SMART VIDEO DETECTION: DEEP LEARNING FOR ENHANCED HOME SECURITY

Dr. A Christy Jeba Malar
Associate Professor,
Department of Information Technology,
Sri Krishna college of Technology,
Coimbatore,India
a.christyjebamalar@skct.edu.in

Dineshkumar M
Student
Department of Information Technology,
Sri Krishna college of Technology,
Coimbatore,India
dineshkumarblack@gmail.com

Gokul A
Student
Department of Information Technology,
Sri Krishna college of Technology,
Coimbatore,India
gokulgs6480@gmail.com

Hemanthkumar M G
Student
Department of Information Technology,
Sri Krishna college of Technology,
Coimbatore,India
hemanthmg1406@gmail.com

***Abstract*— Home video recognition autonomous systems based on traditional approach for face image analysis are inherently unable to provide the required real time/intelligent analytics to adapt to contemporary security threats. This project is centered around utilizing smart technology to bolster home security, more specifically with the help of smart video surveillance. This is the system that learns to recognize the faces and objects in home video to alert homeowners when an intruder enters, but without compromising the privacy of its residents. This has initiated a truly human revolution - introducing deep learning algorithms to enable machines to identify intricate patterns from optical data streams. The purpose of this paper is to provide an insight into the disruptive value of deep learning based algorithms including CNN, ResNet50, VGG16 in the field of home security, exposing the limitations of traditional face image analysis techniques and explaining the revolutionary power of new machine learning approaches. Model accuracy in Normal CNN almost up to 99 percent VGG16 Algorithm too up to 95 percent. This vision recognition system is a substantial leap in improvement in security and surveillance.**

Keywords— Deep learning, CNN, ResNet50, VGG16

I. INTRODUCTION

Consider a smart security system that knew the difference between your family and an intruder. Artificial intelligence-driven facial recognition identifies household residents aware of the system. Having the latest or advancements in computing technology like the CNN, VGG16, ResNet50 but in optical science or in brief, we could do on-device person identification whom it doesn't know i.e, may be a suspicious person who is trying to break-in.

It has been designed to help you protect your house without bothering your family. They have found that the photos it takes assist in the event of a burglary. Grayscale images were used to simplify the data and make the system efficient. These are intensity-only images without colors, which allows faster processing and reduces the risk of errors due to too much detail. Black and white images also means less space in memory as well as storage compared to colour ones.
VGG16 is a simple and powerful neural network design that we are exploiting into our project. This architecture will allow extracting fine-grained visual features from video clip and making accurate face recognition and object detection. Its simplicity means easier to understand and interpret the learned features, a must in the Security Monitoring for accurate detections.

This is accomplished by feeding the familiar faces (e.g family members) into our network and the unfamiliar ones, potential intruders, are also passed into a Convolutional Neural Network (CNN) running on our embedded hardware. These CNNs can learn extremely complex patterns and are what allow our system to raise alerts selectively only for unseen faces.

II. RELATED WORK

Arpit Singh et al.(2023)[1] - The objective of the study is to identify the potential of machines, automatically (facial recognition or computer vision) recognizing and detecting people in live feeds or previously recorded surveillance feeds. With aid of state of the art, neural network algorithms and convolutional neural networks (CNNs) and the VGG Face deep learning architecture developed a real-time system. The

model was then re-trained on a smaller dataset of 26 individuals and 7500 images using transfer learning. In this paper, the authors describe the construction of the dataset and experiments with different machine learning and deep learning methods tailored to improve recognition accuracy. In real-time, they stated that the system identified 26 different individuals with confidence levels ranging from 78.54% of impressions to a perfect 100%; Overall referred to as a 96% level of accuracy.

Ilyas et al. (2019)[2] Facial Recognition System Ilyas Et Al. trained deep learning neural networks for facial recognition They first used Viola-Jones face detection technique to recognize facial regions from the input images and afterward used Histogram Equalization Algorithm (AHE) to improve the quality of the images. In turn, they used Convolutional Neural Networks with the architectures: VGG16 and ResNet50 to extract face features and human face classification. In addition, their proposed methods also outperform other state-of-the-art methods on high dimensional face data sets such as Extended Yale B Face and CMU PIE.

Shakeel and Lam (2019)[3] - Age-Invariant Face Recognition: Shakeel & Lam proposed an age-invariant confront acknowledgment approach that utilized discriminative demonstrate with CNs show of AlexNet as exchange based learning which is a CNN. They attempted to learn exceptionally high-level profound highlights encoded into identity-specific code words that are reliable but one of a kind for all sensor-resolved confront pictures of one single person captured in distinctive times. Its objective was to make a codebook representation which ought to lead to strong acknowledgment over time. Their approach tried with a straight relapse based classifier over a wide extend of datasets, counting the FGNET illustrates the noteworthy advancements that can be gotten by a considering how age extend varieties influence the execution of confront acknowledgment strategies.

Peng et al. (2019)[4] - Peng et al. designed an end-to-end local descriptor learning model for cross-modality face recognition in which compact local discriminative information could be learned from raw facial patches. Their approach, which employs CNNs for the purpose of producing deep local descriptors, achieves a remarkable performance of 98.68% accuracy on average over six frequently used face recognition benchmark datasets across multiple protocols. They showed that their method is effective for their research both with different affect-based facial data and among multiple datasets and modalities.

Harikrishnan et al. (2019)[5] - View of Face Recognition for Attendance System Automation H. Harikrishnan, et.al. launched the recognition technology to the face especially in the cyclic jobs will overlook the usual e.g. attendance system. Applied as a process which uses Artificial Neural Networks (ANNs) on a substantial dataset of images as input to learn to accurately and consistently identify and locate faces in an image. The findings of their work classified the research under the four classes of Detection, Training, Recognition and Data management which was therefore focused on preprocessing modules of resize, enhance and noise filtration with grayscale image for enhanced imagery recognition during attendance-taking as well.

III. Proposed System

It is a proposed system that indicates a revolutionary change in the sector, which is a system that takes advantage of intelligent video recognition. It combines the latest in deep learning with the best of old fashioned security tech, to reinvent home security. The technology in it is that it uses a facial recognition in an advanced way, some complex algorithm or so, in which we can make out that if the person is present at home or unknown one. This selective alerting system contradicts conventional security systems, which alert the user for every facial detection that occurs, including family returning home, etc. In addition to this, the use of Convolutional Neural Networks(CNNs) such as VGG16, ResNet50 etc allow to process the surveillance footage more quickly, with high accuracy and efficiently. The system is specifically privacy-centric, for the monitoring to be targeted at possible threats and treat the privacy of the residents of the household.

IV. Architecture

The neural network architecture of this proposed system is given below. It is taking dataset as input and preprocessing is done. The processed data is given as input to the CNN, ResNet, VGG16 architecture. Figure 1 shows the architecture diagram.

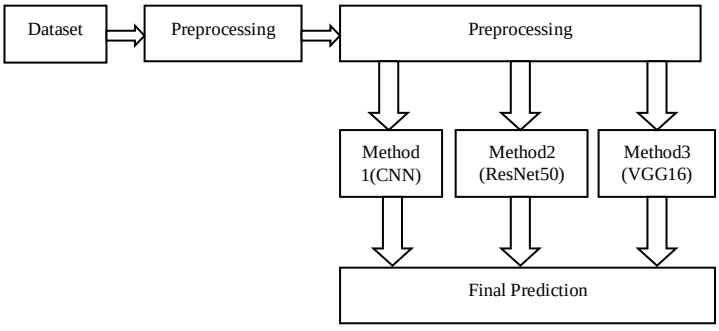

Fig. Architecture Diagram

The flow chart of this project is as follows in Figure 2. It has two phases face registration and video detection

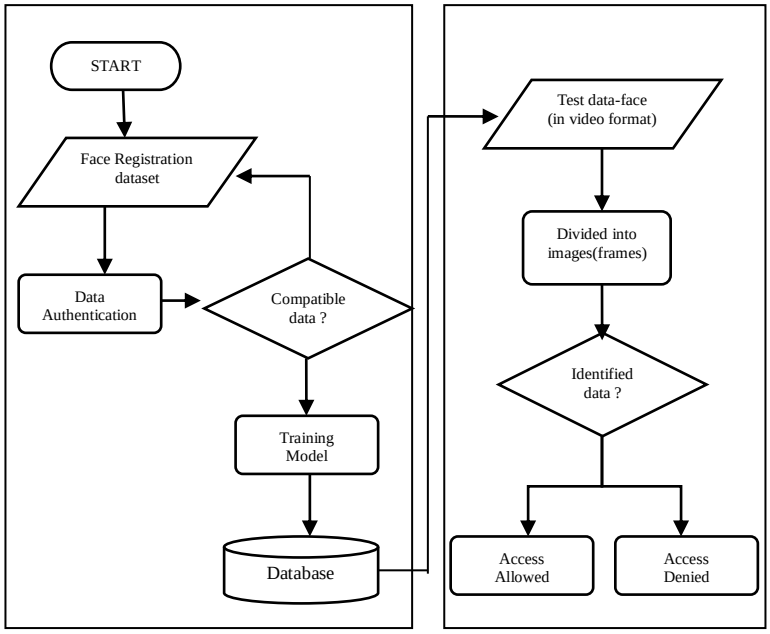

Fig.2 Data Process

V. METHEDOLOGY AND DISCUSSION

A. Importing Dataset

This is in contrast to using the datasets already existing Online, where this system has created an own dataset on the fly. Using the real time photos dataset of persons for this project The dataset included photos capturing faces and expressions of people in real-time. It is a collection with rich, high-resolution images that reveal these dynamic personalities and individual faces in an endless variety of settings. Our dataset consists of updated snapshots demonstrating the rich diversity of poses and emotions inferred over to a range of scenes in everyday lifeThe dataset consists of each of the images capturing patterns of the faces and group individualities which provides an overall picture of their physical appearances. It contains a wealth of data on faces in real time, making it an ideal testbed for systematic study, the paper says, of team members' facial characteristics — an important problem in many areas of facial analysis and recognition. Photos of each human being are stored in separate folder on google drive platform.

This project integrates Google Drive into the Google Colab notebook interface to simplify file management. we can also make Google Drive files/applications a part of the environment in Colab and perform file operations to those files as if they were stored in the native disk. This improves the productivity and convenience to directly interact the data sets, code scripts and resources from Google Drive that are the important piece of the project, thus optimizing the workflow efficiency with in the project.

B. Data Augmentation

Data augmentation: Data Augmentation is going to be applied to the data imported using the methods are important techniques to increasing the performance and generalization capability of machine learning models, especially in cases where little data is available or when overfitting is likely. Data augmentation generates different variations of the training set based on methods such as rotation, flip, scaling and cropping in order to prevent the model from memorizing the same n number of examples and improve its adaptability to the real world variations. This augmentation process helps to not only broaden the dataset but also to help the model to be robust by allowing the model to learn the representation by observing various cases/conditions.

C. Grayscale Preprocessing

Grayscale preprocessing in this wander system is one of the basic steps for course of action of picture data from the data expanded pictures. In this step the handle of alter of pictures from color to dim and white is done .The alteration that was connected to the data expanded the set of pictures to the set of greyscale images[6]. This adjustment has driven to a perish of the number of channels from the three to one fine set of objects. Such viewpoint can be considered advantageous for applying the given pictures for the various assignments, such as picture acknowledgment and protest location. The given preprocessing organize compares with the reason of making utilize of VGG16 and the other convolutional neural systems, makes a difference to extricate the complex visual highlights. In common, grayscale preprocessing plays the part of the storm cellar for the productive and exact investigation.

D. CNN Model

Convolutional Neural Network CNN[7] for detecting and recognition of family members from video, crafted strategically. An example of this is the CNN[8] architecture consisting 2 stages: convolutional layers for first stage feature extraction and subsequent dense layers for classification task designed to poll for intricate patterns and characteristic features of each constituent family member. The network learns hierarchical visual features through the convolutions and max poolings, starting from basic edges to advanced facial features. These learned features are then used by final dense layers, with ReLU activations and dropout regularization for accurate classification among the assumed 3 family members. By contrast, the CNN[9] model is designed to accommodate a robust difference or categorization of a family member in the video, which can be employed for recognition or security systems which require a critical identification of a particular person for a superior monitoring or access control. Figure 3 shows the accuracy of trained CNN model is 99 precent.

.

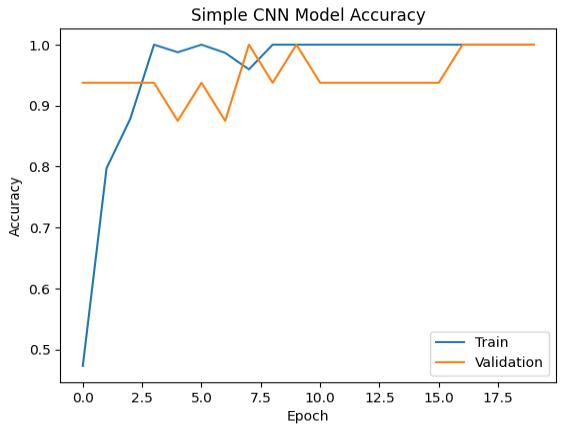

Fig. 3 CNN model Accuracy

E. *VGG 16*

VGG16[10] is a deep convolutional neural network trained on large ImageNet dataset and is being used as a feature extractor here. The convolutional layers are designed to extract complex visual patterns and edges, textures from large images In this hands-on example, the goal is to take advantage of these learned features and turn them into a means of recognizing family members by means of transfer learning. VGG16[11] is strong in a way that it is very deep such that it can learn layers of features. This capability likely makes it better at leveraging the more complex visual information needed to tell family members apart. In Comparison, Convolutional layers essentially act as Feature detectors in the model and helps the model to learn important Discriminant features needed to classify accurately.. The depth and strong learning of hierarchical representations by the model help it to distinguish complex visual features required to identify members of a family in the video. While preserving the fixed convolutional base to retain these learned representations, custom classification layers are added to tailor the model specifically to the task of identifying among the assumed three family members. The project capitalizes on VGG16's robust features to achieve high accuracy in family member recognition, mitigating overfitting by fine-tuning only the added classification layers. The VGG16 also gives an accuracy more than 95 percent as shown in figure 4.

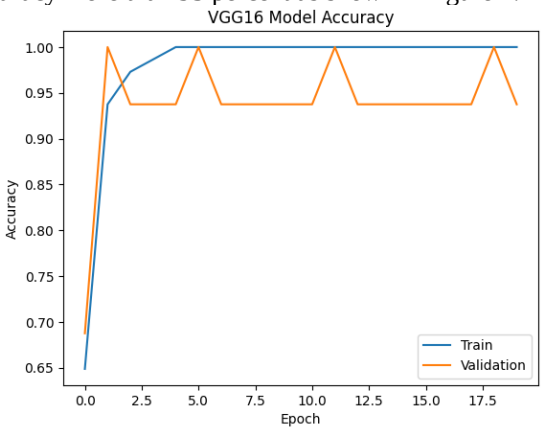

Fig. 4 Accuracy of VGG16 Model

F. *ResNet 50:*

These images are passed through the deep convolutional neural network architecture: ResNet50[12], which has a depth of 50 layers within the Graphic Processing Unit (GPU) so they have the ability to understand even the fine-grained categorization of features of images. It has been trained on the massive ImageNet dataset and learned how to recognize different types of visual patterns, including edges, textures, and other higher-level features found in videos. The project helps to prevent this overfitting to some extent in a small family member dataset by using pre-trained features, thus allowing better generalization to unseen data. ResNet50 has an accuracy of 70 to 80 percent at best as shown in figure 5.

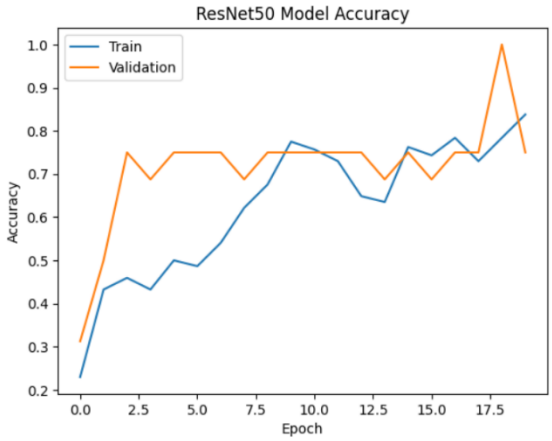

Fig. 5 Accuracy of ResNet50 Model

G. *Video Detection:*

As a result of above results trained CNN and VGG16 models which have performance results as multiple over. The input for this code is video and we then import libraries like OpenCV(cv2), Matplotlib(plt), NumPy(np), pre trained model from TensorFlow Keras and the video is now broken into frames in this code. All videos are loaded using the load files class and each video is iterated over the frames ,reads and processes each frame and displays the frame along with the obtained(class label) using Matplotlib and then releases the video object once all frames are processed and displaying the output as "Access Granted" or "Access Denied" foreach frames.

H. *Sample Output:*

Case 1: If the video has a intruder or not a housemate, the person will not be allowed inside the home as shown in figure 6.

Fig. 6 Stranger- Access Denied.

Case 2: If the video has a Housemate, the person is accepted insoide the house.

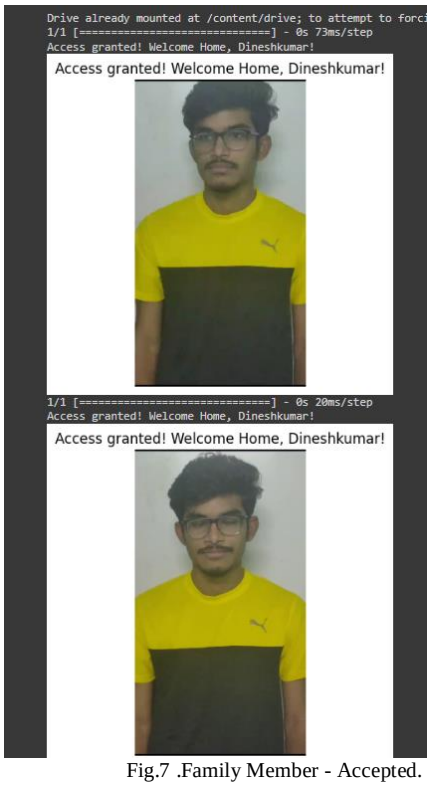

Fig.7 .Family Member - Accepted.

VI. Result Analysis

The project uses live photo data sets that can record expressive and motioned expressions to produce a reliable facial recognition system using deep learning techniques [13].

The data management is optimized by making use of Google drive for storage and Google colab for handling the files. Data is then augmented using several techniques of rotation, flipping, scaling, and cropping, so that the dataset is more effective for the model, and the model learns the variations of the same data improving its performance and also provides it the generalization capabilities to predict the unseen data. Greyscale preprocessing the images from the data collected to get black and white images also means less space in memory as well as storage compared to colored ones. The CNN model is used to recognize family members from video as shown in figure 7, and it is based on convolutional layers for features extraction and dense layers for classification. In particular, the recognizing has more than 99% accuracy on test data. Video detection combines a number of models with OpenCV and other libraries to handle frames in a video and determine what objects in them are, exactly, in real time. In security implementations, the model is used to grant access if the person is recognized as a family member or not. The CNN and VGG16 models, is used to analyze the frames and ensure the highest level of confidence in security tasks.

Accuracy is calculated by

$$\text{Accuracy} = \frac{TP}{(TP+TN)*100}$$

TP → True Positive represents the positive class values that the model predicts accurately.
TN → True negative represents the value that the model predicts the negative class.

Epoch	Loss	Accuracy	Validation Loss	Validation Accuracy
1	3.874	56.631	5.75329	93.657
3	3.459	84.362	3.94532	86.346
7	2.246	94.875	2.73245	85.672
15	0.092	97.638	0.24675	95.627
18	0.08	99.165	0.98244	99.252

Table 1 shows the statistical results from model

VII. Conclusion

The paper introduces a novel proposal in the field of Home Security and Surveillance using the cutting edge deep learning techniques for recognition family member that opens the door to efficient and customized security. The project shows that it is entirely possible and practical to tell whether two images have family members using the current convolutional neural network architecture such as VGG16[14] or ResNet50[15].

These models allow us to not only greatly reduce its model development time but also improve models by using the representations that were learned on other large datasets such as ImageNet. This and the presented inference script in the following gazebo model reinforce the capabilities of the models for actual real-time identification and access control in a home security system. The model allows recognized family members to acquire access abilities while preventing unknown individuals from gaining them, therefore, it shows that the model, can be further used to provide a personalized access control and enhance the security. The demonstrated precision, robustness, and flexibility achieved in the project demonstrate a direction forward to implement this system into real-world surveillance and security application in a scalable and cost-effective way while maintaining a high level of security measures, shareability and customizability for the end-users.

VIII. ACKNOWLEDGMENTS

We would like to express our heartfelt gratitude to our esteemed college and our department for providing us with necessary lab facility and instrument that helped us to accomplish the research. Moreover, we express our earnest thankfulness to my honourable professor who provided us with valuable insights and direction for the completion of this project.

IX. REFERENCES

[1] Singh, A., Bhatt, S., Nayak, V., & Shah, M. (2023). Automation of surveillance systems using deep learning and facial recognition. International Journal of System Assurance Engineering and Management, 14(Suppl 1), 236-245.

[2] Ilyas BR, Mohammed B, Khaled M, & Miloud K (2019) Enhanced face recognition system based on deep CNN. In: Proceedings - 2019 6th International Conference on Image and Signal Processing and Their Applications, ISPA 2019.

[3] Shakeel, M. S., & Lam, K. M. (2019). Deep-feature encoding-based discriminative model for age-invariant face recognition. Pattern Recognition, 93, 442-457.

[4] Peng, C., Wang, N., Li, J., & Gao, X. (2019). DLFace: Deep local descriptor for cross-modality face recognition. Pattern Recognition, 90, 161-171.

[5] Harikrishnan, J., Sudarsan, A., Sadashiv, A., & Ajai, R. A. (2019, March). Vision-face recognition attendance monitoring system for surveillance using deep learning technology and computer vision. In 2019 international conference on vision towards emerging trends in communication and networking (ViTECoN) (pp. 1-5). IEEE.

[6] Smith, K., Landes, P. E., Thollot, J., & Myszkowski, K. (2008, April). Apparent greyscale: A simple and fast conversion to perceptually accurate images and video. In Computer graphics forum (Vol. 27, No. 2, pp. 193-200). Oxford, UK: Blackwell Publishing Ltd.

[7] Ahmed, T., Das, P., Ali, M. F., & Mahmud, M. F. (2020, July). A comparative study on convolutional neural network based face recognition. In 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-5). IEEE.

[8] Coşkun, M., Uçar, A., Yildirim, Ö., & Demir, Y. (2017, November). Face recognition based on convolutional neural network. In 2017 international conference on modern electrical and energy systems (MEES) (pp. 376-379). IEEE.

[9] Smith, P., & Chen, C. (2018, December). Transfer learning with deep CNNs for gender recognition and age estimation. In 2018 IEEE International Conference on Big Data (Big Data) (pp. 2564-2571). IEEE.

[10] Perdana, A. B., & Prahara, A. (2019, November). Face recognition using light-convolutional neural networks based on modified Vgg16 model. In 2019 International Conference of Computer Science and Information Technology (ICoSNIKOM) (pp. 1-4). IEEE.

[11] Dubey, A. K., & Jain, V. (2020). Automatic facial recognition using VGG16 based transfer learning model. Journal of Information and Optimization Sciences, 41(7), 1589-1596.

[12] Li, B., & Lima, D. (2021). Facial expression recognition via ResNet-50. International Journal of Cognitive Computing in Engineering, 2, 57-64.

[13] Mehdipour Ghazi, M., & Kemal Ekenel, H. (2016). A comprehensive analysis of deep learning based representation for face recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 34-41)

[14] Cheng, S., & Zhou, G. (2020). Facial expression recognition method based on improved VGG convolutional neural network. International Journal of Pattern Recognition and Artificial Intelligence, 34(07), 2056003.

[15] Elpeltagy, M.,& Sallam, H. (2021). Automatic prediction of COVID− 19 from chest images using modified ResNet50. Multimedia tools and applications, 80(17), 26451-26463.