DOI: https://doi.org/10.59294/HIUJS.VOL.4.2023.394

Application of artificial intelligence to build a security control software system in local military units

Nguyen Trong The, Nguyen Khac Diep* and Duong Xuan Tra

Information Technology Institute

ABSTRACT

This paper introduces the application of artificial intelligence to build a security control software system in local military units. This software system uses state-of-the-art convolutional neural networks (CNN SOTA) for facial recognition by testing two of the best facial recognition models currently available: the FaceNet model and the VGGFace model. Through testing on the proposed hardware, the FaceNet model meets the accuracy and speed requirements for practical application. The software includes multiple identity management categories to ensure information security and monitor the access of soldiers and other individuals. Additionally, the software features access management functions for protected areas, allowing for audio and visual alerts to ensure safety and security in those areas. The software also enables users to set up connections with other devices for efficient data collection and processing. Simultaneously, it supports synchronized data connection to help users save time and effort in managing information. Moreover, the software includes user-friendly interfaces and customizable settings, ensuring ease of use and adaptability to the specific needs of each military unit. this software system provides a comprehensive and effective solution for ensuring security and monitoring access in local military units. By leveraging artificial intelligence, the system can adapt and improve over time, offering enhanced performance and capabilities to meet the evolving security needs of military organizations. The innovative approach presented in this paper has the potential to significantly improve the overall security and efficiency of local military units, contributing to the safety and well-being of both military personnel and the communities they serve.

Keywords: *artificial intelligence, CNN SOTA, security control, facial recognition*

1. INTRODUCTION

Currently, with the strong development of machine learning research, the application of deep learning models for facial recognition is being used in security systems, tracking systems, surveillance, and in attendance and timekeeping systems more quickly, accurately, and easily [1-2].

Facial recognition problems are divided into two types: face verification and facial recognition (FR). Face verification is a one-to-one comparison problem, which only confirms whether the two input images are of the same person or not, with the output being true or false. This problem is commonly applied in security systems such as door unlocking and mobile device unlocking. Facial recognition is a one-to-many comparison problem. The problem answers the question "who is the person in the photo?", with the input being an image containing a face and the output being the name label of the person in the image. FR is often applied in citizen surveillance systems, facial timekeeping systems, attendance in schools, searching for subjects in public areas, and verifying information in airport and border areas [3].

In the field of national defense and security, with many unique characteristics, there are many sensitive military areas that require authorization to enter and exit, and additional military areas need to monitor people coming in and out. The application of facial recognition to control and monitor helps reduce the guarding force, making operations faster and easier, while ensuring efficiency and safety. However, manual access control or the use of mechanical devices has many limitations and difficulties, such as: labor-intensive, lack of accuracy, untimely, inconvenient, etc. Therefore, a smarter and more accurate software is needed to recognize and identify identities to support the detection and alert of intruders in the Tactical Command Room

Corresponding Author: Dr. Nguyen Khac Diep
Email: diep62@mail.ru

of military units, which has become an urgent need today. The solution of applying artificial intelligence, specifically state-of-the-art (SOTA) models in the field of machine learning and deep learning, to authenticate the identities of entities entering and exiting specific locations (mainly at the Tactical Command Room door, but can be deployed at other locations such as weapon storage, secret document storage, etc.), combined with mechanical and electronic modules that allow the issuance of warning signals when a stranger interacts with the device, and automatically records in the logbook and report, is considered a comprehensive and multi-functional solution.

Therefore, the goal of this paper is to introduce the application of artificial intelligence to build security control software for military units. The research team uses SOTA CNN networks for facial recognition by experimenting with two models currently evaluated as effective: the FaceNet model and the VGGFace model. This software can be applied to military or civilian units with high security and management requirements. The paper is presented in the following order: Part 2 presents the selected solutions to solve the posed problem as well as the initial achieved results and finally, the conclusion is presented in Part 3.

2. SOLUTION RESEARCH AND RESULTS OF DEVELOPING A SECURITY CONTROL SOFTWARE

2.1. Solution Research

2.1.1. Software solution

Based on the above needs, the research team proceeds to design and build artificial intelligence software to control the security for the military unit, which is a comprehensive and multi-functional solution. This software includes several identity management categories to ensure information security and monitor the entry and exit of the military and other subjects. In addition, the software also includes features to manage entry and exit information in protected areas, allowing audio and visual alerts to ensure safety and security for that area. The software also allows users to set up connections with other devices to collect and process data effectively. At the same time, it also supports synchronized data connection to help users save time and effort in managing information.

System function requirements: The system can operate in two modes, including attendance mode and monitoring mode, in which:

- Attendance mode: Requires authentication to enter and exit, save the log. This mode can be scheduled at certain time frames.
- Monitoring and tracking mode: Identify objects entering and exiting normally, then save the log, and simultaneously issue an audio alert when an intruder is detected.

2.1.2. Choosing a CNN Model

The *Convolutional Neural Networks (*CNN) model is a popular deep learning model widely used in tasks such as face recognition, image classification, pattern recognition, and object feature extraction from images. Figure 1 shows that 47% of image processing applications use the CNN model. The progress of the convolutional neural network has led to a significant increase in the performance of modern image processing methods. Research [4] measures the effectiveness of the CNN model compared to three famous recognition methods: Principal Component Analysis (PCA), Local Binary Pattern Histograms (LBPH), and K-Nearest Neighbors (KNN). The experiment was conducted on the ORL database, which includes 400 different objects (40 object classes/10 images per class). The experimental results showed that LBPH performed better than PCA and KNN, but all three methods had lower accuracy than the CNN-based method, which achieved 98.3% accuracy. The research results demonstrate the superiority of CNN models over traditional recognition methods.

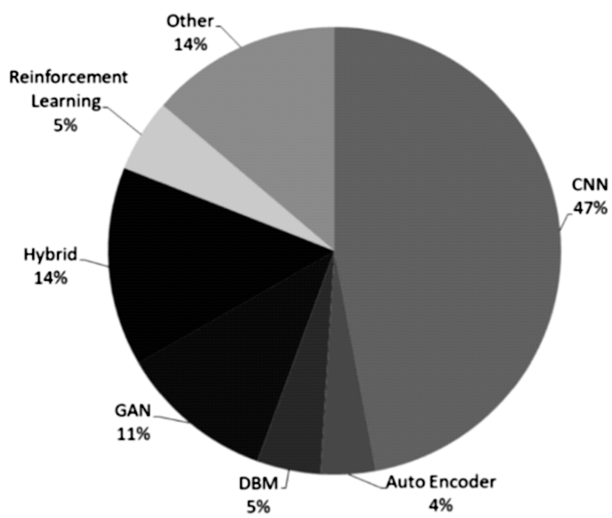

Figure 1. The methods for face recognition

A face recognition system based on CNN follows a standard four-step processing procedure [5-7]:

1. Face detection: collect images from the device and detect faces in the image.

2. Face analysis: analyze the facial features obtained from step 1, including the distance between the eyes, the depth of the eye sockets, the distance from the forehead to the chin, the shape of the cheekbones, and the contour of the lips, ears, and chin. The purpose is to identify key points on the face to distinguish between different people.

3. Convert images to data: the facial features will be represented as numerical data, usually a feature vector for the face.

4. Face recognition: the feature vector obtained will be compared with the stored feature vectors to find the face with the highest similarity.

The CNN model originates from the operation of convolving two matrices. Matrix convolution is an important technique in image processing used for smoothing images, extracting edges, and computing image derivatives. In CNN, the input image is a matrix with dimensions (HxWxD), where H is the height, W is the width, and D is the number of channels in the image. For color images, D=3, for grayscale images, D = 1. The other matrix is called a kernel, which is a matrix with dimensions (MxNxD), where M and N are usually (3x3), and D has the same size as the number of channels in the input image [1].

The CNN network structure consists of layers stacked on top of each other, including four types of layers: convolution layer, pooling layer, non-linear layer, and fully-connected layer. However, the CNN network structure can vary depending on the specific task. During the transformation of the CNN network, many highly effective models have been developed, such as VGGNet, GooLeNet, ResNet, EffecientNet, and SeNet.

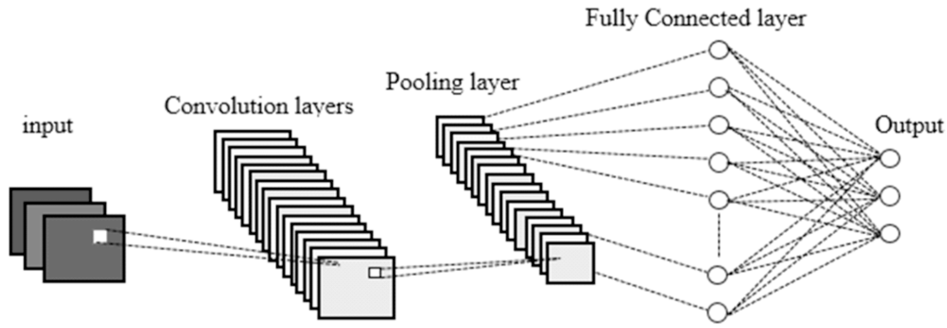

Figure 2. CNN network model

Before the emergence of deep learning and effective deep learning models, traditional feature extraction methods often represented faces as a landmark map. The face landmark map is a set of the most distinctive feature positions on the face, such as the eyes, nose, mouth, chin, and nasal spine. However, it also ignores many other facial features such as cheeks, forehead, skin color, ears, wrinkles, eye color, etc. The landmark vectors will apply classic machine learning classification algorithms such as SVM, KNN, Naive Bayes, Random Forest, MLP.. to determine the identity of the face. Traditional methods often have low accuracy according to research from [4].

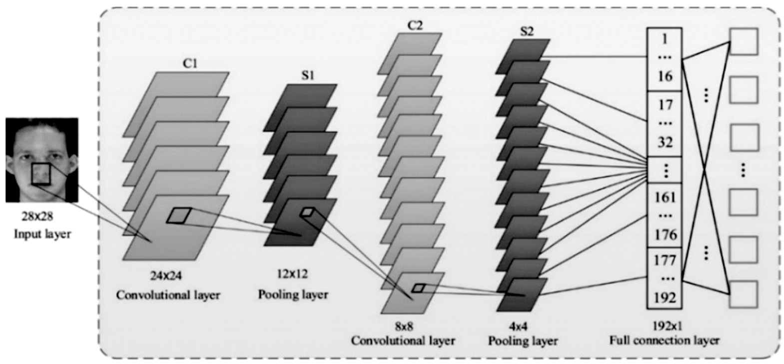

Figure 3. Face recognition using CNN[5]

From 2012 onwards, following the success of the AlexNet architecture model from Professor Hinton's research group, which won the ImageNet competition with a large margin over the following positions and broke the previous stereotypes about the features of machine learning models compared to the features of traditional methods, it opened up the development and application of CNN network models in the field of artificial intelligence. With up to 60 million parameters and many hierarchical layers, each layer will have its own ability to extract unique facial features, and there will be pooling layers and activation functions to speed up the computation process. Later model architectures have been improved to address the issues of previous architectures. Currently, research centers around the world continue to improve and publish many new model architectures.

The main technical solution to the problem is choosing a CNN model for facial recognition that meets the requirements. The state-of-the-art (SOTA) CNN models are those with high accuracy and good performance in solving image processing problems, especially facial recognition. These models are usually trained on large and diverse datasets, capable of extracting deep and discriminative facial features, and can be applied to various situations [8]. In building a practical facial recognition application, using SOTA CNN models has many benefits, such as:

- Increasing the accuracy and efficiency of facial recognition applications compared to traditional methods or older CNN models.
- Reducing time and effort in designing and training a CNN model from scratch by using pre-trained SOTA CNN models or fine-tuning them for specific problems.
- Addressing difficult issues in facial recognition, such as changes in viewing angle, lighting, expression, age, gender... of the face.
- Therefore, to solve our problem, we focus on SOTA CNN models. Highly rated and effective deep learning models currently include Google's FaceNet, Oxford's Resnet-50 trained on the VGGFace2 dataset, CMU's OpenFace, Facebook's DeepFace, and VGGFace by the VGG group. To choose a model, the research team will focus on comparing the performance of FaceNet and VGGFace based on research results [9] to test with the designed hardware system.
- FaceNet and VGGFace are two highly-rated SOTA CNN models in terms of accuracy and performance in facial recognition.
- FaceNet is a CNN model trained to create vector representations for faces, such that the Euclidean distance between vectors of the same person is small, while the distance between vectors of different people is large [8, 10].
- FaceNet represents a state-of-the-art neural network for facial recognition, verification, and clustering. This 22-layer deep neural network is designed to train its output directly as a 128-dimensional embedding. The loss function employed in the final layer is known as triplet loss. FaceNet consists of the aforementioned components, which we will examine sequentially [11].

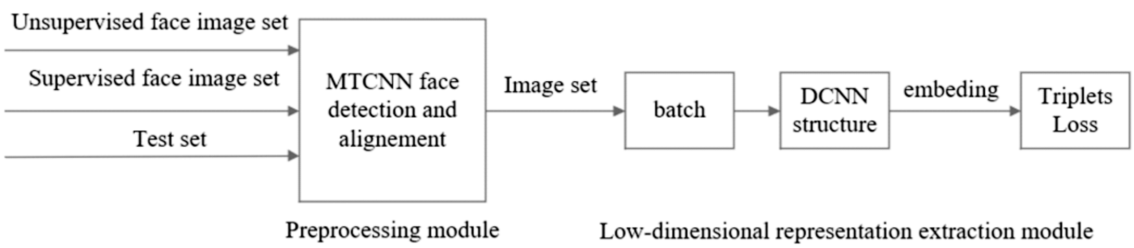

Figure 4. The structure of the FaceNet architecture [11]

The most crucial aspect of the FaceNet model is the end-to-end learning of the entire system. Triplet loss enables a direct reflection of our objectives in facial verification, identification, and classification. The triplet loss of [12] aids in projecting all faces with a similar identity to a single point within the embedding space. Fig. 4 illustrates the structure of the FaceNet architecture.

VGGFace is a deep learning neural network model for facial recognition, researched by the Visual Geometry Group (VGG) from the University of Oxford and built upon the VGGNet architecture, a famous CNN architecture in the field of computer vision. VGGFace is trained on a large dataset of labeled faces and can recognize faces belonging to many different individuals. This model can generate scalar facial representations for each face, making the process of identification and classification of faces faster and more accurate.

The VGGFace model, named later, was described by Omkar Parkhi in the 2015 paper titled "Deep Face Recognition" [13]. The VGGFace model uses the VGGNet architecture with blocks of layers with small kernel sizes, ReLU activation functions, followed by Max Pooling layers, and finally, Fully Connected layers for classification. The VGGFace dataset consists of 2.6 million images containing faces of 2.6 thousand people. The dataset is used to develop CNN models for facial verification and classification tasks. Specifically, popular models are trained on the dataset and evaluated to determine the best state-of-the-art (SOTA) model.

Building upon the large facial dataset of VGG, VGGFace2 was introduced in 2017 as a larger facial dataset, including 3.31 million images of 9131 subjects, with an average of 362.6 images for each subject. The images were collected from Google and exhibit diversity in age, appearance, ethnicity, occupation, and pose [14]. Unlike VGGFace, which uses a CNN architecture based on VGG, VGGFace2 employs the ResNet-50 or SqueezeNet-ResNet-50 models. These models have been trained on the VGGFace2 dataset and have achieved state-of-the-art performance.

Both models can achieve high accuracy on standard datasets such as LFW, YTF, IJB-A... These models can be applied to various facial recognition problems, such as face verification, face recognition, face classification, and face detection. Therefore, testing these two models can help assess the performance of security control software in various situations. In addition, these models can work with different types of data, not limited to high-quality facial images. FaceNet can handle facial images with variations in viewing angle, lighting, and expression... VGGFace can handle facial images with differences in age, gender, and ethnicity... Thus, testing these two models can help assess the flexibility and sustainability of the software in managing access control for military units.

Data collection method: In this study, we focus on testing methods for face detection, feature extraction, and face classification. The data used for testing is collected by the research team from facial images of colleagues to train and test. For each person, an IMX219 8MP camera is used to capture facial images from various angles, states, and lighting conditions (Fig.5). The proposed hardware configuration is the NVIDIA Jetson Nano B1 embedded computer (4-core ARM A57 CPU, 128-core Maxwel GPU, 4 GB Memory).

Figure 5. Each face is collected from various various angles, states, and lighting conditions

2.2. Functions of modules

As presented in section 2.1.1, with advanced and efficient features, this software system is a perfect solution for units that want to ensure security and manage access intelligently and innovatively. The main functions of the software subsystem include (Fig. 6):

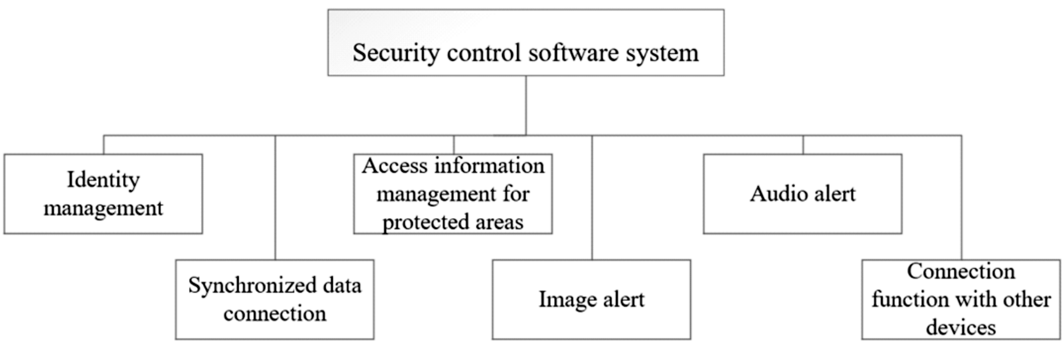

Figure 6. The functional decomposition model of the software system

Identity management: allows adding new identities, deleting existing identities, changing identity information, and updating identity image changes (Fig.7).

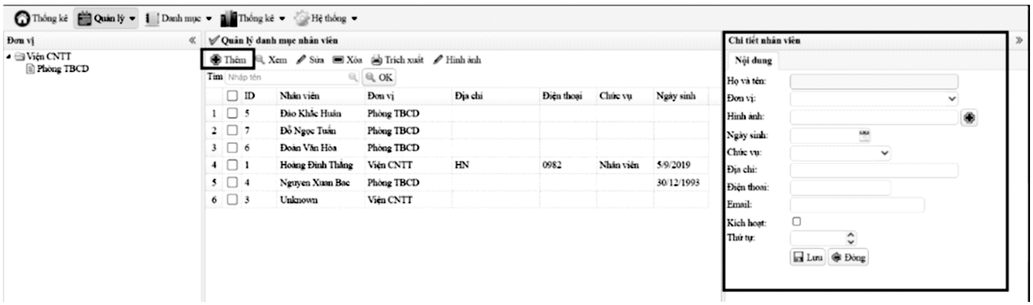

Figure 7. Screen interface of the identity management function

Access information management for protected areas: allows automatic updating of entry and exit information, checking entry and exit information by time, checking entry and exit information by object, printing entry and exit reports by time, and printing entry and exit reports by object (Fig.8).

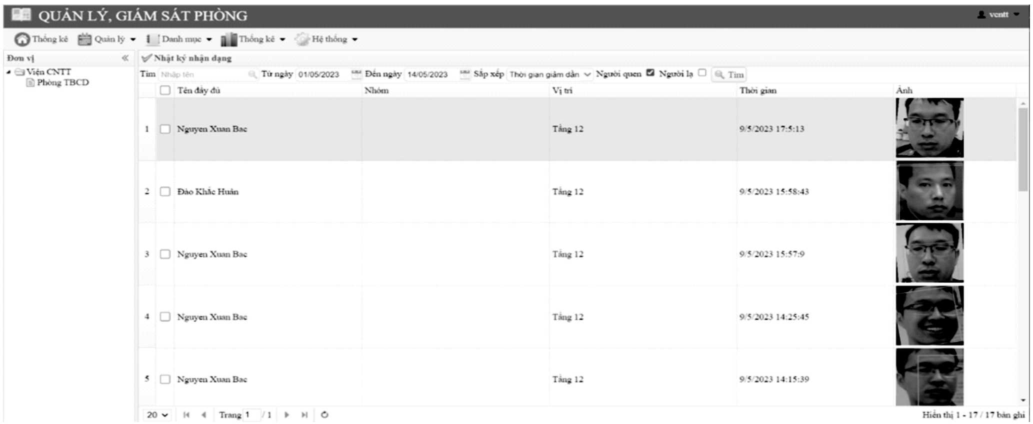

Figure 8. Screen interface of the Access information management for protected areas

Audio alert function: allows setting up an audio alert mode when an unidentified person is detected, and setting up alert disabling.

Image alert function: allows displaying personal login information on the connected screen (Fig.9).

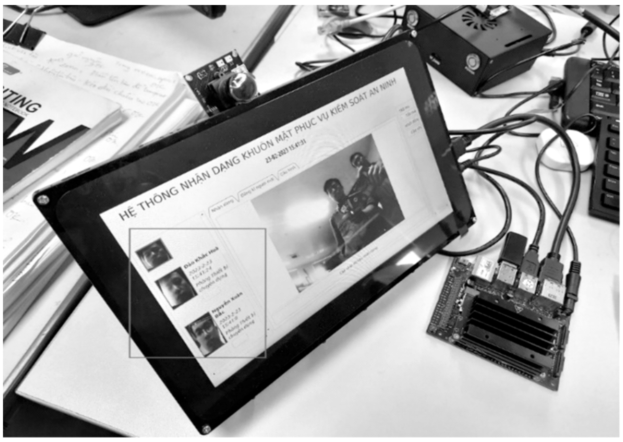

Figure 9. Screen interface of the image alert function

Connection function with other devices: allows adding new connections from devices, deleting connections with devices, and updating connection status with devices. Synchronized data connection function: allows updating connection status with devices, checking connections, and extracting identity data.

2.3. Test results

2.3.1. The CNN SOTA model

We tested two face recognition models, FaceNet and VGGFace, on the proposed hardware, NVIDIA Jetson Nano B1, with an IMX219 8MP camera module. Both models achieved high accuracy on the LFW dataset, a standard dataset in the field of face recognition. In addition, these models can also recognize the faces of colleagues around them after being trained on images taken from various angles, states, and different lighting conditions. This demonstrates the flexibility and robustness of the two models in handling facial variations. Accuracy is calculated as the ratio of the number of correctly identified face image pairs to the total number of face image pairs tested. Speed is calculated as the average time to recognize a pair of face images. The test results are shown in the following table:

Table 1. Test results of two face recognition models on the proposed hardware

Model	Accuracy (%)	Speed (seconds)
FaceNet	99.63	0.6
VGGFace	98.95	1.2

Thus, the experiment on the proposed hardware shows that the FaceNet model meets the accuracy and speed requirements for practical applications.

2.3.2. The software

Currently, the research team is in the testing phase of this software system at the Combat Duty Room of the Command Headquarters of the Military Command of Ba Ria-Vung Tau Province, and has obtained certain results. The software has demonstrated the ability to accurately identify objects entering and exiting through identity management categories; automatically and continuously manage information entering and exiting protected areas; sound and image alerts when unidentified individuals are detected; set up connections with other devices to efficiently collect and process data; and set up synchronized data connections to help users save time and effort in information management. The software has also received satisfaction and trust from users in managing the entry and exit of the unit.

Compared to traditional methods of managing entry and exit of the unit, such as notebooks and observing cameras with the naked eye, the security control software system of the unit is a newer and more advanced solution, with outstanding advantages such as accuracy, safety, convenience, and efficiency in managing, monitoring, and supervising entry and exit in critical areas of the military.

3. CONCLUSION

In this article, we have introduced the security control software system of the local military unit. This software uses the SOTA CNN network to recognize faces by testing the two best face recognition models currently available, FaceNet and VGGFace. The tests on the proposed hardware show that the FaceNet model meets the accuracy and speed requirements for practical applications. This article has presented in detail the functions of the security control software and evaluated the expected effectiveness of the software in the actual testing process. Currently, the product is in the process of continued testing and improvement; after completion, this system will be applied in many local military units in the southern region.

While the article focuses on the security control software system for military units, there are potential broader implications for society as a whole. The use of advanced face recognition technology can improve security measures in various settings, such as airports, government buildings, and public events. However, it is important to consider the ethical implications of such technology, including privacy concerns and potential misuse.

In terms of future research and development, there is potential for civilian use of the system. For example, the software could be adapted for use in private security systems or for access control in corporate settings. Additionally, further research could explore the use of other advanced technologies, such as biometric authentication or machine learning algorithms, to enhance security measures.

In general, the security control software system discussed in this paper has the capability to considerably enhance security measures in both military and civilian environments. Further research and development will be required to guarantee the ethical and efficient utilization of this technology.

ACKNOWLEDGEMENT

This article is part of a project funded by the Department of Science and Technology of Ba Ria - Vung Tau province.

REFERENCES

[1] S. Kumar, D. Kathpalia, D. Singh, M. Vats, "Convolutional Neural Network based Automated Attendance System by using Facial Recognition Domain," in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, IEEE, 2020.

[2] K. G. Saravanan, J. Rani, D. C. Jullie Josephine, M. Parameswari, Hridya Venugopal, J. Ramal, "Deep Learning based Facial Recognition System for Attendance Maintenance", in *2022 International Conference on Edge Computing and Applications (ICECAA)*, IEEE, 2022.

[3] M. T. H. Fuad et al., "Recent Advances in Deep Learning Techniques for Face Recognition," *IEEE Access*, vol. 9, 2021.

[4] M. Benco, T. Mizdos, R. Radil, Roman, "A New Method for Face Recognition Using Convolutional Neural Network," *Advances in Electrical and Electronic Engineering*, vol. 15 (4), 2017.

[5] P. Lu, B. Song, L. Xu, "Human face recognition based on convolutional neural network and augmented dataset", *Systems Science & Control Engineering*, vol. 9 (sup2), 2021.

[6] Y. Zhang, Y. Liang, J. Wang, "Face Recognition System Based on CNN", in *2020 International Conference on Computer Information and Big Data Applications (CIBDA)*, IEEE2, pp. 44-47, 2020.

[7] Z. Xie, J. Li, H.Shi, "A Face Recognition Method Based on CNN", *Journal of Physics: Conference Series*, vol. 1368(3), 2019.

[8] A. Sajjanhar, Z. Wu, Q. Wen, "Deep Learning Models for Facial Expression Recognition", in *2018 Digital Image Computing: Techniques and Applications (DICTA)*, IEEE, 2019.

[9] S. I. Serengil, A. Ozpinar, "LightFace: A Hybrid Deep Face Recognition Framework", in *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, 2020.

[10] S. Florian, D. Kalenichenko, J. Philbin, "Facenet: A unified embedding for face recognition and clustering", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015.

[11] S. Ammar, T. Bouwmans, N. Zaghden, M. Neji, "Towards an Effective Approach for FaceRecognition with DCGANs Data Augmentation", in *15th International Symposium on Visual Computing*, 2020.

[12] Sun Y., Chen Y., Wang X., and Tang X, "Deep learning face representation by jointidentification-verification", *NIPS'14:International Conference on Neural Informa-tion Processing Systems*. 2:1988–1996, December 2008.

[13] O. M. Parkhi, A. Vedaldi, A. Zisserman, "Deep Face Recognition". *In Xianghua Xie, Mark W. Jones, and Gary K. L. Tam, editors, Proceedings of the British Machine Vision Conference (BMVC)*, pages 41.1-41.12. BMVA Press, 2015.

[14] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, A. Zisserma, "VGGFace2: A dataset for recognising faces across pose and age", in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition*, IEEE, 2018.

Ứng dụng trí tuệ nhân tạo xây dựng hệ thống phần mềm kiểm soát an ninh ở các đơn vị bộ đội địa phương

Nguyễn Trọng Thể, Nguyễn Khắc Điệp và Dương Xuân Trà

TÓM TẮT

Bài báo này giới thiệu việc ứng dụng trí tuệ nhân tạo để xây dựng hệ thống phần mềm kiểm soát an ninh của đơn vị bộ đội địa phương. Hệ thống phần mềm này sử dụng mạng nơron tích chập (CNN) SOTA để nhận diện khuôn mặt bằng cách thử nghiệm hai mô hình nhận diện khuôn mặt tốt nhất hiện nay là mô hình FaceNet và mô hình VGGFace. Qua thử nghiệm trên phần cứng đề xuất cho thấy mô hình FaceNet đáp ứng được độ chính xác và tốc độ cho ứng dụng thực tế. Phần mềm này bao gồm nhiều hạng mục quản lý định danh để đảm bảo an ninh thông tin và giám sát ra vào của bộ đội và các đối tượng khác. Ngoài ra, phần mềm còn bao gồm các tính năng quản lý thông tin ra vào khu vực bảo vệ, cho phép cảnh báo âm thanh và hình ảnh để đảm bảo an toàn và bảo mật cho khu vực đó. Phần mềm cũng cho phép người dùng cài đặt kết nối với các thiết bị khác để thu thập và xử lý dữ liệu hiệu quả. Đồng thời, nó cũng hỗ trợ kết nối đồng bộ dữ liệu để giúp người dùng tiết kiệm thời gian và công sức trong việc quản lý thông tin. Hệ thống phần mềm này cung cấp một giải pháp toàn diện và hiệu quả để đảm bảo an ninh và giám sát ra vào tại các đơn vị bộ đội địa phương. Bằng cách tận dụng trí tuệ nhân tạo, hệ thống có thể thích ứng và cải tiến theo thời gian, mang lại hiệu suất và khả năng nâng cao để đáp ứng nhu cầu an ninh đang phát triển của các tổ chức quân sự. Hơn nữa, phần mềm có giao diện thân thiện với người dùng và cho phép tuỳ chỉnh, đảm bảo dễ sử dụng và thích ứng với nhu cầu cụ thể của từng đơn vị. Hệ thống phần mềm này cung cấp một giải pháp toàn diện và hiệu quả để đảm bảo an ninh và giám sát ra vào tại các đơn vị bộ đội địa phương. Nhờ ứng dụng trí tuệ nhân tạo, hệ thống có thể thích ứng và cải tiến, đáp ứng nhu cầu kiểm soát an ninh ở các đơn vị bộ đội. Phương pháp cải tiến được trình bày trong bài báo này có thể cải thiện đáng kể tình hình kiểm soát an ninh của các đơn vị bộ đội địa phương, góp phần đảm bảo an toàn cho cả quân nhân và nhân dân.

Từ khóa: *trí tuệ nhân tạo, CNN SOTA, kiểm soát an ninh, nhận diện khuôn mặt*

Received: 01/06/2023
Revised: 14/06/2023
Accepted for publication: 15/06/2023