# Enhancing video face recognition through Illumination and pose compensation models

*Gajanan Joshi*[1*]

[1]Department of computer science VPASC College Baramati, India

**Abstract.** This work envisions a strong framework for improving video-based face recognition by addressing simultaneously the issues of pose and illumination changes. Since even exhaustive efforts in face recognition research have not eliminated the impact of dynamic environments in real-world applications on recognition accuracy caused by nonlinear facial distortions and lighting differences, this work is timely. To circumvent these challenges, a hybrid deep learning model is constructed that combines SENet and channel attention mechanisms for efficient spatial feature extraction and a Transformer network and cross-attention for temporal dependency modeling. The new system begins with face detection via SENResNet, feature refinement via Transformers, and stable tracking via the Regression Network-based Face Tracking (RNFT) model. This end-to-end system enables effective learning of invariant representations over diverse poses and lighting conditions. Testing on benchmark datasets shows substantial improvement in recognition accuracy and robustness, confirming the effectiveness of the proposed system in real-world applications of video surveillance and human-computer interaction.

Keywords—Face Recognition, Video Surveillance, Pose Variation, Illumination Compensation, SENRes Net, RNFT (Regression Network-based Face Tracking), Deep Learning, Transformer Network, Cross-Attention Mechanism, Temporal Dependency.

## 1. Introduction

Face recognition has been a computer vision focus for 30 years. Pose and lighting fluctuations make practical applications difficult despite significant advances. Some systems handle stance variations [1] or illumination changes [2][3], but few handle both [4]. Face recognition reliability is greatly affected by such variances, especially in dynamic video applications. Models that handle position and light fluctuations will improve video face recognition in realistic and uncontrolled conditions. Pose variation complicates face recognition because to the complicated 3D human head. Some features are partly or totally covered by rigid head rotations, whereas the viewable sections alter form and look

* Corresponding author: research.joshigajananr@gmail.com

nonlinearly with posture. Traditional algorithms struggle to recognize faces in varied positions because these variations typically outweigh identification distinctions. This study uses contemporary pose compensation models to manage appearance changes from posed photos, improving video-based face recognition system dependability.

Images are significantly affected by illumination because the intensity of each pixel $I(x, y)$is determined by the product of incident lighting $L(x, y)$ and reflectance $R(x, y)$, meaning that $I(x, y) = R(x, y) \times L(x, y)$. Consequently, acquired photos are affected by lighting conditions. Modeling the light distribution or extracting illumination-invariant features like Local Binary Pattern (LBP) and Histogram of Oriented Gradients (HOG) [5] are the two primary methods utilized to accomplish illumination-invariant face recognition [6], [7].

Face recognition systems must be able to manage both position and lighting changes at the same time since they often happen concurrently in real-world situations. Deep learning techniques have shown a great capacity to model nonlinear data distributions, attaining state-of-the-art performance in applications such as object identification [8] and object classification [9]. They excel in learning hierarchical features, where each network layer recognizes patterns ranging from global structures to local gradients. Traditional hand-crafted characteristics are less strong and informative than these learnt attributes [10], [11].

Even if the face seems different in different postures, the same person may have important linkages. Photo of the same person in various lighting matches. Proper instruction may teach posture and lighting invariance. Deep convolutional networks' feature learning capabilities enable this study's location and lighting-based face detection method.

This study improves video face recognition with location and illumination changes. Use SENet for channel attention and a Transformer with cross-attention for temporal dependency modeling to gather spatial and temporal data. Hybrid designs compensate for self-occlusions and nonlinear feature fusion in informative zones. End-to-end learning enhances identification accuracy on benchmark datasets, making it a reliable and scalable real-world video face recognition technology.

## 2. Literature review

Recently, many promising approaches have been proposed to tackle the poses and illumination challenge in face recognition, leveraging advancements in deep learning and computational techniques. Here some studies are given below.

| Reference | Title | Method | Result |
|---|---|---|---|
| [12] | Application Research on Face Image Evaluation Algorithm of Deep Learning Mobile Terminal for Student Check-In Management | face image detection model for student check-in management is constructed, and a deep learning network is used to realize face detection. | showed that the face detection based on the improved deep learning network can shorten the retrieval time and improve the accuracy of face image classification. |
| [13] | Face Recognition and Identification Using Deep Learning | Convolutional neural networks are capable of extracting diverse characteristics from photos. An uncomplicated approach for face recognition is used. | The findings underscore the efficacy and accuracy of the method in differentiating persons based on face features, making a notable addition to the domain of biometric identification. |

| | | | |
|---|---|---|---|
| [14] | Multi-view Face Detection Using Deep Convolutional Neural Networks | deep learning object detection methods such as bounding-box regression, SVM, or image segmentation | suggested technique can recognize faces from various angles and can partially manage occlusion; nonetheless, a link seems to exist between the distribution of positive instances in the training set and the scores. about the planned facial recognition system. |
| [15] | Illumination Invariant Face Recognition Using Near-Infrared Images | Local binary pattern characteristics counteract monotonic transformation, producing illumination-invariant facial representation. Techniques for face identification using NIR pictures, statistical learning algorithms extract discriminative features, and develop precise face matching engine. | a system for precise, rapid face identification in practical applications, addressing specular reflections of active NIR lights on eyeglasses, a significant challenge in active NIR image-based face recognition. |
| [16] | Classical and modern face recognition approaches | variability in illumination, pose, aging, cosmetics, scale, occlusion, and background. Along with classical face recognition techniques, most recent research directions are deeply investigated, i.e., deep learning, sparse models and fuzzy set theory. | presents future aspects of face recognition technologies and its potential significance in the upcoming digital society. |
| [17] | Real-time Face Recognition and Emotion Detection Using Deep Learning | Deep learning methods with Keras convolutional neural network for real-time identification and emotion detection | The system achieved high accuracy in detecting facial characteristics and emotions, validated through experiments using OpenCV and Python. |
| [18] | Local feature hierarchy for face recognition across pose and illumination | Proposed a convolutional network method to extract nonlinear features invariant to pose and illumination using 1 × 1 convolutions. | Achieved 96.9% recognition accuracy on multi-PIE dataset, improving the state-of-the-art by 7.5%. For profile views, accuracy improved by 19%. |
| [19] | Face Recognition from Video using Deep Learning | Face recognition using pre-trained FaceNet (FN) model to identify offenders in video | Demonstrated efficient offender identification with high accuracy, though requiring substantial data. The system is beneficial for law enforcement and administrative applications. |

| | | | |
|---|---|---|---|
| [20] | An Effective Face Recognition under Illumination and Pose Variations | Two-dimensional view-based face recognition and shadow compensation method for simultaneous pose and illumination handling | Proposed method effectively compensates for pose and illumination variations with less computational effort compared to 3D models, providing a comparable recognition rate. |
| [21] | Combine Image Quality Fusion and Illumination Compensation for Video-Based Face Recognition | Classification of illumination orientation, illumination direction mapping, plane-fitting technique, and picture quality fusion criteria. | Achieved significant performance improvement in face recognition under varying illumination conditions. The method reduced the impact of illumination and image quality degradation, with satisfactory recognition rates on various face video databases. |
| [22] | Efficient face detection and tracking in video sequences based on deep learning | SENResNet face recognition and RNFT face tracking models | Videos for face identification and tracking were improved by the SENResNet and RNFT models over state-of-the-art approaches. |

## 3. Research gap

Despite improvements in face recognition under different stances and lighting, video-based methods are limited. Especially in 3D models, image quality fusion, CNNs, FaceNet, and SENResNet handle just one problem or need large computing resources. Temporal correlations and self-occlusion are infrequently studied, and end-to-end spatial-temporal frameworks are inefficient and unscalable. A hybrid model that accommodates position and lighting fluctuations while using channel attention and temporal dependency modeling is required for real-world video face recognition.

## 4. Background

*SENResNet face detection model*

Face identification and tracking technology is used in video surveillance, safe driving, human–computer interaction, and medical diagnostics. Ambient light, posture, and occlusion interfere with most face identification and tracking systems in video sequences. We present an efficient deep learning-based face recognition and tracking framework for video sequences to achieve accurate face tracking.

*System architecture*

The proposed efficient architecture for face identification and tracking consists of a SENResNet model for face detection and an RNFT model for face tracking. The overall process of the framework is demonstrated in Fig. 1. The process of implementation is described below:

**Initialization:** SENResNet is used by the framework to automatically identify the target face in the first video frame; no human annotation is necessary. Relevant face traits and templates are computed, and the tracking window is established by the identified facial part.

**Face Detection:** SENResNet, which combines SEN and ResNet, gives the target face position for tracking initialization after detecting facial characteristics in every frame. By

anticipating probable face locations, a decision process manages transient occlusions and automatically resumes tracking.

**Face Tracking:**

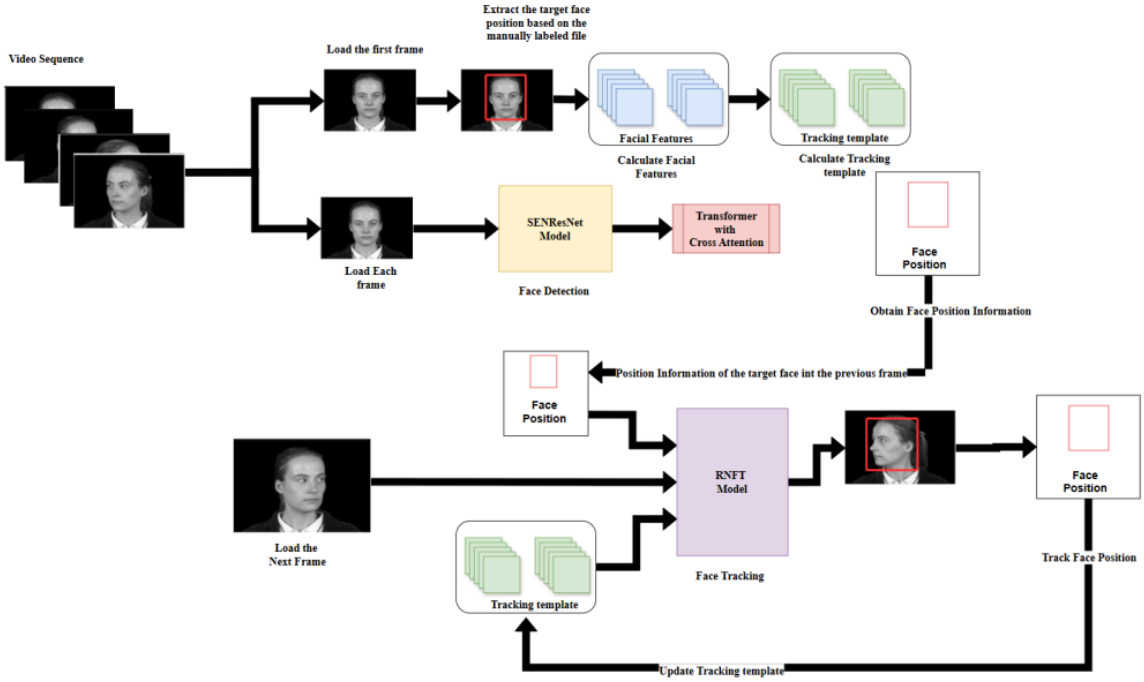


Fig. 1.An image of a flowchart for a face detection and tracking system, showing the different steps involved in the process

The RNFT model uses the tracking template and SENResNet inputs to track the face. Through frame-by-frame prediction of candidate face positions, it guarantees precise and consistent tracking throughout the video sequence.

Illustration depicts facial tracking and identification. After loading the first video frame, SENet detects faces. Cross-attention Transformers enhance identified properties. The optimum face position updates the RNFT model's tracking template for the following frame. This method monitors face position throughout video. Face recognition and feature extraction are used by most components to gather video sequence information. A good example shows face detection and tracking.

*Face detection based on SENResNet*

CNNs' accuracy and multi-level feature learning make them good face detectors, however deeper networks have gradient vanishing difficulties. ResNet by He et al. efficiently trains deep networks by preserving gradients during backpropagation. In the proposed SENResNet, SEN and ResNet enhance significant traits and decrease unimportant ones. Figure 2 shows that SENResNet has three convolutional layers, one max-pooling layer, 16 ResNet blocks, and 3 SEN blocks. Resized images are processed using 3x3 convolutions, double channel pooling layers, and a global average layer for prediction. Table 1 shows SENResNet setup.

*ResNet Module*

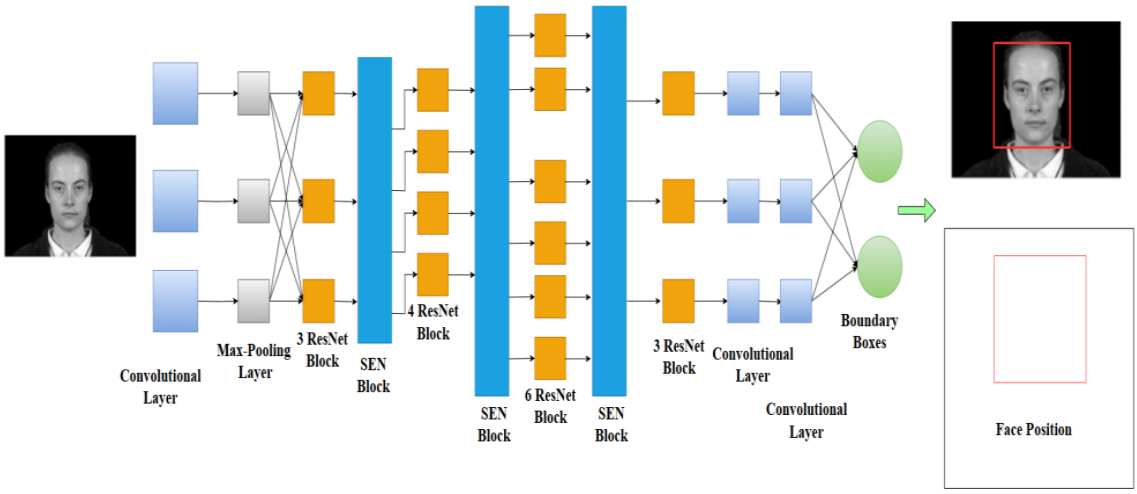


Fig. 2. architecture of the SENResNet model for facial detection

In the SENResNet model, for an input $x$, the feature extracted by the model is $(x)$. The residual mapping is characterized by the acquisition of an additional mapping using stacked non-linear layers, where:

$$F(x) = H(x) - x \qquad (1)$$

The learned residual mapping becomes $(x) + x$ instead of a straight mapping link. A ResNet module is suggested to learn a residual mapping. Figure 2 depicts a ResNet module setup.

ResNet bypasses convolutional layers via short-circuit connections. The function H(x) = F(x) + x is accomplished using short-circuit connections. The short-circuit link becomes an identity mapping, enabling input data to be directly added to the stacked convolutional layers' output. Definition of residual block:

$$u = F(x, W) + x \qquad (2)$$

where $x$ and $y$ denotes the input and output of the residual block respectively while $F(x,W)$ depicts the residual mapping the block learns to.

In case each residual block has three layers, the residual mapping defines as:

$$F(x) = W_3.F(W_2x) \qquad (3)$$

F(x) + x is the shortcut connection notation, and element-wise addition yields the network feature map. A linear projection $s$ is used in the shortcut connection if $x$ and the feature map after operation $F$ have mismatched dimensions:

$$y = F(x, W) + W_s x \qquad (4)$$

This ensures the dimensions of x and the feature map align.

*SEN Module*

We build the SENResNet model by adding SEN to ResNet, which combines information across convolution channels. The SEN module clearly models convolutional properties' exchange interactions and recalibrates them to emphasize important features and reduce unnecessary ones. distinct neural network layers perform distinct functions when the SEN module is set up.

The SEN module is in three components:

**Convolutional Layer:** This part carries out the convolution operation

$F\ t\ : X \rightarrow U$ F tr:X→U, where $X \in R\ H\ 0\ W\ 0\ C\ 0$ X∈R H 0W 0C 0

The input is denoted as U where U ∈ R^(HWC), and the output is specified accordingly. Convolution operation is denoted as

$$u_c = v_c * X = \sum_{S=1}^{C} vS_c . x_s \qquad (5)$$

**Squeeze Operation:** This operation reduces multiple spatial dimensions of feature maps into a channel descriptor to summarize the global context. Squeeze operation yields a channel descriptor

$\boldsymbol{z \in R\ C}$ **z∈R C** that can be obtained as

$$Z_c = F_{sq}(u_c) = \frac{1}{HW} \sum_{i=1}^{H} . \sum_{j=1}^{W} u_c(i,j) \qquad (6)$$

**Excitation Operation:** The excitation function delineates the relationship between channels and employs a gating mechanism to readjust the feature maps. The excitation function is characterized as:

$$s = F_{ex}(z, W) = \sigma\big(W_2 . ReLU(W_1 . z)\big) \qquad (7)$$

where $W_1 \in \mathbb{R}^N$

$W_1 \in \mathbb{R}C_rC$, $W_2 \in \mathbb{R}C_rC$

W ∈ R^C_r^C, where σ denotes the sigmoid activation function.

The ultimate result after excitation is the recalibrated feature map $ec$.

$$e_c = s_c . u_c \qquad (8)$$

where $sc$ is the scalar output of the excitation function, and $uc$ is the original feature map.

*Loss Functions for Face Detection*

The effectiveness of face identification in the SENResNet model is affected by several loss functions. During training, the model uses several loss functions, such as SoftMax, Large-Margin SoftMax (L-SoftMax), Normalized SoftMax Loss (NSL), Angular SoftMax (A-SoftMax), and Additive Margin SoftMax (AM-SoftMax), to improve accuracy in complex face identification scenarios.

**SoftMax Loss:** This often-used loss function for classification problems optimizes the posterior probability of the accurate label. It is characterized as:

$$L_{Softmax} = -\frac{1}{m} \sum_{i=1}^{m} \log\left(\frac{e^{W^T{}_{yi}\, xi + b_{yi}}}{\sum_{j=1}^{n} e^{W^T{}_j\, x_i + b_j}}\right) \qquad (9)$$

**Large-Margin SoftMax Loss (L-SoftMax):** This function transforms the SoftMax loss into an angular representation, hence improving class separation. The L-SoftMax is characterized as follows:

$$L_{L-softmax} = -\frac{1}{m} \sum_{i=1}^{m} \log\left(\frac{e^{k\cos(h_{y_i} x_i)}}{e^{k\cos(h_{y_i} x_i)} + \sum_{j \neq y_i} e^{k\cos(h_j x_i)}}\right) \qquad (10)$$

**Normalized SoftMax Loss (NSL):** NSL standardizes the weight values to $kWj$=1 and mitigates the impact of feature vector norms. It is characterized as:

$$L_{NSL} = -\frac{1}{N}\sum_{i=1}^{m} \log\left(\frac{e^{\cos\left(h_{y_i}x_i\right)}}{\sum_{j=1}^{n} e^{\cos\left(h_j x_i\right)}}€\right) \quad (11)$$

**Angular SoftMax Loss (A-SoftMax):** This function incorporates an angular margin to enhance feature distinguishability. It is characterized as:

$$L_{A-softmax} = -\frac{1}{m}\sum_{i=1}^{m} \log\left(\frac{e^{k\cos\left(mh_{y_i}x_i\right)}}{e^{k\cos\left(mh_{y_i}x_i\right)}+\sum_{j\neq y_i} e^{k\cos\left(h_j x_i\right)}}\right) \quad (12)$$

**Additive Margin SoftMax (AM-SoftMax):** This loss function modifies the angular margin to enhance convergence. It is characterized as:

$$L_{AM-softmax} = -\frac{1}{m}\sum_{i=1}^{m} \log\left(\frac{e^{\cos\left(h_{y_i}x_i\right)-m}}{e^{\cos\left(h_{y_i}x_i\right)-m}+\sum_{j\neq y_i} e^{\cos\left(h_j x_i\right)}}\right) \quad (13)$$

*RNFT face tracking model*

The RNFT model improves SENResNet by learning facial characteristics from successive frames to predict face positions. It tracks robustly under rotation, translation, and illumination changes using initialization and synchronized current-frame window inputs. In difficult video settings, a correction network retains face proportions from the previous frame, enhancing accuracy, stability, and tracking.

## 5. Methodology

*Data Preprocessing*

In the RNFT model, facial tracking begins with an automatically identified bounding box around the face specified by center coordinates ((c_x, c_y)), width (w), and height (h), increased by a factor (k_1 > 1) to gather contextual data. The model trims the current frame at each frame (t) using the face position from frame (t-1) to reduce computation and background noise. Since frames seldom change, reliable tracking is guaranteed if the face doesn't move or disappear. For fast motions, increasing (k_1) yields more robustness but increases complexity. Model produces the monitored face's updated location and dimensions ((c_x', c_y', w', h', k')) in the current frame.

*Model training*

A hybrid approach for tracking, feature improvement, and face recognition in real-time inside video sequences is presented in this study. It incorporates RNFT for precise frame-to-frame tracking, transformers with cross-attention for enhanced feature representation, and SENResNet for initial face identification. With a thorough step-by-step breakdown of the procedure, the system tackles issues including motion blur, occlusion, and changes in face position, lighting, and expression.

*Input Processing:* The system receives a video series $V = \{ F 1 , F 2 , \cdots , F N \}$ V={F 1,F 2,⋯,F N} containing frames $F t \in R H \times W \times C$ F t∈R H×W×C with dimensions $H \times W$ H×W and color channels C. A lightweight pre-processing module recognizes humans in each frame for efficiency. Without persons, frames are eliminated; those containing identified individuals continue. This allocates computational resources to

relevant frames, resulting in a subset $F\ t\ '\ \subset V$ F t ′ ⊂V containing at least one identifiable individual each frame.

*Face Detection Using SENResNet:* The SENResNet model recognizes and locates faces in frame $Ft$. ResNet's powerful feature extraction is used with Squeeze-and-Excitation (SE) blocks to emphasize important features and reduce less important ones in SENResNet. The network can focus better on facial characteristics like the eyes, nose, and mouth.

Extraction of Features

The input frame $Ft$ is processed via a ResNet backbone, yielding a feature map $x$:

*Squeeze-and-Excitation Module:* The SE block readjusts the feature map $x$ using two processes: global average pooling and two completely linked layers with non-linear activations.

Where:

*GAP(x)*

Global average pooling (GAP(x)) condenses spatial data into channel descriptors.

$W_1$ and $W_2$ are SE block trainable weight matrices. δ is the ReLU activation function, whereas σ is the sigmoid function. Recalibrated features $z$ improve the original feature map by highlighting relevant channels. Input modified attributes into the detecting head to predict a bounding box (Bt=(x1,y1,x2,y2)) around the detected face.

*Refinement of Facial Features Using Transformers:* Convolutional approaches may prevent SENResNet from detecting certain facial features, such as wrinkles and emotions. Facial regions $B\ t$ are taken from the frame and treated using a transformer to improve features.

*Refinement of Transformer Features:*

Self-attention helps transformers model visual connections. Feature vectors are encoded from the clipped face area $Bt$ for the transformer. The transformer enhances feature representation by analyzing these characteristics:

$Ft$ polished.

Local and global facial dependencies are captured in the polished facial feature map (F t refined). Refining ensures nuanced and context-sensitive features are accurately represented.

*Enhancement Using Cross-Attention Mechanism:*

To improve features, the transformer has a cross-attention mechanism. To increase accuracy, cross-attention uses temporal data to match current frame features with nearby frames.

Inter-Attention

Cross attention operation compares query (current frame characteristics) to keys and values (adjacent frames). Formula for it:

Q indicates the query matrix based on the refined characteristics of the current frame Ft. K and V are key and value matrices, derived from refined properties of consecutive frames Ft−1 and Ft+1. Key vector dimension is dk.

The cross-attention module improves feature maps:

Using contextual information from nearby frames, the model may correct missing or distorted features.

*Person Tracking Using RNFT:*

Upgraded feature maps The RNFT module monitors the person via frames, receiving updates. The RNFT is designed to address short-term and long-term dependency. This makes RNFT sturdy against quick motion and occlusion.

An RNN component

Transient dependencies between frames are captured by the RNN: $h\ t$ h t represents the hidden state at time $t$ t, capturing the temporal context up to frame $t$ t.

*Transformer Element:* The transformer component analyses the succession of hidden states $\{h_1, h_2, \ldots, h_t\}$ to capture long-term dependencies.

*Output:* The RNFT module outputs the predicted bounding box or position of the person in the next frame: This step ensures robust and consistent tracking, irrespective of variations in the person's position or motion.

*Final Output: Continuous Tracking:* The system's final output consists of a series of bounding boxes and enhanced feature maps that continuously monitor the identified individual throughout the video frames. This output is robust against changes in position, occlusions, and other difficulties and ensures reliable performance in real-world applications.

*Evaluation Metrics*

The proposed methodology is evaluated using both detection and tracking metrics. For face detection, precision and recall are computed as:

*Accuracy:* Accuracy is the most straightforward method for gauging how often the classifier produces accurate predictions. It may be seen as the ratio of all accurately anticipated favorable events to the entire number of predictions.

$$Accuracy = \frac{TP + TN}{S} \qquad (1)$$

*Precision:* Contrary to this ratio, which also subtracts one from it (1—precision) and displays the proportion of false negatives, 1/Precision produces recall.

$$Precision = \frac{TP}{TP + FP} \qquad (2)$$

*Recall:* However, in contrast to true negatives, they are also known as false negatives.

$$Recall = \frac{TP}{TP + FN} \qquad (3)$$

*F1-Score:* F1-Score is computed using the recall and precision ratings' symmetrical mean.

$$F_1 = \frac{2 * Pr\,e\,cision * Re\,c\,all}{Pr\,e\,cision + Re\,c\,all} \qquad (4)$$

# 6. CONCLUSION

This work effectively implemented discrete quadrotor control algorithms on onboard hardware. Static output feedback is needed for control design since ordinary sensors cannot accurately monitor state variables like propeller rotation speeds. To improve robustness and reduce disturbances, the closed-loop system's $L_2$-gain was reduced. Unlike continuous-

system approaches, parametric synthesis was applied to discrete-time systems, closing a gap. The method stabilizes UAV tracking under turbulence, making it useful for testing and stabilization. For better performance, enhanced navigation filtering, loop-shaping, and mixed sensitivity will be studied.

## References

1. X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. image Process.*, vol. 19, no. 6, pp. 1635–1650, 2010.
2. C. Ding, J. Choi, D. Tao, and L. S. Davis, "Multi-directional multi-level dual-cross patterns for robust face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 3, pp. 518–531, 2015.
3. M. Günther *et al.*, "The 2013 face recognition evaluation in mobile environment," in *2013 International Conference on Biometrics (ICB)*, IEEE, 2013, pp. 1–7.
4. W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, 2003.
5. X. Wang, T. X. Han, and S. Yan, "An HOG-LBP human detector with partial occlusion handling," in *2009 IEEE 12th international conference on computer vision*, IEEE, 2009, pp. 32–39.
6. K.-C. Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 684–698, 2005.
7. X. Jiang, Y. O. Kong, J. Huang, R. Zhao, and Y. Zhang, "Learning from real images to model lighting variations for face images," in *Computer Vision–ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part IV 10*, Springer, 2008, pp. 284–297.
8. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2016.
9. C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.
10. Z. Zhu, P. Luo, X. Wang, and X. Tang, "Multi-view perceptron: a deep model for learning face identity and view representations," *Adv. Neural Inf. Process. Syst.*, vol. 27, 2014.
11. M. Kan, S. Shan, and X. Chen, "Multi-view deep network for cross-view classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4847–4855.
12. Y. Tan, "Application Research on Face Image Evaluation Algorithm of Deep Learning Mobile Terminal for Student Check-In Management," *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/3961910.
13. V. C. R, V. Asha, B. Saju, N. S., T. R. M. Reddy, and S. K. M, "Face Recognition and Identification Using Deep Learning," in *2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*, 2023, pp. 1–5. doi: 10.1109/ICAECT57570.2023.10118154.
14. S. S. Farfade, M. Saberian, and L. J. Li, "Multi-view face detection using Deep convolutional neural networks," *ICMR 2015 - Proc. 2015 ACM Int. Conf. Multimed. Retr.*, pp. 643–650, 2015, doi: 10.1145/2671188.2749408.
15. S. Z. Li, R. Chu, S. Liao, and L. Zhang, "Illumination Invariant Face Recognition Using Near-Infrared Images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 627–639, 2007, doi: 10.1109/TPAMI.2007.1014.

16. W. Ali, W. Tian, S. U. Din, D. Iradukunda, and A. A. Khan, *Classical and modern face recognition approaches: a complete review*, vol. 80, no. 3. Multimedia Tools and Applications, 2021. doi: 10.1007/s11042-020-09850-1.
17. S. A. Hussain and A. S. A. Al Balushi, "A real time face emotion classification and recognition using deep learning model," in *Journal of physics: Conference series*, IOP Publishing, 2020, p. 12087.
18. X. Jiang, D. Zhang, and X. Feng, "Local feature hierarchy for face recognition across pose and illumination," in *2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, IEEE, 2016, pp. 1–5.
19. S. Manna, S. Ghildiyal, and K. Bhimani, "Face recognition from video using deep learning," in *2020 5th International conference on communication and electronics systems (ICCES)*, IEEE, 2020, pp. 1101–1106.
20. S.-I. Choi and C.-H. Choi, "An effective face recognition under illumination and pose variations," in *2007 International Joint Conference on Neural Networks*, IEEE, 2007, pp. 914–919.
21. C. Wang and Y. Li, "Combine image quality fusion and illumination compensation for video-based face recognition," *Neurocomputing*, vol. 73, no. 7–9, pp. 1478–1490, 2010.
22. G. Zheng and Y. Xu, "Efficient face detection and tracking in video sequences based on deep learning," *Inf. Sci. (Ny).*, vol. 568, pp. 265–285, 2021.