refers to the operation performed on U to achieve a squeeze effect, which is essentially global average pooling (GAP). The formula for F_{sq} is as follows:

$$z_c = F_{sq}(u_c) = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} u_c(i,j)$$

……[Formular 2]

The formula in (2) applies to each channel in C , where the average is computed over all pixel values in the H*W feature map. This results in a 1*1*C matrix, as illustrated in Figure 6. This matrix represents the distribution of feature map values across the C layers, providing a statistical summary that encapsulates the essential information of the entire image.

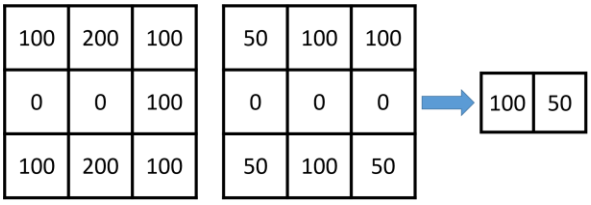

Figure 6. Illustration of Global Average Pooling

4.4 In Excitation:

To better utilize the information aggregated during the Squeeze operation and to further capture the dependencies between feature channels, the formula for F_{ex} is as follows:

$$s = F_{ex}(z, W) = \sigma\big(g(z, W)\big) = \sigma\big(W_2 \delta(W_1 z)\big)$$

……[Formular 3]

This section contains two fully connected layers with dimensions $W_1 \in \mathbb{R}^{\frac{C}{r} \times C}$ and $W_2 \in \mathbb{R}^{C \times \frac{C}{r}}$. After multiplying the pooled data z by W_1 , the dimension decreases: $W_1 z \in \mathbb{R}^{1*1*\frac{C}{r}}$. Here, δ is the activation function ReLU. Subsequently, after multiplying by W2W_2W2, the dimension increases again: $W_2 \delta(W_1 z) \in \mathbb{R}^{1*1*C}$. Finally, the Sigmoid activation function is applied to distribute the results within the range of 0 to 1, as shown in Figure 7.

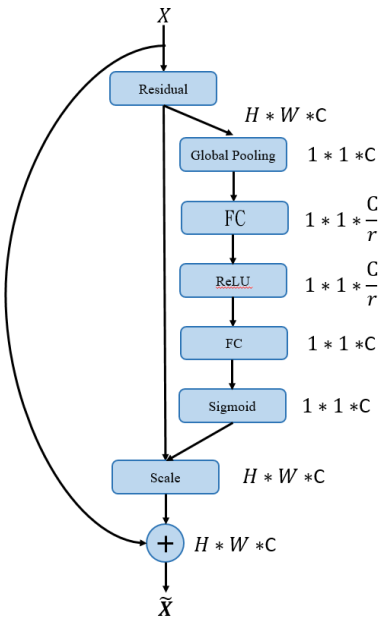

Figure 7: SE-Resnet Diagram

4.5 In Scale:

This stage involves multiplying the learned weights S by the initial U. By utilizing the previous stages, the model learns the importance of inter-channel relationships and updates the weights, enhancing the model's ability to evaluate features for each channel. Ultimately, ResNet (Residual Neural Network) incorporates the SE block module into its structure. As shown in the confusion matrix in Figure 8, the model trained with this architecture demonstrates a relatively high accuracy in predicting facial expressions. The AffectNet-sourced dataset includes 456,349 images, while the RAF-DB holds 15,339 images. Notably, the number of happy facial data accounts for approximately 30% of the entire dataset, and the results indicate that the accuracy exceeds that of other facial expressions.

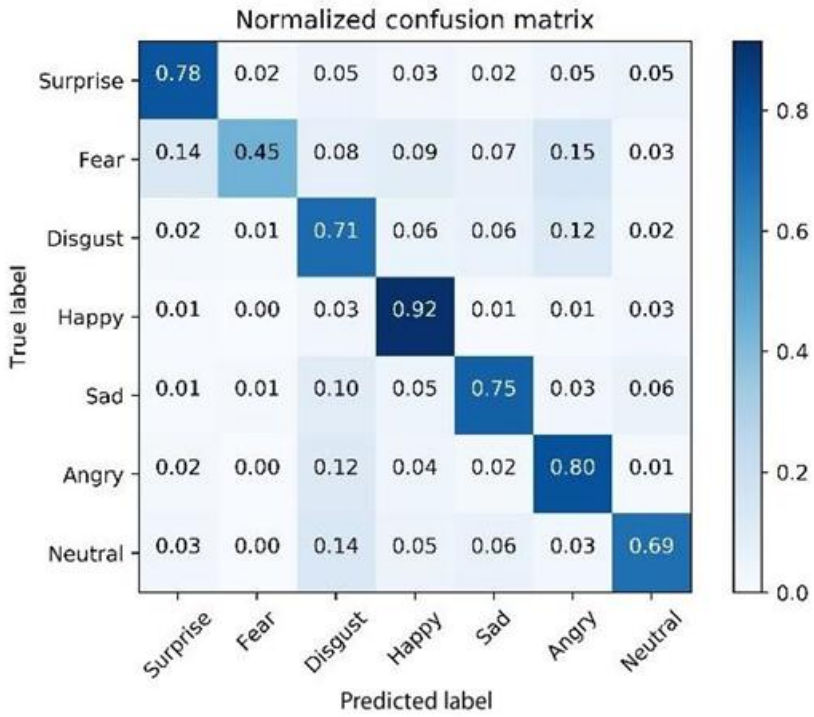

Figure 8. Normalized Confusion Matrix (Data sourced from [7], Figure 8(c)).

To evaluate the performance of the model, classification evaluation metrics such as Precision, Recall, and F1 Score are used, as shown in Table 1.

Table 1. Model Classification Evaluation Metrics.

Class	Precision	Recall	F1 Score
Surprise	0.788	0.804	0.796
Fear	0.703	0.417	0.523
Disgust	0.710	0.710	0.710
Happy	0.736	0.920	0.818
Sad	0.688	0.750	0.718
Angry	0.696	0.721	0.708
Neutral	0.663	0.690	0.676

4.6 Handheld Weapons:

This research plans to collect images of various weapons commonly associated with campus safety concerns on the internet to create a dataset. We will use a Convolutional Neural Network (CNN) for model training to help determine whether an individual is holding a dangerous item. Dangerous items may include weapons such as butterfly knives and spring knives, which should not be present on campus.

The training will utilize ResNet101 (Residual Neural Network), with a total of 18,000 images primarily focused on knives. Each of the positive and negative datasets will contain 9,000 images. To evaluate the model's performance, 50 images will be selected from both the positive and negative datasets as a test set to validate the model's classification capability. The remaining images will be used for the training set. Figure [9] displays the final classification

confusion matrix. If the model classifies an item as a handheld dangerous object, it will trigger a notification or require careful observation.

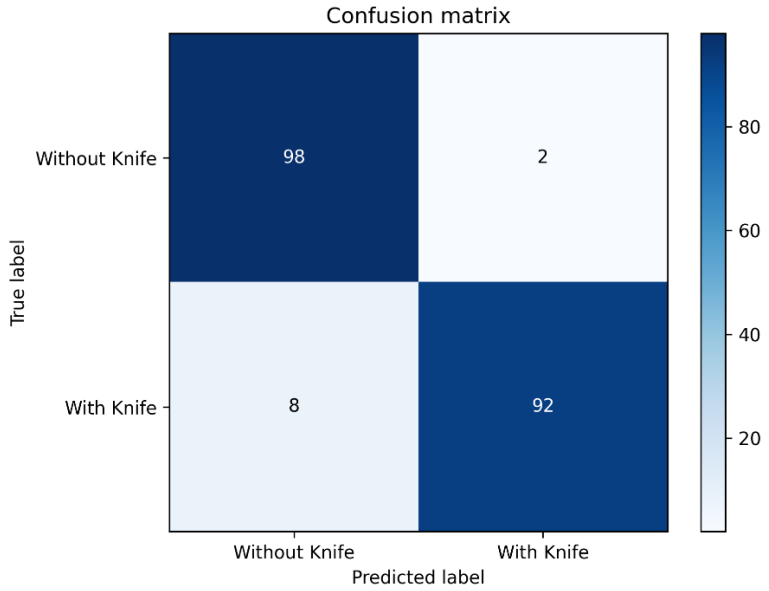

Figure 9. Confusion Matrix for Handheld Weapons Classification

5. Results and Discussion

This study proposes a campus safety management system integrating Artificial Intelligence (AI) and Internet of Things (IoT) technologies. Using deep learning models like SE-ResNet and ResNet101 along with object detection methods like YOLOv5, the system has shown promise in simulated environments in terms of accuracy. It aims to enhance the efficiency of campus safety management and foster a secure learning environment. However, several challenges were encountered during the research process, which impacted the system's accuracy and stability. These challenges include issues with the camera angles of surveillance systems, limitations in image resolution, and insufficient diversity in data sources. Overcoming these challenges will enable the system to achieve greater potential in practical applications and contribute to the advancement of smart campus safety management technologies.

References

[1] Lin, Z. The rise of school bullying: Over 1,000 cases annually. The Epoch Times News, 2023. Retrieved from: https://www.epochtimes.com/b5/23/7/11/n14032325.htm.

[2] Liao, X. Observe these 5 signs to detect if your child may be experiencing school bullying. Liberty Health Network, provided by Ton-Yen General Hospital, 2023. Retrieved from: https://health.ltn.com.tw/article/breakingnews/4339705.

[3] Zhang, Q., Hang, J., Zhang, B., Wei, M., Zhu, Q. and Li, X. Campus safety monitoring system based on deep learning. IEEE Xplore, 2021. DOI: 10.1109/DCABES52998.2021.00036.

[4] Yang, Y. Ministry of Education Campus Safety and Disaster Prevention Reporting and Handling Center: Various data on campus safety reports from January to June, 2023. 2024. Retrieved from: https://csrc.edu.tw/filemanage/detail/4beb2020-5d2d-4f62-9ca4-7af1d0eacd6f.

[5] Yang, Y. Ministry of Education Campus Safety and Disaster Prevention Reporting and Handling Center: Various data on campus safety reports from July to December, 2023. 2024. Retrieved from: https://csrc.edu.tw/filemanage/detail/4230bb4f-ddfd-4a05-918d-9a935e0f7277.

[6] Mellouka, W. and Handouzia, W. Facial emotion recognition using deep learning: review and insights. ScienceDirect, 2020, 175, 689-694. DOI: 10.1016/j.procs.2020.07.101

[7] Huang, Z., Chiang, C., Chen, J., Chen, Y., Chung, H., Cai, Y. and Hsu, H. A study on computer vision for facial emotion recognition. Scientific Reports, 2023, Article number: 8425. DOI: 10.1038/s41598-023-35446-4.

[8] John. Quick paper review: Squeeze-and-excitation networks. Medium, 2020. Retrieved from: https://meetonfriday.com/posts/79fdff34/Vaibhav Khandelwal.Medium.The Architecture and Implementation of VGG-16.

[9] Hu, J., Shen, L., Albanie, S., Sun, G. and Wu, E. Squeeze-and-excitation networks. IEEE Xplore, 2018. DOI: 10.1109/CVPR.2018.00745

[10] Zheng, L., Li, Q., Wang, H. and Zhang, J. A new mutual authentication protocol in mobile RFID for smart campus. IEEE Access, 2018, 6, 60996-61005. DOI: 10.1109/ACCESS.2018.2875973.

[11] Wan, W. Research on intelligent video analysis technology in smart campus security scenario. 5th Asian Conference on Artificial Intelligence Technology (ACAIT), 2021, 362-365. DOI: 10.1109/ACAIT53529.2021.9731339

[12] Gawande, U., Hajari, K. and GolSam, Y. Novel person detection and suspicious activity recognition using enhanced YOLOv5 and motion feature map. Springer Link, 2024, 57, Article number: 16. DOI: 10.1007/s10462-023-10630-0

[13] Gawande, U., Hajari, K. and GolhaE, Y. Real-time deep learning approach for pedestrian detection and suspicious activity recognition. ScienceDirect, 2023, 218. DOI: 10.1016/j.procs.2023.01.219

[14] Kaya, V., Tuncer, S. and Baran, A. Detection and classification of different weapon types using deep learning. MDPI, 2021. DOI: 10.3390/app11167535

[15] Zhao, Q., Li, G., Cai, J., Zhou, M. and Feng, L. A tutorial on internet of behaviors: Concept, architecture, technology, applications, and challenges. IEEE Communications Surveys & Tutorials, 2023, 25(2), 1227-1260. DOI: 10.1109/COMST.2023.3246993

[16] Hui, H., Zhang, X., Liu, Y., Li, T. and Wang, J. Affective computing model with impulse control in internet of things based on affective robotics. IEEE Internet of Things Journal, 2022, 9(21), 20815-20832. DOI: 10.1109/JIOT.2022.3176323

[17] Fong, A.C.M., Zhou, B., Hui, S.C., Hong, G.Y. and Do, T.A. Web content recommender system based on consumer behavior modeling. IEEE Transactions on Consumer Electronics, 2011, 57(2), 962-969. DOI: 10.1109/TCE.2011.5955246