Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 7-Issue 1, January 2024
Paper: 74

Person Identification from Video with Image Enhancement using Hybrid Approach

Aman Nawaz Manjith
Department of Computational Intelligence
SRM Institute of Science and Technology, Chennai, India,
am5937@srmist.edu.in

Dr. S. Karthick *(Associate Professor)
Department of Computational Intelligence
SRM Institute of Science and Technology, Chennai, India,
karthiks2@srmist.edu.in

Vikranth Bandaru
Department of Computational Intelligence
SRM Institute of Science and Technology, Chennai, India,
vb7224@srmist.edu.in

***Abstract*— With the advancement in technologies, it is now possible to use enormous image databases more and more frequently. Person identification is a useful technique for managing and retrieving facial data from a video. Machine Learning algorithms such as Haar Cascade algorithm is used to scan and capture the images, while picture annotation plays an essential function in a pre-processing phase of images having a purpose of dividing the picture into components or areas of interest for a deeper examination of one or more of these parts. YOLO is the most common model architecture and object detection technique. It employs one of the finest neural network designs to provide excellent accuracy and overall processing speed, which is the key reason for its popularity. Image enhancement is an extra process involved in increasing the accuracy of the output. Real-ESRGAN employs a new design that incorporates a feature extraction network and a reconstruction network. The feature extraction network is designed to extract high-level features from low-resolution images, while thereconstruction network uses the extracted features to generate high-resolution images. Through this, the computer will be able to identify the face of the person with an even better accuracy rate.**

I. INTRODUCTION

Person Re-identification is a demanding job in machine vision that includes identifying a person across various non-overlapping cameras. Existing person identification systems have limitations such as changes in illumination, occlusion, pose variations, and low resolution, which can result in misidentification and hinder the effectiveness of video surveillance systems. This project aims to overcome these limitations byusing hybrid approaches that combine deep learning- based feature extraction with image enhancement techniques. project involves developing a system that can extract robust and discriminative features from images using deep learning techniques. The system will also apply image enhancement techniques to enhance the level of detail of the photographs, making it easier to distinguish between different individuals. The projecthas significant potential for improving the accuracy androbustness of person identification systems, which can have a positive impact on various domains such as public safety, law enforcement, and transportation.

A. Problem Statement

To create a Person Identification application with better accuracy from images using Hybrid Approaches that aims to capture the images from the video then compare the preprocessed images to the targeted image.

B. Motivation

The motivation for creating this project is to address the problem of person identification (re-ID) in footage monitoring equipment. The motivation for this project is to improve the accuracy of person identification in practical monitoring systems, which is an important task for law enforcement and public safety. Using a hybrid approach that combines image enhancement and state-of-the-art re-ID algorithms, the project aims to enhance the precision of person identification, reduce false positives, and provide more reliable results for real-world applications.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 7-Issue 1, January 2024
Paper: 74

C. *Innovative idea of our project*

1 An attempt to create a better working model using YOLO along with Real-ESRGAN in order to receive a better accuracy.
2 A comparison between the output from the model with using Real-ESRGAN and without the usage of Real-ESRGAN will be made.

II. Literature Survey

Person identification from video with image enhancement using hybrid approaches is a complex topic that involves several different fields, including computer vision, image processing, machine learning and deep learning. In computer vision, object detection and identification are critical tasks. Many studies have been proposed in order to improve the accuracy and speed of these tasks. We review several papers in this literature review that have contributed to the development of these tasks using various approaches:

The YOLO algorithm ([1], [4]) a unified model for object detection that achieves real-time performance. The YOLO algorithm divides the image into a grid and predicts bounding boxes and class probabilities for each grid cell. This approach is efficient and has been widely adopted in subsequent research. Moreover, a similar work done by Geethapriya et al [4] proposed a modified YOLO algorithm for object detection using a single neural network. Their proposed algorithm achieves real-time performance and high accuracy on multiple datasets. This algorithm is generalized, it out performs different strategies once generalizing from natural pictures to different domains. The algorithm is simple to build and can be trained directly on a complete image.

The look-based suggested a three-shot identification of individuals issue was investigated ([2]) by presenting an extensible framework to capture a person's look. Their proposed model achieves state-of-the-art performance on two benchmark datasets. To back this problem, another work by Ye and Yuen proposed PurifyNet ([6]), a resilient identification of individuals model that can handle label noise. Their proposed model achieves competitive performance with noisy labels on multiple datasets. They explain that the model optimizes the network parameters and refines the labels in a progressive manner using limited training samples for each identity. Instead of filtering out wrongly labeled samples, they re-use them. To fine-tune the network, the authors develop a hard-aware case re-weighting technique by allocating big loads to hard samples with valid labels. The efficacy of PurifyNet is proven by extensive trials on four sets of data in both simulated and real-world scenarios.

Ha et al. presented a component-based enhanced super resolution ([5]) (PESR) network for low-resolution pictures person identification. Their suggested network achieves outstanding results on two datasets that serve as benchmarks. This research primarily examines Directly Low Resolution Person Re-identification (DLRPR) problem and comes forward with a Part-based Improved Superior Resolution (PESR) network to ease the pixel-to-pixel oversight in DLRPR problems. Despite the fact that in the past certain potential techniques, such as ([1], [4]) there is still a lack of attaining precise results and hence Ahmad et al. proposed a modified YOLOv1 ([7]) neural network for object detection by modifying the loss function and adding spatial pyramid pooling layer and inception module with convolution kernels of $1 * 1$. Their proposed algorithm achieves high accuracy on multiple datasets.

And after the invention of several known technologies along with several very well reputed algorithms, a paper by Chandana and Ramachandra ([8]) provided a brief evaluation of YOLO algorithm and other prevalent variations for object detection. They discussed the advantages and limitations of each algorithm and provided insights for future research.

III. Methodology

A. *OpenCV*

OpenCV is a robust freely accessible toolkit featuring computer vision, machine learning, and photographic processing capabilities. It provides support for different programming languages, making it possible to execute its cv, ml, and image processing libraries in a wide range of applications. OpenCV is capable of analysing still photographs as well as moving videos to recognize various objects, including faces, handwriting, and artifacts. Face detection is a critical task in computer vision, and OpenCV provides a range of algorithms to achieve this objective. One of the earliest established methods for detecting facial data in OpenCV is the Haar Cascade method. This algorithm was developed in the 1980s and is still widely used today due to its effectiveness.

The Haar Cascade examine works by training a classifier using a collection of both beneficial and

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 7-Issue 1, January 2024
Paper: 74

unpleasant photos. The flattering images show faces, whilst the unflattering ones do not. The classifier uses this training data to identify the features that are most indicative of a face, such as the eyes, nose, and mouth. These features are represented as Haar-like features, which are essentially rectangular areas with different brightness levels. Once the classifier is trained, it can be used to detect faces in new images or videos. The classifier scans the input image or video using a sliding window approach, looking for areas that match the Haar-like features. If a match is found, the classifier labels that region as a potential face. The classifier then applies a set of rules to eliminate false positives and refine the face detection results. Overall, the Haar Cascade classifier is one of the fundamental and most widely used methods for detecting faces in OpenCV. It provides a reliable and effective approach for identifying facial data in various contexts, including still images, recordings, and live video streams.

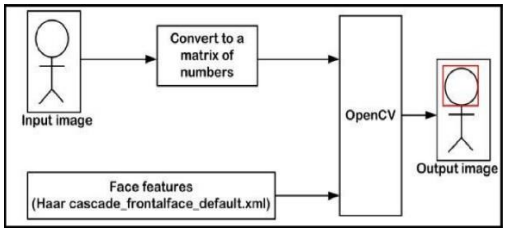

Fig. A (i) Block Diagram of Face Detection

B. Haar Cascade

A method of ML called Haar-cascading is employed to coach a classifier by exposing it to an enormous range of examples each +ve & -ve of the target category. Those that have contributed to the event of the formula are paul Viola Cascade and Michael jones. Classifiers that will support Haar options are the kind of classifiers that are utilized for detecting object. This classifier uses a way referred as ML, that has the cascade operation which uses the photos to find things in vivid images. A no-hit try at face recognizing in addition because the recognition of facial expressions in photographs. When that the experiment is finished off by showing the classifier each +ve & -ve examples of the pictures. When that the characteristics are taken from the image every feature has its own individual value which might be calculated by the formula during which it's able to recognize the faces of a range of individuals whereas they are in a very different kind of locations.

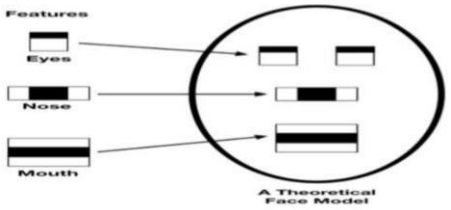

Fig. B (i). Haar-Cascade Classifier

A totally new, more efficient method of analysing photographs and locating faces through the use of rectangular characteristics is depicted in the above diagram. The rectangular features, which mimic the kernel and can be seen in the picture, are used to identify other facial features such as the eyes and the notes. This can be seen in the figure.

Integral pictures make it possible to calculate in constant time the Haar-like feature of any size.

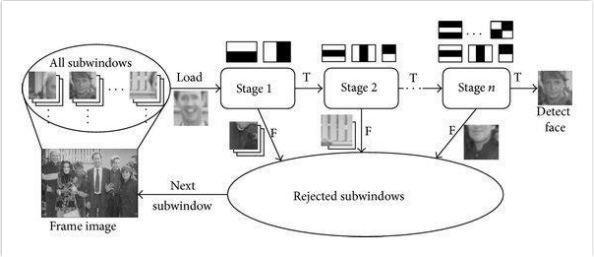

Fig. B (ii). Stages involved in Face Detection with HAAR Cascade in OpenCV Python

Steps involved in HAAR Cascade Method:

- The image is fragmented into smaller parts before being sent to the classifier in order to improve accuracy (or sub windows as shown in the illustration)
- We organise N detectors in a cascade arrangement, with each one learning a combination of multiple different feature types from images that are sent through the cascade (such as lines, edges, circles, and squares). After the feature extraction process is finished, a confidence rating is supposedly assigned to each individual sub-part.
- Only the images (or sub-images) in which there is the highest degree of certainty that they represent face images are accepted and sent to the accumulator. As a direct consequence of this, the

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 7-Issue 1, January 2024
Paper: 74

cascade obtains the subsequent frame or image, if there is one, and then restarts the process.

C. *Real ESRGAN*

Real-ESRGAN is a state-of-the-art image super-resolution algorithm that uses a deep convolutional neural network (CNN) to enhance the resolution and quality of images. It is an extension of ESRGAN and is designed to produce more realistic and natural-looking images. Real-ESRGAN uses a novel architecture that includes a feature extraction network and a reconstruction network. The feature extraction network is designed to extract high-level features from low-resolution images, while the reconstruction network uses the extracted features to generate high-resolution images.

Fig. C (i). Comparison of several image enhancement techniques.

Real-ESRGAN also incorporates several advanced techniques such as progressive up-sampling, residual-in-residual dense blocks, and adversarial loss to further improve the quality of the generated images. It has been shown to outperform existing state-of-the-art super-resolution algorithms on standard benchmark datasets and is widely used in various applications such as medical imaging, remote sensing, and computer vision.

D. *YOLO*

YOLO v3 is an item classification solution that employs just one neural network to do object recognition and categorization in real-time. It was created by Joseph Redmon and his colleagues at the University of Washington and was debuted in 2018. Compared to its predecessors, YOLOv3 offers superior accuracy and speed thanks to a few major improvements in the design. These include the usage of leftover blocks in feature pyramid networks, and a novel categorization loss function named cross-entropy binary with logistic loss. The topology of the network is formed around a Darknet-53 core that is utilised to extract attributes from the input picture. These characteristics are then provided into a detection head that predicts box bounds, objectness scores, and class probabilities for each object in the image. Overall, YOLOv3 is a sophisticated object identification system that has earned tremendous popularity in both academic and commercial contexts thanks to its real-time performance and high accuracy.

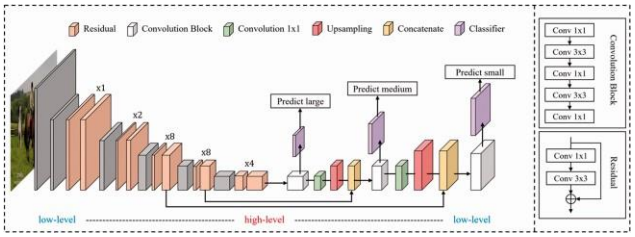

Fig. D (i). YOLOv3 Architecture Diagram

IV. DATASET

We utilized the Queen Mary Multiview Dataset which is made up of 48 face pictures that have been aligned, cropped and normalized by a computer programme. Out of these 38 are in grayscale and have an image size of 100 by 100 pixels, while 10 are in RGB color and have a size of 56 by 56 pixels. In addition to this we have included a bespoke dataset that is made up of 3500 grayscale face photos that have been taken in 350 various positions. In the beginning, the pictures are taken using a high-definition webcam that has a resolution of 1080 pixels.

V. EXPERIMENTAL RESULTS

A. *Experimental Setup:*

All architectures were coded using the YOLOv3 [1] framework and were trained on the NVIDIA RTX 3050. The training dataset comprised of 685 data points.

Figure 1, illustrates the experimentation specifics of each of the deep learning models known with ours. It is likely that Figure 1 include information such as the accuracy rate together with their respective model. These facts are critical for understanding how the models were built and trained and give a basis for comparing the performance of the different models.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 7-Issue 1, January 2024
Paper: 74

Overall, the experimental setup demonstrates the necessity of choosing proper algorithms and software when training deep learning models, as well as the requirement for a well-defined experimental design and unambiguous reporting of results.

B. Discussions:

Real-ESRGAN is a visual enhancement technological advances that employs deep learning algorithms to increase the quality and quality of pictures. Compared to previous image optimizing approaches, Real-ESRGAN has various benefits. Real-ESRGAN generates high-quality photos with improvedresolution, texture, and detail while retaining the naturalness of the original image. Real-ESRGAN is a flexible approach that may be utilized for numerous image enhancing applications, including super-resolution, denoising, and deblurring. This implies thatit may be applied to a wide range of photos and can deliver high-quality results across multiple domains.

C. Results:

We compare our final models on several public standard datasets with state-of-the-art techniques including Bicubic, EnhanceNet, HaarCascade and ESRGAN. With the goal to evaluate the effectiveness of the techniques, we have compared the persisting algorithms mainly along with their accuracy rates. Each and every algorithm when evaluated with their respective dataset consisting of unique images, hence produce results that show how efficient the algorithm is.

As is evident from Table 2, we observe spontaneous results of improvement. Starting from the bicubic implementation, to the modern ESRGAN implementation. Our analysis showed that YOLOv3 framework when combined with Real-ESRGAN performed the best among all the remaining algorithms. The accuracy rate had a good amount of 90% - 95% for the same. These results suggest that YOLOv3 with Real-ERGAN is the best algorithm for audio and visual data tasks among the three evaluated.

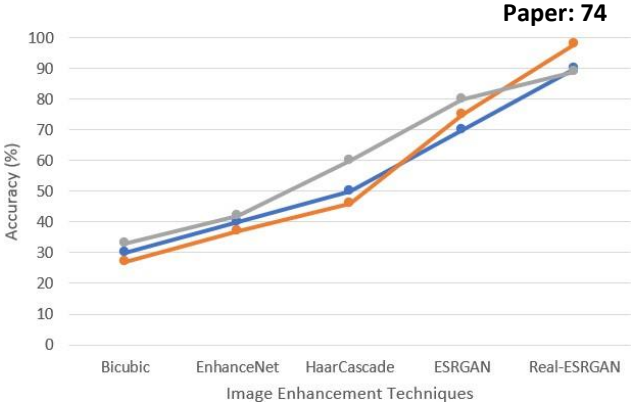

Fig. 1 (i). Comparison of different image enhancement algorithms with Real-ESRGAN.

As previously stated, they are numerous more picture improvement strategies. According to the study, we have acquired more effective accuracies when the approach underwent training with Real-ESRGAN than under a regular ESRGAN. The following are the results received by the model.

Model	Accuracy
YOLOv3 without Real - ESRGAN	80% - 85%
YOLOv3 with Real - ESRGAN	90% - 95%

Table 1 (i) accuracy based on respective algorithm

VI. CONCLUSION

The research study evaluates the performance and efficacy of numerous computer vision algorithms, including Haar Cascade, YOLOv3, LabelImg, and Real-ESRGAN. The report discusses the benefits and disadvantages of each approach and gives insights into their prospective use cases. YOLOv3 is a real-time object identification method that has demonstrated excellent performance in recognizing objects in diverse environments. The article indicates that YOLOv3 can achieve high accuracy while retaining quick processing speeds, making it an attractive candidate for real-time applications. The research discusses Real-ESRGAN, an image restoration technique that leverages a deep learning approach. The article indicates that Real-ESRGAN is successful in increasing the quality of low-resolution photos, making it a valuable tool for image improvement applications. Finally the research presents a comparison between the accuracy supplied by utilizing Real-ESRGAN for picture identification and the accuracy offered without using Real-ESRGAN.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 7-Issue 1, January 2024
Paper: 74

VII. REFERENCES

[1] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

[2] Khan, F. M., & Brémond, F. (2016). Person identification for real-world surveillance systems. *arXiv preprint arXiv:1607.05975*.

[3] Van Ranst, W., De Smedt, F., Berte, J., & Goedemé, T. (2018, November). Fast simultaneous people detection and identification in a single shot network. In 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) (pp. 1-6). IEEE.

[4] Geethapriya, S., Duraimurugan, N., & Chokkalingam, S. P. (2019). Real-time object detection with Yolo. International Journal of Engineering and Advanced Technology (IJEAT), 8(3S).

[5] Ha, Y., Tian, J., Miao, Q., Yang, Q., Guo, J., & Jiang, R. (2020). Part-based enhanced super resolution network for low-resolution person identification. IEEE Access, 8, 57594-57605.

[6] Ye, M., & Yuen, P. C. (2020). PurifyNet: A robust person identification model with noisy labels. IEEE Transactions on Information Forensics and Security, 15, 2655-2666.

[7] Ahmad, T., Ma, Y., Yahya, M., Ahmad, B., Nazir, S., & Haq, A. U. (2020). Object detection through modified YOLO neural network. Scientific Programming, 2020, 1-10.

[8] Chandana, R. K., & Ramachandra, A. C. (2022). Real Time Object Detection System with YOLO and CNN Models: A Review. arXiv preprint arXiv:2208.00773.

[9] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes challenge: A retrospective. International Journal of Computer Vision, 111(1):98–136, Jan. 2015. 2

[10] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(9):1627–1645, 2010. 1, 4

[11] S. Gidaris and N. Komodakis. Object detection via a multiregion & semantic segmentation-aware CNN model. CoRR, abs/1505.01749, 2015. 7

[12] Zhang, L., Yang, M., Feng, X.: Sparse representation or collaborative representation: which helps face recognition? In: ICCV. (2011)

[13] Bialkowski, A., Denman, S., Sridharan, S., Fookes, C., P.Lucey: A database for person identification in multi-camera surveillance networks. In: DICTA. (2012)

[14] Hirzer, M., Beleznai, C., Roth, P.M., Bischof, H.: Person identification by descriptive and discriminative classification. In: Image Analysis. Springer (2011) 91–102

[15] B. Kumar, G. Carneiro, I. Reid, et al. Learning local image descriptors with deep siamese and triplet convolutional networks by minimising global loss functions. In CVPR, pages 5385–5394, 2016. 2

[16] V. B. Kumar, B. Harwood, G. Carneiro, I. Reid, and T. Drummond. Smart mining for deep metric learning. arXiv preprint arXiv:1704.01285, 2017. 2

[17] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and ´ S. Belongie. Feature pyramid networks for object detection. In CVPR, volume 1, page 4, 2017. 2

[18] Joseph Redmon, Ali Farhadi,"YOLO9000: Better, Faster, Stronger", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 7263-7271.

[19] Jifeng Dai, Yi Li, Kaiming He, Jian Sun, "R-FCN: Object Detection via Region-based Fully Convolutional Networks", published in: Advances in Neural Information Processing Systems 29 (NIPS 2016).

[20] Karen Simonyan, Andrew Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", published in Computer Vision and Pattern Recognition (cs.CV)