Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 2 (2024)

A New ID Verification and Image Recognition Framework for Risk Management

Zhicong Chen [1], Alice Xiaodan Dong [2]
[1, 2] *The University of Technology Sydney, Ultimo, Australia*

Abstract:- In the era where identity information is widely applicable in modern society, accurate identity authentication has become a focus of social attention, whether it is in the fields of finance, transportation, insurance, etc. In Australia, driver's licenses are widely used as a form of identification. Nevertheless, there is a growing trend of fraudsters creating counterfeit driver's licenses, deceiving both government entities and financial institutions, and causing substantial financial losses. This paper presents a new framework tailored for driver's license identity portrait verification. Utilizing a deep learning model and advanced preprocessing techniques, the framework is specifically designed to enhance risk management effectiveness. Experimental results indicate that the proposed Convolutional Neural Network modeling framework with Error Level Analysis (ELA) preprocessing exhibit notably higher accuracy compared to the traditional models.

Keywords: Risk Management, Deep learning, Convolutional neural network.

1. Introduction

Accurate identification of image authenticity holds considerable importance in risk management in the financial sector. It not only aids in preventing fraudulent activities and ensuring the authenticity of financial transactions and documents but also enhances document processing efficiency and fulfills compliance requirements. This means image authenticity resolution technology brings higher security, compliance, and efficiency to the financial system.

Having a driving license is an important part of life for many Australians, as it facilitates social contacts and access to employment, education, and health services [1]. At the same time, in many countries, as technology has developed, criminal activities related to driving licenses have also increased. In many interactive scenarios, especially financial scenarios, users' identity information needs to be authenticated to prevent identity theft [2]. Acceptable ID forms are created by various authorities and include features such as light-sensitive stripes, ghosting, and material properties to help examiners distinguish between real and fake IDs [3]. In Australia, a driver's license is often one of the available forms of identification. However, even if the document's portrait photo is edited or faked, it is often considered a real document because verification tools often lack image analysis functionality and cannot detect the authenticity of the image [3]. Editing real photos via computer software or mobile apps is now one of the easiest things people can do [4]. Although most people do this for fun, there are also cases where ID images are maliciously altered for criminal purposes.

The natural question is whether it can help relevant personnel verify identity information while determining whether the portrait image on the driver's license has been maliciously altered or processed. In practical scenarios, fraudsters commonly engage in identity theft by manipulating the portrait image. For instance, when the risk management system requests individuals to provide a selfie photo with facial features matching their driver's licenses, fraudsters may attempt to alter the portrait image on the driver's licenses. To safeguard against such fraudulent activities, it is essential to intervene and scrutinize the facial image on the user's driver's license to ensure it has not been tampered with maliciously. We intercept the facial image on the driver's license provided by the user using deep learning techniques and check whether the image has been maliciously altered. Most people in Australia hold driver's licenses. Depending on when the driver acquired the license, the license used, including the holder, differs. ID card information and facial image information, as shown in Figure 1 and Figure 2.

Figure 1. The sample of NSW Full driver licence

Figure 2. The sample of NSW P1 driver's licence

Efforts to streamline application processes, led by institutions such as financial entities and government departments, involve automating the verification of customer-provided selfie photos against driver's license images. However, a critical hurdle emerges as fraudsters seek to manipulate driver's license images to exploit these automated systems. This paper focuses on advancing risk management in automated application processes, representing the primary focus of our research. The rest of this article is organized as follows.

In the next section, related work is discussed. Section 3 describes the proposed authentication framework in detail. Section 4 presents the implementation details and experimental results. Section 5 concludes our work.

2. Related Works and Dataset

Facial recognition is one of the research areas of computer vision and has received more and more attention with the development of the field of artificial intelligence [5]. Face verification is a subfield of face recognition [6]. The focus is on verifying whether two facial images belong to the same person, which is also known as one-to-one face verification [7]. When the research direction includes document information or research based on an identity verification framework, facial recognition and recognition technology are required [8]. The reason is that the main purpose of face detection and recognition is to compare the face images present in the dataset with the intercepted face and determine whether the intercepted face is real or not [9]. This is to ensure that the owner of the intercepted portrait is the right person and to prevent the portrait image from being used maliciously.

With the spread of computer graphics processing technology and image editing software, people can modify and retouch images in a short time and at a low cost [10]. This poses challenges for the authenticity of the identification. Some criminals use these technological means to create fake identification documents or driver's licenses and use them for various criminal activities. For example, you may use someone else's identity to commit fraud or use a fraudulent driver's license to evade law enforcement. This has become an important issue worldwide [11]. However, by building a large-scale facial image database, a large number of real facial samples are collected, including various fake facial samples [12]. The deep learning model is then trained to classify these real and fake facial samples. This method is considered feasible and makes this area an important sub-area in the field of facial recognition.

Authenticity detection of portrait images based on a convolutional neural network can also be viewed as a binary classification problem [13]. First, a large number of real portrait images and various manipulated portrait images are collected to create a dataset. Typical sources of portrait images are facial recognition datasets, social media avatars, etc. The subject of this research is portrait images of driver's licenses. Because ID images are sensitive information, they are difficult to find on database websites (Kaggle) or social media. As an alternative, we found a sample image of a driver's license online that did not contain any sensitive or private information. We leveraged insights gained from practical experiences in fraud investigation to generate modified driver's licenses. Utilizing editing software, we manipulated images, simulating the techniques employed by fraudsters. Finally, these two image sets are combined to serve as the base for training and validation data sets for the Convolutional Neural Network Classifier (CNN Classifier).

3. Methodology

To improve the reliability and tamper-proofness of driver's license authentication, the key is to verify whether the face in the uploaded image belongs to the driver. The goal of this research is to design an end-to-end convolutional neural network framework to solve recognition and authentication problems between drivers and driving documents. Given that this accomplishment can address challenges like enhancing image forgery techniques and minimizing manual verification expenses, it is imperative to advance automated deep learning models for achieving automatic visual recognition of the authenticity of driving license facial images.

In recent years, convolutional neural networks have made great progress in visual recognition tasks due to their advantages in automatic feature learning [14]. We believe that with this progress, it will be possible to combine novel data processing methods and create a reasonable convolutional neural network architecture [15]. Through a variety of preprocessing methods with different functions, the model is supplied with pre-processed data from different technologies, and this part of the data is used to train the model to overcome the shortcomings of the CNN model that cannot distinguish images alone. Finally, we used three different preprocessing techniques to build a CNN model to identify a dataset consisting of about 2,500 images, and through one-to-one and one-to-many comparisons, the model was able to initially distinguish authentic images and fake portraits.

3.1. Preprocessing Techniques

3.1.1. Face Detection and Cropping Module

To accurately locate the facial area in the driver's license image, we built a cascade classifier based on Haar features and AdaBoost training [16]. Haar features use pixel differences to encode visual patterns such as edges and lines. The classifier consists of a cascade of multiple weak classifiers that can efficiently filter non-face areas. In OpenCV, we load a data file with facial features on the front to recognize driving license images. The specific process is that the classifier iteratively scales the detection window with a ratio of 1.1, calculates the Haar feature similarity, and determines whether it contains a facial pattern. We set the minimum number of coincidence detections to 5 to filter out false positive samples. Finally, the rectangular frame area of the detected face is intercepted and resize to 128x128, which is used as the default image size of the input layer of the CNN model. This ensures that the input surface area is uniform and favours model training.

3.1.2. Image Enhancement Module

To deal with the impact of changes in the driver's head pose and distance in practical applications, we developed a data enhancement process based on the cropped face to improve the robustness of the model. Specifically, we created a custom image enhancement function based on affine transformation, including: 1. Rotation transformation: The input surface is randomly rotated by -15 to 15 degrees; 2. Scaling transformation: The input area is randomly scaled from 0.8 to 1.2 times; 3. Translation transformation: The input surface is randomly translated along the horizontal and vertical axes by a distance of ±20 pixels. The above three transformations can generate new samples with different angles, distances, and positions of the face, which is very helpful for checking the consistency of model judgments. To verify the transformation effect, we visualized and saved the improvement results separately. This diversified dataset generation strategy can substantially enhance the model's robustness in determining the authenticity of faces, marking a key innovation in this method.

3.2. Modelling Methodology

3.2.1. CNN Extraction

The first convolution layer contains 16 three-dimensional convolution kernels with a size of 3 × 3 × 3, where 3 × 3 is the spatial dimension and 3 is the RGB-3 channels of the input image. The step is set to 1, the padding strategy is undefined, and the default is the original setting of TensorFlow. The convolution kernel calculates the element-wise clear product of each region of the input image and its corresponding weight, then sums it and outputs it as a feature map after nonlinear transformation of the ReLU activation function. This layer extracts basic visual features such as low-level edges, textures, and corners from the input image.

$$y = f(w * x + b)$$

Figure 3. Convolution layer formula

Connect a maximum pooling layer with a kernel size of 2×2 and a stride of 2. The maximum value of each 2×2 region in its space is used as the corresponding position of the output feature map to obtain the function of obtaining key features to achieve image down sampling reduction. Improve feature robustness to deformation and lighting changes and reduce network parameters.

$$y = down(max(x))$$

Figure 4. Pooling layer formula

The number of filter channels in the second convolutional layer increases to 32, indicating that the learned feature types or patterns increase. The kernel size remains 3x3, step 1 remains unchanged, the padding strategy is undefined, and the default values are the same. Continue the folding calculation of the sliding window. Different filters can separately learn different feature patterns in local areas of the image, mainly obtaining DETAILED FEATURES of eyes, nose, mouth, and other parts. The number of convolution kernel parameters is significantly increased, improving the network's ability to remember complex functions.

The number of filters in the third convolution layer is reduced to 16 and other parameters such as kernel size 3x3, step size 1 and padding mode are retained by default. This convolutional layer extracts highly abstract overall contour features of the face shape as well as more global features such as missing subtle location features.

Then another maximum pooling layer with the same core parameters. By compressing feature representative information, the robustness of the response to position and deformation can be improved, which has a positive effect on the positioning of the facial area. Collapse all feature map data connections and enter them into the fully connected layer. The number of neurons in this layer is 256. The fully connected structure strengthens the nonlinear combination interaction between the overall features and combines the scattered local features into the overall feature pattern required for the discrimination task. Significantly increase network parameters.

$$y = f(Wx + b)W$$

Figure 5. Fully connected layer formula

The final output layer is fully connected to the single neuron and is linked to the sigmoid activation. Based on the comprehensive features, a prediction probability of 0-1 is given to determine the authenticity of the face.

$$L(y, y') = -(ylog(y') + (1 - y)log(1 - y')$$

Figure 6. The loss function uses the cross-entropy function

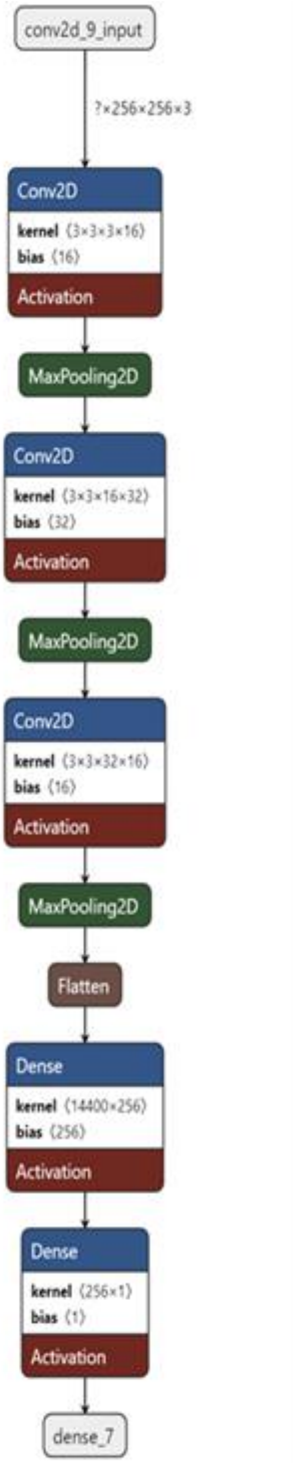

Figure 7. The flow chart of modelling

3.2.2. Feature Extraction Module of Convolutional Neural Network

The convolutional neural network is responsible for learning the distinguishing features of facial images. We built a network structure containing three groups of convolutional computing units. In the first and second group of units, two consecutive 3x3 convolution operations are first performed, and the number of convolution kernels is 16 and 32, respectively. As a result, low- to high-level features of the face are captured hierarchically through a different number of filters. Additionally, we set the step to 1 to ensure that the size of the feature map remains unchanged. After each set of convolutions, a 2x2 max pooling operation is performed to halve the down-sampled feature map resolution. This can both reduce the number of parameters and improve the spatial invariance of features. All convolutional layers consistently use the ReLU activation function to accelerate network convergence. Finally, a global max-pooling layer outputs a 512-dimensional feature vector of the entire network, representing the semantic features of the input face image.

3.2.3. Dual Classification Output Layer

After the features extracted by the CNN are processed by the fully connected layer, the model assigns these features to a binary classification output layer. The role of this output layer is to use the previously extracted features as well as the weights and biases learned by the model to ultimately produce a binary prediction value indicating whether the input image is real or fake. This prediction ability is achieved through the model's learning and feature extraction of image data, allowing it to accurately classify the authenticity of the image.

3.2.4. Optimization Module for Model Training

To optimize the performance of the model, the binary cross-entropy loss function is used as the loss measure of the model during the training process. The difference between the model prediction and the actual truth is measured by this loss function. For parameter updating, the Adam optimization algorithm is used to update the model parameters during each epoch iteration. Depending on the learning rate setting, the model is optimized to improve the accuracy of its prediction. This training process was repeated for 20 epochs, allowing the model to better learn image features and gradually improve its ability to distinguish the authenticity of images.

3.2.5. Facial Recognition and Interception

In the module dedicated to face detection, we use OpenCV's classic Haar feature cascade classifier model to detect faces in images [16]. This model is designed to leverage the capabilities of a pre-trained facial feature model and uses techniques such as zooming and sliding over the image to identify potential facial areas. Once these potential facial areas are identified, they are subjected to a filtering process based on a predefined minimum number of adjacent elements to meticulously eliminate any possible false positives that could occur.

The haarcascade_frontalface_default.xml is a classifier trained to recognize faces in images. It is based on a cascade classifier of Haar features, a method that can identify objects in images. This special XML file contains a model for frontal face recognition. It works by moving windows of different sizes in the image and applying pre-trained feature templates (Hair features) to detect the presence of faces. These features are created based on changes in pixel brightness in different areas. The cascade classifier applies a series of features to gradually exclude areas that are not faces and finally determines the position of the face. This XML file is part of the OpenCV library and can be used in many programming languages for facial recognition in images or videos and is the basis of many facial recognition systems and applications.

Following this phase, the Facial Cropping and Preservation submodule uses the precise coordinates of the detected facial areas from the previous step to perform precise cropping of the image while respecting the positional information from the rectangular frame. These cropped facial images then undergo a standardization process to ensure they conform to a consistent size of 128 x 128 pixels before being systematically archived into a specific folder. The nomenclature assigned to these images strictly follows the format of the face_number_source file name and is carefully structured for future use and reference. In addition, a range of data augmentation features, involving sophisticated techniques such as rotations, scaling, and translations, have been carefully implemented to significantly increase the size and diversity of the dataset. These augmentation methods have significantly contributed to strengthening the training efficiency of the model, thereby significantly improving its ability to recognize and generalize facial images.

3.2.6. Image Enhancement

To recognize images well and ensure the efficiency of the proposed framework, a representative image dataset is necessary. However, access to driver's license images is very limited and difficult because driver's licenses are personal items and store very sensitive information. To achieve this goal, we considered two options. Firstly, search for suitable existing datasets on the World Wide Web. Secondly, we use artificial intelligence to create a portrait, and then add the portrait image to the driver's license photo board. Additionally, we also include driving license templates from the USA and India that we found on the internet. Each image has a different pixel size and needs to be standardized in subsequent steps to facilitate the training process.

The cropped face photos undergo of several changes within the picture improvement module. These conversions include rotational modifications, scaling operations, and positioning changes. Interestingly, these diverse transformation techniques serve a dual purpose: significantly expanding the pool of available samples and introducing heightened complexity and diversity into the model, thereby enhancing its robustness and adaptability.

Through the implementation of these data augmentation techniques, the model experiences a significant transformation, improving its ability to handle facial variances in a range of positions and orientations. As a result, this strengthens the model's ability to analyse images in different environments. The increased diversity of datasets greatly aids in improving the model's training process, giving it stronger generalization skills and an improved capacity to manage a wide range of real-world image scenarios.

3.3. Image Processing

Within the realm of our third module dedicated to image preprocessing, our strategy encompasses a multifaceted approach integrating ELA (Error Level Analysis), and gradient analysis techniques for meticulous and comprehensive handling of image data and comparison to the original dataset at the same time.

As a foundational method, ELA serves as a forensic instrument to detect any possible picture changes or alterations. This is achieved by looking at the differences in error levels between the compressed and re-saved version of the image and the original. This thorough examination helps to confirm an image's legitimacy and reveals the editing history of the image, providing important context for any modifications that may have been performed.

Our strategy for addressing photo noise relies on statistical calculations to evaluate the standard deviation of pixel values. This statistical analysis accomplishes two goals: it measures the amount of noise in the picture and opens the door for future noise reduction techniques. We can spot irregularities that may impact further processing and obtain a more profound comprehension of the image's integrity by identifying these noise patterns.

Furthermore, the incorporation of gradient analysis techniques improves our comprehension of picture features, specifically edges, and changes in the image. We carefully examine these gradients using operators such as Sobel and Laplacian, which provide fine details about edges and small changes in the picture. This thorough examination gives us a better understanding of the composition and content of the image.

The data pre-processed using these three methods is placed in different class folders. Each class folder has two subfolders, “real” and “fake,” which store data from different sources. Each set of real and fake data is trained by the CNN model. The purpose is to observe different results through different training models to assess the quality of the preprocessing techniques.

3.4. Model Training and Evaluation

3.4.1. Model Training

Based on the constructed convolutional neural network, the model parameters are iteratively updated according to the epoch based on the training set batch samples. Choose Adam optimization algorithm to dynamically adjust the parameter update amplitude to avoid gradient disappearance and explosion problems. Set the binary cross-entropy loss function to measure the difference between the predicted output and the ground truth. The training goal is to minimize the loss function. After 20 rounds of Epoch training, the model can remember representative facial features and perform parameter adjustments.

3.4.2. Performance Evaluation

To evaluate the model performance, we use an independent test dataset. As depicted in Table 1, the best model involves combining CNN with ELA. The model achieves a recall rate reaches 96.875%. This confirms the robustness of the model when detecting fake samples. The accuracy rate is 98.3%, which overall confirms the efficient learning and memory ability of the convolutional neural network structure on driving license facial recognition features. Furthermore, we also consider the P-R curve and the confusion matric. The P-R curve comprehensively describes the dynamic trade-off relationship between precision and recall; The confusion matrix intuitively represents the prediction distribution of different sample types. It effectively demonstrates the detailed performance of the model in distinguishing real and fake facial images, offering valuable insights for model optimization.

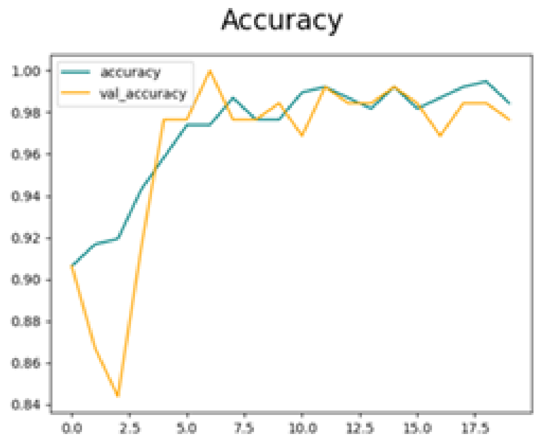

Figure 8. Accuracy of ELA

Using the dataset augmented by gradient analysis, we re-evaluated the model effect and obtained the following indicators: The accuracy rate is 55.56%, indicating that almost half of the samples judged by the model to be real faces are incorrect. The low precision value reflects that the model has certain false negatives when assessing this type of image. The recall rate reached 100% and the model successfully recognized all real face samples. This shows that after improving the data based on gradient information, the completeness of the model in judging real facial images is improved. Overall, the accuracy rate is 66.67% and there is still room for further improvement. Further analysis of image and gradient response features will help improve the model's ability to distinguish complex samples and achieve higher face recognition accuracy.

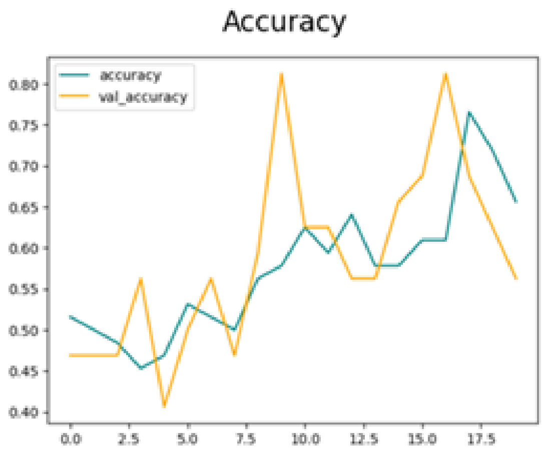

Figure 9. Accuracy of gradient analysis

To evaluate the effect of enhanced learning, we test the model on original face data and gradient-enhanced data respectively. We found that the model loss was 0.68 and the accuracy was only about 56% on the original image data. In comparison, when re-evaluated on the data set enhanced by gradient analysis, the model's recall rate increased to a perfect 100%. This shows that the use of additional visual features mined in image processing and analysis methods indeed enhances the model's coverage of judging real faces.

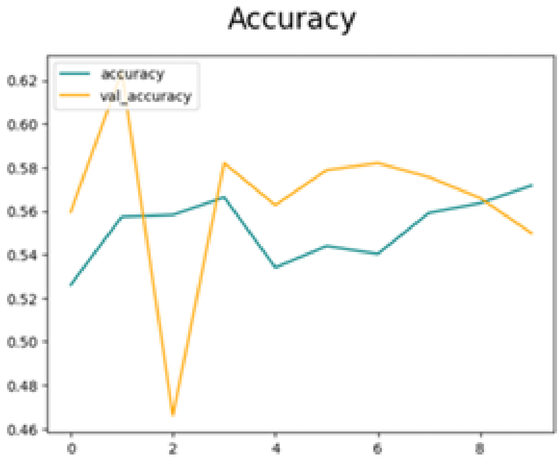

Figure 10. Accuracy of original datasets

Table 1. Model Permormance

Dataset	Accuracy	Lost	Recall
ELA	0.98	0.0103	0.96
Gradient Analysis	0.66	0.6325	1.0
Original	0.56	0.6952	1.0

4. Discussion of Future Work

The framework proposed in this paper for recognizing the facial authenticity of driver's licenses, built upon a deep convolutional neural network, has exhibited precise and resilient classification and discrimination capabilities compared to existing models. Considering the promising outcomes, we suggest that future research can further enhance and fortify these results by exploring the following dimensions:

Firstly, we could use a larger facial image dataset to verify the generalization performance of the model. This is attributed to the necessity of addressing diverse facial features in real-world application scenarios. By leveraging extensive datasets, the model undergoes rigorous testing for robustness against intricate and altered facial images, ensuring its suitability for practical deployment.

Secondly, assessing the model's stability with recent samples becomes imperative, given the rapid evolution of fraudster techniques. This involves simulating adversarial attacks of a specific scale, such as fine-tuning parameter attacks and occlusion attacks, to gauge the sensitivity of intrinsic features learned by the CNN model to noise and interference stimuli. Furthermore, we can introduce an attention mechanism to improve the model structure. The attention structure allows the network to autonomously focus on the core areas of the recognition sample, such as key points on the face. This can improve hard sample detection and achieve the goal of improved recall.

Moreover, the creation of a fraudster dataset and the application of transfer learning technology are viable strategies. This involves leveraging large-scale pre-trained model parameters to optimize and adjust the model for adaptation to the distribution of facial data in this specific field. Utilizing existing model knowledge not only reduces training requirements but also synergistically enhances performance.

To expand practical application scenarios, an efficient deployment solution for the model on mobile devices or IoT devices will finally be developed. This requires optimization technologies such as model compression, compiled deployment, and inference acceleration. Achieving high-precision real-time facial recognition has diverse application prospects.

5. Conclusion

To prevent the driver's license from being used maliciously, the identity must be authenticated carefully and carefully. However, the traditional authentication method is very cumbersome. To find a solution, an identity authentication framework based on machine learning algorithms is proposed. The execution mode of this framework is to compare the holder's ID portrait with the original portrait data in the database to ensure the authenticity of the identity. The advantages of the framework are mainly reflected in reducing the interaction process between people, accurately identifying documents through machine learning, detecting malicious changes that are difficult to detect with the naked eye, and comparing pre-processed data through the CNN model, to quickly achieve detection results. The experiment outcomes affirm the framework's superior accuracy compared to traditional models, ensuring robust identity verification reliability. This also attests to the practicality and success of the proposed framework.

References

[1] Angell, B., Cullen, P., Laba, T., Lung, T., Shanahan, M., Sakashita, C., Eades, S., Ivers, R., & Jan, S. (2018). What is the value of a driver licence? A contingent valuation study of Australian adults. *Transportation Research Part A: Policy and Practice*, *108*, 25–34. https://doi.org/10.1016/j.tra.2017.12.010

[2] Nokhbeh Zaeem, R., Manoharan, M., Yang, Y., & Barber, K. S. (2017). Modeling and analysis of identity threat behaviors through text mining of identity theft stories. *Computers & Security*, *65*, 50–63. https://doi.org/10.1016/j.cose.2016.11.002

[3] Weatherford, D. R., Erickson, W. B., Thomas, J., Walker, M. E., & Schein, B. (2020). You shall not pass: how facial variability and feedback affect the detection of low-prevalence fake IDs. *Cognitive Research: Principles and Implications*, *5*(1). https://doi.org/10.1186/s41235-019-0204-1

[4] Zheng, L., Zhang, Y., & Thing, V. L. L. (2019). A survey on image tampering and its detection in real-world photos. *Journal of Visual Communication and Image Representation*, *58*, 380–399. https://doi.org/10.1016/j.jvcir.2018.12.022

[5] Hangaragi, S., Singh, T., & N, N. (2023). Face Detection and Recognition Using Face Mesh and Deep Neural Network. *Procedia Computer Science*, *218*, 741–749. https://doi.org/10.1016/j.procs.2023.01.054

[6] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/cvpr.2015.7298682

[7] Sengupta, S., Chen, J.-C., Carlos Fernandez-del Castillo, Patel, V. M., Rama Chellappa, & Jacobs, D. R. (2016). Frontal to profile face verification in the wild. *Workshop on Applications of Computer Vision*. https://doi.org/10.1109/wacv.2016.7477558

[8] He, T.-C., Tian, Z., Huang, W., Shen, C., Qiao, Y., & Sun, C. (2018). *An End-to-End TextSpotter with Explicit Alignment and Attention*. https://doi.org/10.1109/cvpr.2018.00527

[9] Mo, S., Lu, P., & Liu, X. (2022). *AI-Generated Face Image Identification with Different Color Space Channel Combinations*. *22*(21), 8228–8228. https://doi.org/10.3390/s22218228

[10] Kee, E., O'brien, J., & Farid, H. (2013). Exposing photo manipulation with inconsistent shadows. *ACM Transactions on Graphics*, *32*(3), 1–12. https://doi.org/10.1145/2487228.2487236

[11] Farid, H., & Bravo, M. J. (2010). Image forensic analyses that elude the human visual system. *SPIE Proceedings*. https://doi.org/10.1117/12.837788

[12] Wang, X., Wang, K., & Lian, S. (2020). A survey on face data augmentation for the training of deep neural networks. *Neural Computing and Applications*. https://doi.org/10.1007/s00521-020-04748-3

[13] Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Niessner, M. (2019). FaceForensics++: Learning to detect manipulated facial images. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. https://doi.org/10.1109/iccv.2019.00009

[14] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

[15] Tarasiuk, P., & Szczepaniak, P. S. (2021). Novel convolutional neural networks for efficient classification of rotated and scaled images. *Neural Computing and Applications*, *34*(13), 10519–10532. https://doi.org/10.1007/s00521-021-06645-9

[16] Mohamed, A., Issam, A., Mohamed, B., & Abdellatif, B. (2015). Real-time Detection of Vehicles Using the Haar-like Features and Artificial Neuron Networks. *Procedia Computer Science*, *73*, 24–31. https://doi.org/10.1016/j.procs.2015.12.044