

Enhancing automotive interior automation through face analysis techniques

Hanzhen Qin

Department of Computer Science, Rensselaer Polytechnic Institute, 110 8th St, Troy, NY, 12180

Contributing author: lilryan988@gmail.com; qinh2@rpi.edu

Abstract. In contemporary society, the progress and widespread adoption of automotive automation, particularly in autonomous driving technology, have been remarkable. The rapid evolution of this technology has equipped numerous vehicles with high-stability autonomous capabilities, significantly enhancing convenience for users. However, as autonomous driving continues its developmental trajectory, it is crucial to give equal attention to the advancement of interior automation technologies within vehicles. Exploring avenues that make these automated technologies smarter, safer, and more conducive to delivering heightened convenience and support to users is imperative. This realm presents a domain ripe with potential for innovation and substantial advancements that cater to the evolving needs and expectations of modern transportation.

Keywords: Convolutional Neural Network (CNN), Principal Component Analysis (PCA), FaceNet, Autonomous Driving Technology

1. Introduction

In contemporary society, autonomous driving technology stands out as a breakthrough innovation with far-reaching significance in the transportation field. This transformative technology holds the promise of delivering advantages beyond efficiency, sustainability, and safety[1]. Firstly, the autonomous driving system has the potential to significantly reduce traffic accidents by efficiently navigating complex traffic scenarios, mitigating human errors such as ignoring traffic signals, maintaining inadequate distance between vehicles, speeding, distracted driving, drunk driving, and fatigue driving [2]. Additionally, it extends increased mobility to individuals with physical disabilities and the elderly. Moreover, autonomous driving technology, utilizing state-of-the-art machine learning algorithms, can make optimal choices on each road section, achieving the lowest fuel consumption and energy-saving effects through calculated smooth driving speeds.

In critical situations, particularly those involving fatigue driving, autonomous driving can entirely avoid such problems. Individuals can utilize their time more effectively during autonomous driving, engaging in activities such as work, meetings, watching videos, and enjoying the scenery. In conclusion, autonomous driving technology holds great importance for today's society.

As a result, in the future of the automotive field, automated systems are expected to receive increased attention and continuous development, ultimately bringing more convenience to people. The automotive interior automation system, in particular, holds significant development potential when compared to

autonomous driving techniques. The automotive interior automation system encompasses various aspects. By recognizing drivers' facial features, the system can analyze the interior requirements expressed by the driver's face and make predictions using a face analysis algorithm, incorporating face recognition, facial expression recognition, fatigue detection, and more. For example, if the driver appears uncomfortable, the system can adjust the seat and inquire about the preference, offering options to restore or readjust it. Simultaneously, the system can consider the actual environment; for instance, when the driver sneezes, and the external temperature is measured at 5 degrees Celsius, the system adjusts the air conditioner temperature accordingly. All the functions mentioned about the automotive interior automation system are closely tied to the face analysis technique, as individuals can express most of their feelings through their faces. The system's ability to accurately predict drivers' emotions and make appropriate adjustments to the car's interior undoubtedly enhances the driving experience.

The three most commonly used and mainstream face analysis algorithms familiar to us are Convolutional Neural Network (CNN), Principal Component Analysis (PCA), and FaceNet.

Convolutional Neural Network (CNN) stands out as a deep-learning neural network extensively employed for face analysis. It discerns features in images through convolution and pooling layers, subsequently performing classification through fully connected layers. Many advanced face analysis systems, including those based on ResNet or VGG network structures, leverage convolutional neural networks. The concept of CNN originated from the evolution of artificial neural networks, which exhibit exceptional self-learning and adaptability. As known, biological nervous systems possess characteristics such as parallelism, robustness, nonlinearity, and high fault tolerance[3]. These attributes have led to the widespread use of artificial neural networks in fields such as image processing and pattern recognition. The fundamental idea involves learning and analyzing the correlation pattern between input and output, subsequently applying this rule to make comparatively accurate predictions for the corresponding outputs of new input information. A key simplification in image processing is that CNN can directly input original images. CNN employs techniques such as local perceptual field analysis, weight sharing, and pooling to significantly minimize the number of trainable parameters.

Simultaneously, CNN demonstrates remarkable capabilities in handling image translation, rotation, and distortion, proving invaluable in deep learning research, particularly in computer vision. The field of computer vision encompasses tasks such as image classification, target recognition, and face analysis [4][5].

Principal Component Analysis (PCA), also known as Karhunen-Loeve (KL) transformation, is employed to address classification and image recognition problems, representing a fundamental technique in the realm of face analysis. The roots of PCA can be traced back to the early 20th century when Karl Pearson laid the mathematical foundations based on linear algebra and statistical concepts. Pearson's work involved identifying the eigenvectors and eigenvalues of the covariance matrix of the data, such as pixel values, to capture the maximum variance in the data[6]. In the context of face analysis, PCA is utilized to reduce the dimensionality of face images and preserve the most crucial information. This is achieved by projecting face images onto principal components, effectively reducing dimensionality and facilitating easier differentiation between different faces using the reduced feature set[7]. PCA exhibits two primary advantages: the capacity to extract and represent significant facial features and the enhancement of face analysis systems' robustness to variations in lighting, pose, and facial expressions. Its focus on the most significant patterns in the data contributes to high efficiency in face analysis by reducing the computational complexity of the recognition process, thereby significantly advancing computer vision and pattern recognition. However, traditional PCA has limitations in handling non-linear variations of face images, prompting the development of more advanced techniques such as Kernel PCA[8] and deep learning methods in the field of face analysis. Nevertheless, PCA remains a foundational building block in the development of face analysis systems.

FaceNet represents a cutting-edge face analysis technique that leverages deep learning and neural network architectures to map facial features into a high-dimensional space, known as a face embedding. Developed by researchers at Google, FaceNet employs a triplet loss function during training to optimize the embedding space, minimizing the distance between embeddings of the same person's face while

maximizing the distance between embeddings of different individuals. In alignment with contemporary studies utilizing deep networks, this methodology is entirely data-centric, deriving its representation directly from the pixel-level information of facial images[9]. FaceNet excels in crafting precise and condensed facial representations, enabling adept face analysis in diverse scenarios. An outstanding attribute is FaceNet's ability to perform both face verification and identification tasks using a unified model seamlessly, rendering it a versatile solution applicable to a spectrum of uses such as security systems, surveillance, and access control.

As facial recognition and automotive automation technologies continue their steady advancement, the prospect of integrating these innovations into a more intelligent automated interior system for vehicles emerges as a promising endeavor. The convergence of these technologies offers fertile ground for the development of a sophisticated interior system within automobiles. Such advancements hold the potential to substantially elevate the user experience, safety standards, and overall comfort levels within the automotive landscape.

2. Automotive Interior Automation System Design

The automotive interior automation system comprises various modules, including the image input module, facial analysis module, safe driving module, and comfort adjustment module. The subsequent subsections will provide a detailed explanation of these modules.

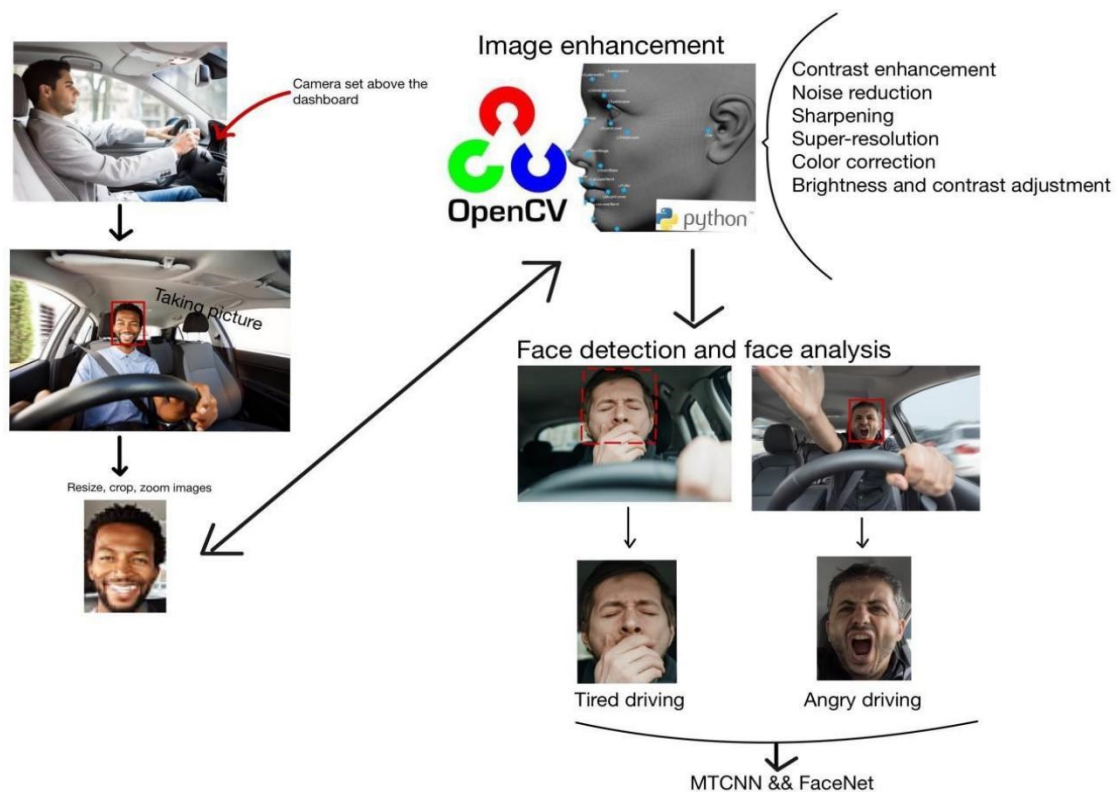


Figure 1. The flow diagram of the image input module and the facial analysis module.

2.1. Image Input Module

The image transmission module serves as a foundational component within the overall system framework. This critical module is responsible for acquiring, pre-processing, and subsequently transmitting photographs captured by in-car cameras. Its primary objective is to deliver these images to the system's face analysis module, which, in turn, performs a sequence of operations including face detection, positioning, feature extraction, and final feature matching or classification.

The camera designated for continuous driver monitoring is strategically positioned above the dashboard to ensure optimal and unobstructed image acquisition for precise face analysis. After image capture, the system engages in crucial preprocessing tasks to refine the acquired images. Preprocessing involves multiple stages, notably image adjustment, scaling, and image enhancement.

Image adjustment refines captured visuals to achieve optimal clarity and quality, ensuring that subsequent face analysis processes receive clear and standardized inputs. Scaling procedures aim to resize images to conform to specific dimensions or resolution criteria, facilitating standardized processing within the system. Furthermore, image enhancement techniques are employed to improve the overall quality of the acquired visuals, mitigating factors like poor lighting conditions or image noise, thereby enhancing the efficacy of subsequent face analysis processes. The utilization of widely acclaimed tools such as OpenCV significantly streamlines image adjustment and scaling processes within the system[10]. OpenCV, a renowned and extensively used computer vision library, offers a diverse array of image processing functionalities crucial to our objectives. Among its myriad capabilities, OpenCV provides tools for resizing, cropping, and scaling images, presenting a robust suite of functions essential to our image preprocessing workflow.

The library provides a user-friendly interface that enables the straightforward implementation of operations, such as resizing images to meet the specific input size requirements stipulated by the model. Notably, OpenCV's `cv2.resize()` function[11][12] emerges as a particularly valuable resource within this context, facilitating seamless resizing of images to align with the precise dimensions expected by the system's face analysis model[13][10][12]. The system's image enhancement procedures are designed to augment image quality and optimize visual impact through a comprehensive set of considerations[14]:

1. Contrast Enhancement: This involves emphasizing brightness differences across the image using methods like histogram equalization.
2. Noise Reduction: Addressing noise resulting from sensor limitations or acquisition processes is achieved through filters (e.g., Gaussian, median) and advanced denoising models.
3. Sharpening: Enhancing clarity and detail is accomplished through edge refinement using techniques like sharpening filters.
4. Super-Resolution: Visual fidelity is improved by upscaling low-resolution images for enhanced spatial resolution.
5. Color Correction: Accurate adjustments in color and tone are made to faithfully represent the original scene, enhancing realism.
6. Brightness and Contrast Adjustment: Fine-tuning overall brightness and contrast helps manipulate visual perception and aesthetic appeal.

The previously mentioned image enhancement processes can be efficiently achieved by leveraging the robust capabilities of OpenCV[11]. The comprehensive OpenCV database encompasses a multitude of photo preprocessing methods essential for enhancing image quality. For instance, the `cv2.equalizeHist()` function is effective in elevating image contrast through histogram equalization. Similarly, the `cv2.GaussianBlur()` function plays a crucial role in reducing noise, effectively addressing various sources of interference such as sensor limitations. Additionally, for precise color correction and standardization, the `cv2.cvtColor()` function stands as a reliable tool within the OpenCV framework. Employing these advanced tools within our system ensures a refined and comprehensive approach to image enhancement, thereby augmenting the quality and visual impact of captured images for subsequent analysis and processing stages.

2.2. Facial Analysis Module

Following the processing of driver-captured images by the image input module, attention shifts to the central core of the system: the facial analysis module. At this critical stage, the primary goal is to detect and localize faces within the image, subsequently extracting individual features using the FaceNet-based classification algorithm model[9]. This comprehensive procedure culminates in final feature matching and classification, determining system intervention in specific driving scenarios. For precise face

detection and localization, the multi-task cascaded convolutional neural networks (MTCNN) technique [15] has emerged as a highly efficient methodology. This approach integrates FaceNet, employing an improved loss function to achieve exceptional accuracy rates, reportedly reaching 99.85%.

MTCNN comprises three pivotal components, each specializing in different facets of face detection [16]. The Proposal Network (P-Net) constitutes the initial phase, generating candidate face frames within the image by proposing potential face locations and their respective confidence scores. Subsequently, the Refinement Network (R-Net) filters these candidate frames, refining the accuracy of bounding box locations and confidence scores. Finally, the Output Network (O-Net), the ultimate stage of MTCNN, precisely localizes remaining candidate frames, providing detailed bounding boxes and keypoint locations, thereby finalizing the detection process.

Leveraging multi-scale transformations, MTCNN constructs an image pyramid to facilitate multi-scale processing, effectively detecting faces of varying sizes and proportions. MTCNN's distinctive advantages stem from its ability to execute multi-task learning, concurrently handling face detection, keypoint localization, and bounding box regression. Moreover, its end-to-end training methodology enables comprehensive learning and optimization directly from raw image data.

MTCNN

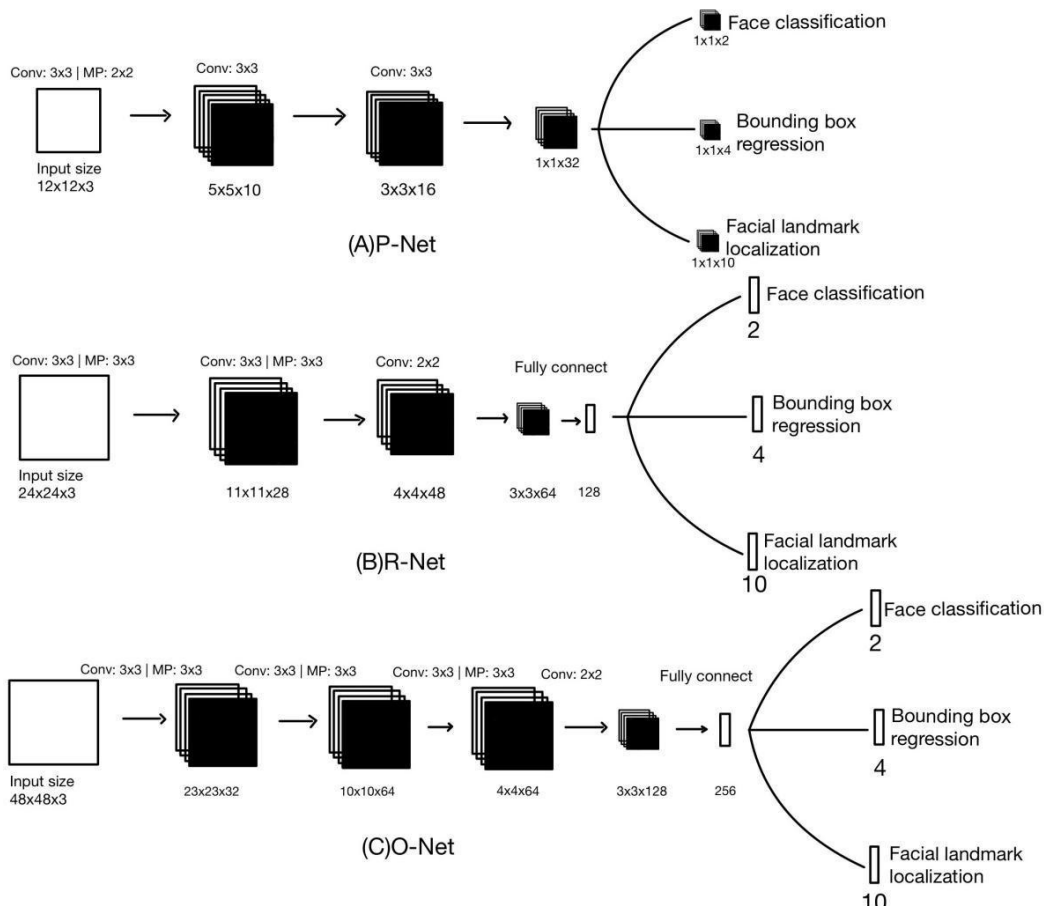


Figure 2. Three structures of MTCNN: P-Net network structure, R-Net network structure, and O-Net network structure. [16]

Upon the successful identification of faces within the driver's image by MTCNN, FaceNet assumes a pivotal role in conducting comprehensive facial analysis. Trained with an optimized loss function, FaceNet strives to map input images to high-dimensional vector representations, ensuring that images

of the same individual converge closely in this multidimensional space. This model adeptly transforms new facial images into feature vectors, thereby facilitating a nuanced understanding of facial characteristics.

Crucially, the foundation of this system rests on an ever-expanding dataset comprising labeled images that encapsulate various driver states, including but not limited to drowsiness, temperature discomfort, and emotional expressions such as anger. The strategic incorporation of FaceNet at a critical juncture in the program is geared towards achieving exceptional classification success rates. This, in turn, empowers the automated interior system of the vehicle to accurately discern the driver's current state. These advancements pave the way for practical applications in modules dedicated to ensuring safe driving and seamless comfort adjustments.

3. Application Modules

3.1. Safe Driving Module

Within the safe driving module, the interior automation system primarily focuses on various driving conditions, particularly when the driver is fatigued or faces an emergency. Advances in autonomous driving technology are poised to revolutionize the automotive industry, offering enhanced convenience and safety for drivers. A notable scenario involves the intelligent utilization of automation systems within vehicles. For example, an internal automation system, equipped with an image input module, captures and analyzes images of the driver. Upon detecting a yawning expression through its facial analysis module, the system identifies it as a potential sign of fatigue. Subsequently, it prompts the driver via the onboard display to inquire if autonomous driving intervention is desired. If the driver opts for intervention by selecting "Yes" from the displayed options, the system initiates the autonomous driving program.

However, fatigue-induced impairment may not always manifest through yawning; instances where a driver is in the process of falling asleep without overt signs like yawning are plausible. To address this issue, the system is designed to undertake increasingly nuanced assessments of driver photographs. It scrutinizes subtle cues such as blurred or lackluster eyes, a dull facial expression, relaxed muscles exhibiting diminished energy, or an unconsciously drooping head.

The system interprets facial characteristics within the driver's image as indicators of drowsiness and fatigue during driving. Upon recognizing these telltale signs, the system initiates a standardized procedure to validate the driver's state before intervening in the autonomous driving mode, contingent upon the driver's consent. However, in instances where the system fails to elicit a response from the driver within a three-second window, it autonomously intervenes in the autonomous driving mode as a precautionary measure to avert potential accidents. This proactive approach underscores the system's commitment to prioritizing driver safety by preemptively addressing instances of driver fatigue or drowsiness to mitigate the risks associated with impaired driving conditions.

The system conducts a parallel assessment of the driver's emotional state, with a particular focus on the possibility of anger, given its potential to trigger erratic driving behaviors leading to severe accidents. Through facial image analysis, the system identifies anger-related cues such as furrowed brows, teeth or lip clenching, tense jaw muscles, intense stares, and aggressive facial expressions. Detecting these indicators prompts the system to activate the autonomous driving mode and display a prompt urging calmness. This mode remains engaged until the system confirms that the driver's emotional state has stabilized, ensuring safe driving conditions. This proactive measure aims to mitigate risks associated with emotionally charged driving, aligning with the system's paramount objective of prioritizing driver safety.

This proactive approach exemplifies the potential of autonomous driving technology to intervene in critical situations, mitigating risks associated with driver fatigue and enhancing road safety. Through real-time monitoring and swift intervention, such systems aim to prevent accidents.

3.2. *Comfort Adjustment Module*

Within the comfort adjustment module, the system's primary function is to make suitable alterations based on the driver's indications of thermal discomfort or variations in physical comfort. This module predominantly oversees adjustments to the vehicle's air conditioning system. For air conditioning adjustments, the image input module continuously captures and transmits images of the driver to the facial analysis module to discern indications of thermal discomfort, whether related to heat or cold.

To address the system's evaluation of elevated temperatures, it relies on detecting perspiration on the driver's face, indicative of heightened internal car temperatures. In such instances, an initial assessment is made regarding the status of the air conditioning unit. If the system detects the air conditioner as inactive, it prompts the driver to activate it. Subsequently, if the system confirms the air conditioner's activation, it autonomously adjusts the temperature settings downwards, aiming to optimize the driver's comfort. Furthermore, in scenarios where the system identifies facial reddening in conjunction with an outdoor temperature exceeding 20 degrees Celsius, it interprets this as an indication of elevated internal temperatures, thereby implementing similar adjustments to enhance the driver's comfort.

Conversely, the system's evaluation of coldness relies on recognizing specific facial cues. These cues encompass muscular rigidity and involuntary trembling, predominantly around the cheekbones, chin, and lips. Additionally, responses to cold such as wrinkled eyebrows and tightened or trembling lips are considered. Cold conditions constrict blood vessels, resulting in a pallid appearance of facial skin. Facial flushing may also occur in response to cold, triggered by heightened blood circulation to maintain bodily warmth.

Upon identifying these facial manifestations indicative of cold discomfort, the system deduces that the internal car temperature is excessively low. It promptly initiates appropriate adjustments to raise the vehicle's temperature correspondingly, aiming to alleviate the driver's discomfort. This systematic approach highlights the sophistication of the comfort adjustment module in discerning and responding to drivers' thermal comfort needs, thereby contributing to an enhanced driving experience.

4. **Conclusions**

In the rapidly evolving domain of computer vision and machine learning, facial analysis algorithms and image detection systems have made significant strides in terms of both potency and stability. The reliability of these innovations is underscored by rigorous testing and real-world applications, confirming their robustness through successive iterations. Within this context, the feasibility of implementing an automated interior system for automobiles is becoming increasingly promising. The envisioned automated interior system for cars, as expounded in this discourse, represents not only a standalone innovation but also a pioneering platform for further evolution. It serves as a foundational canvas upon which a multitude of indispensable functionalities can be developed and integrated. As this technology matures, its widespread adoption across various sectors beyond automotive applications promises heightened convenience and efficiency.

In summary, the introduction of an automated interior system within automobiles marks a significant stride, heralding a new era of amplified functionality and efficacy. This evolutionary shift is poised to redefine the driving experience, enhancing it with unprecedented capabilities and innovations, paving the way for an era of transformative advancements across diverse fields.

References

- [1] Yurtsever, E., Lambert, J., Carballo, A., & Takeda, K. (2020). A survey of autonomous driving: Common practices and emerging technologies. *IEEE access*, 8, 58443-58469.
- [2] Rosenzweig, J., & Bartl, M. (2015). A review and analysis of literature on autonomous driving. *E-Journal Making-of Innovation*, 1-57.
- [3] Wang, D., Yu, H., Wang, D., & Li, G. (2020, April). Face recognition system based on CNN. In *2020 International Conference on Computer Information and Big Data Applications (CIBDA)* (pp. 470-473). IEEE.

- [4] Kamencay, P., Benco, M., Mizdos, T., & Radil, R. (2017). A new method for face recognition using convolutional neural network. *Advances in Electrical and Electronic Engineering*, 15(4), 663-672.
- [5] Xie, Z., Li, J., & Shi, H. (2019, November). A Face Recognition Method Based on CNN. In *Journal of Physics: Conference Series* (Vol. 1395, No. 1, p. 012006). IOP Publishing.
- [6] Shah, J. H., Sharif, M., Raza, M., & Azeem, A. (2013). A Survey: Linear and Nonlinear PCA Based Face Recognition Techniques. *Int. Arab J. Inf. Technol.*, 10(6), 536-545.
- [7] Kolenikov, S., & Angeles, G. (2004). The use of discrete data in PCA: theory, simulations, and applications to socioeconomic indices. Chapel Hill: Carolina Population Center, University of North Carolina, 20, 1-59.
- [8] Sarma, P., Durlofsky, L. J., Aziz, K., & Chen, W. H. (2007, February). A new approach to automatic history matching using kernel PCA. In *SPE Reservoir Simulation Conference?* (pp. SPE-106176). SPE.
- [9] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 815-823).
- [10] Bradski, G., & Kaehler, A. (2000). OpenCV. *Dr. Dobb's journal of software tools*, 3(2).
- [11] Bradski, G., & Kaehler, A. (2008). *Learning OpenCV: Computer vision with the OpenCV library*. " O'Reilly Media, Inc."
- [12] Howse, J. (2013). *OpenCV computer vision with python* (Vol. 27). Birmingham: Packt Publishing.
- [13] Kopf, J., Neubert, B., Chen, B., Cohen, M., Cohen-Or, D., Deussen, O., ... & Lischinski, D. (2008). Deep photo: Model-based photograph enhancement and viewing. *ACM transactions on graphics (TOG)*, 27(5), 1-10.
- [14] Girgensohn, A., Adcock, J., & Wilcox, L. (2004, October). Leveraging face recognition technology to find and organize photos. In *Proceedings of the 6th ACM SIGMM international Workshop on Multimedia information Retrieval* (pp. 99-106).
- [15] Wu, C., & Zhang, Y. (2021). MTCNN and FACENET based access control system for face detection and recognition. *Automatic Control and Computer Sciences*, 55, 102-112.
- [16] Ku, H., & Dong, W. (2020). Face recognition based on mtcnn and convolutional neural network. *Frontiers in Signal Processing*, 4(1), 37-42.