Facial Feature Classification of Drug Addicts Using Deep Learning

Chi-Hung Chuang, Cheng-Tan Tung, Yuan-Song Chang, Edward Lin, Chih-Ping Yen *

Abstract—**Today, closed-circuit television (CCTV) is widely used in various fields. If authorized by law, CCTV can be used to identify drug users in public places or specific areas. It can estimate the drug population and determine the hot spots of drug proliferation to comprehend and control the overall drug situation in a region. Based on the GoogLeNet network of deep learning, this work integrated image processing, face detection, data augmentation, and transfer learning to establish a special face image recognition framework that could differentiate drug users from nonusers. The facial database with drug addicts and nonaddicts used in the experiment was obtained from the Internet, face benchmarks, and police agencies. This facial database was diverse and included different genders, ages, and races. The experimental results showed that the model accuracy in this study reached 87.14%, implying that the model could learn the low-dimensional facial discrimination features of drug addicts. The three classification performances, precision, recall, and F1-score reached 87.32%, also representing a good overall classification performance of the model. In addition, we also confirmed that adjusting the activation function and the dropout probability could further optimize the model's performance.**

Index Terms—**Closed-Circuit Television (CCTV), deep learning, data augmentation, transfer learning, feature classification**

I. Introduction

In recent years, the number of closed-circuit televisions (CCTVs) in private areas (stores, banks) and public areas (stations, public roads) has continually increased. Using the videos and photos of different crimes captured by the CCTVs to conduct facial image comparisons has been gradually brought to the forefront of crime investigations and judicial scenes. CCTV footages help us to maintain social security, such as assistance in finding missing elderly and children, finding perpetrators of violence in public places, investigating contacts during disease control, and hunting down wanted criminals [1-3]. Under the authority of the law, CCTV is used to identify drug abusers in public places or in certain designated areas to estimate the number of drug users, determine the hot spots of drug abuse, and assist police in pursuing drug sources. This helps in grasping and controlling the overall drug situation in an area, and provides a new way of cross-domain drug law enforcement and governance.

In Fig. 1, the catastrophic effects of drug abuse on human health are obvious, and over time, the chemicals in the drugs (methamphetamine, cocaine, heroin, ecstasy, cannabis, or painkillers, etc.) can cause a serious damage to various parts of the human body, especially the facial appearance led by the loss of facial collagen, fat, muscle loss, and tooth deterioration, further resulting in skin wrinkles, loss of elasticity, depression, redness, and rash [4], etc.

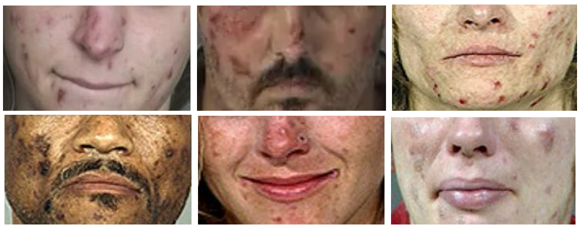


Fig. 1. Facial features of drug addiction.

II. Related work

Over the past few years, the convolutional neural network (CNN) of deep learning has been widely regarded as a reliable method for object detection and classification. Especially the various winning trainers proposed in ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [5] over the years, such as AlexNet [6], ZFNet [7], VGGNet [8], GoogLeNet [9], ResNet, SENet [10]. These trainers have continuously reduced the error rate and have also brought about a wave of applications in various research fields.

A. Deep Learning Models for Face Recognition

As mentioned above, great progress has been made in face recognition. And among these trainers, GoogLeNet and ResNet have better accuracy [11], but the complexity of the ResNet model is much higher than that of GoogLeNet. Based on this, we will conduct this research on the basis of GoogLeNet.

GoogLeNet is a deep and wide deep learning architecture proposed by Christian Szegedy [9]. As shown in Table I, the GoogLeNet architecture consists of 22 layers, the input data size is 224×224×3, and the final output can predict 1,000 categories.

Manuscript received September 04, 2022; revised April 09, 2023. This research was funded by National Science and Technology Council under grant number MOST 111-2410-H-015-001-SS2.

Chi-Hung Chuang is an Associate Professor of Information and Computer Engineering Department, Chung Yuan Christian University, Taoyuan 320314, Taiwan, ROC (e-mail: chchuang640509@gmail.com).

Cheng-Tan Tung is an Assistant Professor of Information Management Department, Central Police University, Taoyuan 333322, Taiwan, ROC (e-mail: tung@mail.cpu.edu.tw).

Yuan-Song Chang is an Assistant Professor of Foreign Affairs Police Department, Central Police University, Taoyuan 333322, Taiwan, ROC (e-mail: una189@mail.cpu.edu.tw).

Edward Lin is a graduate student of Electrical Engineering at Stanford University, Stanford, CA 94305, USA (e-mail: elin24@stanford.edu).

Chih-Ping Yen is an Associate Professor of Information Management Department, Central Police University, Taoyuan 333322, Taiwan, ROC (corresponding author, e-mail: peter@mail.cpu.edu.tw).

TABLE I
THE ARCHITECTURE OF THE GOOGLENET

Type	Filter size/stride	Output size	Depth	#Params
Convolution	7x7/2	112x112x64	1	2.7K
Max pooling	3x3/2	56x56x64	0	
Convolution	3x3/1	56x56x192	2	112K
Max pooling	3x3/2	28x28x192	0	
Inception (3a)		28x28x256	2	159K
Inception (3b)		28x28x480	2	480K
Max pooling	3x3/2	14x14x480	0	
Inception (4a)		14x14x512	2	364K
Inception (4b)		14x14x512	2	437K
Inception (4c)		14x14x512	2	463K
Inception (4d)		14x14x528	2	580K
Inception (4e)		14x14x832	2	840K
Max pooling	3x3/2	7x7x832	0	
Inception		7x7x832	2	1,072K
Inception		7x7x1,024	2	1,388K
Avg pooling	7x7/1	1x1x1,024	0	
Dropout (40%)		1x1x1,024	0	
Fully connection		1x1x10,575	1	1,000K
Softmax		1x1x10,575	0	
Classification output				

The main features of GoogLeNet include the following:

1) the introduction of nine inception module structures to fuse feature information of different scales, as shown in Fig. 2. First, perform a 1×1 convolution operation on the previous layer feature maps to reduce the number of channels (dimensionality reduction) and reduce training parameters. Then use different sizes of convolution kernels for convolution operations. Finally, merge all channel feature maps.

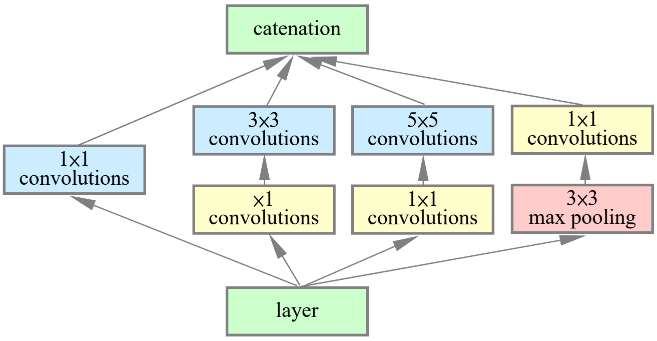


Fig. 2. Inception module with dimension reductions.

2) Two auxiliary classifiers are added to help training to prevent vanishing gradient in the network structure, as shown in Fig. 3. The auxiliary classifier performs a softmax operation on the output of the Inception (4a) and Inception (4d) modules, and then calculates the auxiliary loss. Finally, the weighted sum of the auxiliary loss and the real loss is equal to the total loss. In addition, the auxiliary loss is only used for training, not for testing.

3) The average pooling layer is used instead of the fully connected layer, which greatly reduces the amount of model parameters.

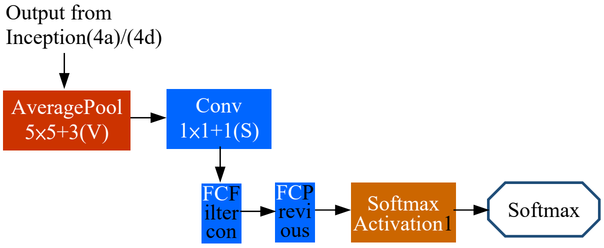


Fig. 3. The architecture of auxiliary classifier.

B. Activation Functions

The function of the activation function is to be able to add some non-linear factors to the neural network, so that the neural network can better solve more complex problems. Activation functions can be divided into two categories, saturated and non-saturated [12]. Saturated activation functions are commonly used in Sigmoid and Tanh, and non-saturated activation functions are commonly used in ReLU, Leaky ReLU, PReLU, RReLU, Maxout, ELU, Clipped ReLU etc. The advantage of saturation activation function is to squash the input between two fixed values, which means that it can be used as a good classifier. However, the saturated activation function has the disadvantage of vanishing gradient problem. Moreover, the advantage of the non-saturated activation function is that it can kill the gradient, thereby solving the vanishing gradient problem.

Sigmoid. The Sigmoid activation function [13] takes real values and squashes them to a range between 0 and 1. The Sigmoid is given by

$$f(x) = \frac{1}{(1 + exp^{-x})}. \tag{1}$$

ReLU. Rectified Linear Units (ReLU) [14] performs a threshold operation on each input value, where values less than zero are set to zero, and values greater than zero maintain the original input value, so the ReLU function is as follows:

$$f(x) = max(0, x) = \begin{cases} x_i, & if\ x_i \geq 0 \\ 0, & if\ x_i < 0 \end{cases}. \tag{2}$$

Leaky ReLU. ReLU sets all negative values to zero, whereas Leaky ReLU [15] assigns a fixed non-zero slope a (Normally, a=0.01) to all negative values. Thus the Leaky ReLU is computed as

$$f(x) = ax + x = \begin{cases} x, & if\ x > 0 \\ ax, & if\ x \leq 0 \end{cases}. \tag{3}$$

ELU. Exponential Linear Units (ELU) [16] has a small slope for negative values like the leaked ReLU. It uses curves instead of straight lines to smooth the gradient near the origin. They are both in functional form as follows:

$$f(x) = \begin{cases} x, & if\ x > 0 \\ a\ exp(x) - 1, & if\ x \leq 0 \end{cases}, \tag{4}$$

where a is a constant value initialized to 1. Another, the derivative of ELU equation is given by

$$f'(x) = \begin{cases} 1, & if\ x > 0 \\ f(x) + a, & if\ x \leq 0 \end{cases}. \tag{5}$$

Clipped ReLU. To prevent the output from becoming too large, Clipped ReLU [17] sets any input value less than zero to zero, and any value higher than the clipping ceiling (Normally, ceiling=10) is set to that clipping ceiling.

$$f(x) = \begin{cases} 0, & if\ x < 0 \\ x, & if\ 0 \leq x < ceiling\ . \\ ceiling, & if\ x \geq ceiling \end{cases} \tag{6}$$

In addition, researchers can still look for other better features to help neural networks converge faster or perform better.

C. Loss Function

The loss function can optimize the parameters of the neural network. Calculate the loss value between the predicted label and the real label through a neural network. Then, the gradient descent method is used to optimize the weights of the network until the loss is minimal. In this paper, the following cross-entropy equation is used to make convergence faster when approaching accurate classification [18].

$$\text{Cross-entropy}=-\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{k} t_{i,c}\log(p_{i,c}), \tag{7}$$

where

N: number of samples,

k: number of classes (k=2, drug and non-drug two classes),

log: natural logarithm,

$t_{i,c}$: a binary indicator (0 or 1), 1 if sample i is in class c and 0 otherwise,

$p_{i,c}$: the predicted probability that sample i is in class c.

The ideal cross entropy value is 0, but the experiment is set to be less than 10^{-4} or epoch up to 6 times to stop training.

III. Proposed Classification Framework

The task of this study is defined as the classification of face images of drug users and non-drug users. Therefore, we employ face detection, GoogLeNet, and transfer learning technology to establish a basic framework, as shown in Fig. 4. The details about important steps of the proposed framework are given in the following subsections.

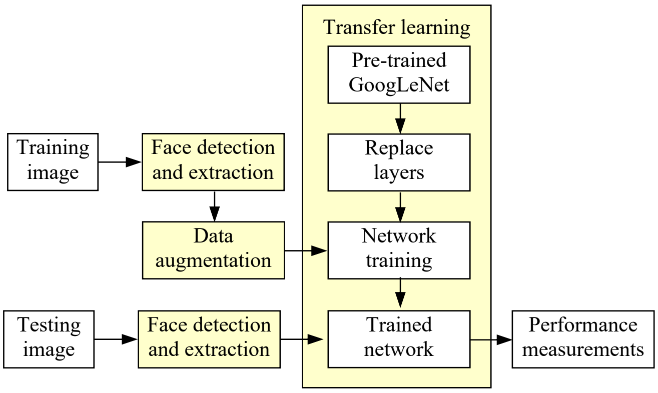


Fig. 4. The overall framework for the proposed classification system.

A. Face Detection And Extraction

The Viola-Jones algorithm [19] is an effective and fast face detection method. This research uses it to perform Haar feature selection, create integral images, Adaboost training, and cascade classifier steps, then extract frontal face images of drug users and non-drug users.

B. Data Augmentation

The number and imbalance of experimental samples are highly correlated with the accuracy of deep learning [20-21]. Therefore, in order to improve the accuracy, this study used data augmentation to increase the number of training samples. The data augmentation methods for face recognition are divided into two categories, generic and face specific transformation [22]. The former includes: (1) Geometric transformation: translation, rotation, scaling, mirroring, flipping, reflection, cropping, padding, perspective transformation; (2) Photometric Transformation: noise adding, contrast adjustment, color jittering, grayscaling, filtering, lighting perturbation, vignetting effect, random erasing. The latter includes: (1) Component transformation: hairstyle, makeup, accessory; (2) Attribute transformation: pose, expression, age.

In this work, the collected drug and non-drug face images conduct data augmentation of generic transformation, which commonly use include horizontal/vertical flip, translation, rotation, scale, brightness, blur, and shearing transformation to enlarge training image database. The seven transformations are described as follows:

Horizontal/Vertical flip: Mirror the image along the center vertical/horizontal line.

Translation: To move the image in the X or Y direction (or both), which can indicate everywhere in the image.

Rotation: Rotate the image at different angles clockwise or counterclockwise.

Scale: Resize the image outward or inward. When facing outward, the image size is enlarged, and when facing inward, the image size is reduced.

Brightness: Adjust the lightness and darkness of the image.

Blur: Using blurred images may result in higher resistance to motion blur during testing. Commonly used are Gaussian, average and median.

Shearing transformation: Moves each point horizontally or vertically by an amount proportional to its coordinates.

The drug and non-drug training datasets mentioned in the previous section will perform data augmentation. First, the two training datasets with a total of 1,470 face images will be flipped vertically, and 2,940 will be obtained. These images are then subjected to six types of data augmentation, including translation, rotation, scale, brightness, blur, and shearing transformation, and each type extends 4 images. Finally, a total of 70,560 face images in the drug and non-drug datasets were used in the experiment. Table II shows the appropriate parameter values for the six data augmentations.

TABLE II
LIST OF DATA AUGMENTATION PARAMETERS

Data augmentation	Parameter description	#obtained images
translation	10% displacement of image size in 4 directions up, down, left, and right	4
rotation	rotate 10, 20 degrees clockwise and counterclockwise, respectively	4
scale	scale factor 0.8, 0.9, 1.1, 1.2	4

brightness	brighten 10%, 20%, and darken 10%, 20%	4
blur	Gaussian filter with standard deviation σ = 0.5, 1; motion blur with angle $\theta=0^0$, 10^0 and corresponding length L=5 pixels	4
shearing transformation	horizontal and vertical shearing with shear factor m=-0.5 and 0.5, respectively	4

C. Transfer Learning

Transfer learning is a technique that uses an existing learning model to solve another new task. It only requires minimal retraining and avoids the complexity and time consuming of restarting the training model [23]. It is suitable for situations with insufficient training data or large-scale data collection and labeling. According to [11], it is pointed out that GoogLeNet [9] and ResNet [24] have better accuracy of face recognition, but the complexity of ResNet model is much higher than that of GoogLeNet. Based on this, we use GoogLeNet as the basis for transfer learning.

First load the Pre-trained GoogLeNet, and then we modify his last three layers, while the parameter values of other layers are fixed. As shown in Table III, the fully connected (FC) layer in the original GoogLeNet has been deleted. Instead, a new FC layer with two output sizes is inserted. The FC layer is followed by the softmax layer, and the last layer is replaced with two classes of output.

TABLE III
THE REPLACED LAYERS OF GOOGLENET BY TRANSFER LEARNING

No. of layers	Original	Replaced
142	Fully connected layer (1,000 nodes) with pre-trained weights and biases	Fully connected layer (2 nodes) with random initialization
143	Softmax layer	Softmax layer
144	Classification layer (1,000 classes)	Classification layer (2 classes: drug and non-drug)

IV. Experimental evaluation

A. Experimental Settings

Finally, the experiment was implemented in Matlab R2021a programming based on Windows 10 (x64) operating system, and was run on a PC-based machine with an Intel Core i5-6500 CPU, 3.2GHz, and 8G RAM.

B. Databases

We collected face images from the Internet and police agencies' databases to build two datasets of non-drug addiction and drug addiction. These two datasets are shown in Table IV, while detailed descriptions as below.

Drug Addiction Dataset

This database collects 520 images of drug addiction from the Internet, all of which are Western faces. Another 530 images of the police agency database, all of which are Oriental faces. A total of 1,050 of the above two, which include men and women different genders, as well as young and mature.

Non-drug Addiction Dataset

In order to increase the diversity of research data, such as different datasets, gender, age, and ethnicity. We collected three matching still image face datasets, including VGGFace2 [25], CAFR [26] and MORPH Album 2 [27]. VGGFace2 is a large-scale face recognition dataset with 3.31 million images of 9,131 people. The images were downloaded from Google Image Search, and meet the cross-age, different-racial face images required for this article. Cross-Age Face Recognition (CAFR) benchmark dataset, which mainly provides research on age-invariant face recognition. It contains 1,446,500 facial images from 25,000 subjects and is annotated with age, identity, gender, ethnicity and landmark labels. And, MORPH Album 2 contains 78,207 facial images of 20,569 subjects, spans 5 years, and ranges in age from 16 to 77.

Due to the large difference in the sample numbers of drug addiction and non-drug addiction databases, this binary classification of imbalanced data will cause distortion of the evaluation indicators [28]. Therefore, the experimental database randomly sampled a total of 1,050 images from VGGFace2, CAFR, and MORPH Album 2 with different people, and divided the gender, east and west, young and mature in half.

TABLE IV
AN OVERVIEW OF DRUG ADDICTION AND NON-DRUG ADDICTION DATASETS

Dataset	Source of collection	#images	#total	Description
Drug addiction	Internet	520	1050	Including Eastern and Western face images of different genders, as well as young and mature.
	Police agency	530		
Non-drug addiction	VGGFace2	-	1050	Face images were randomly selected from the three face databases, and divided into two halves by gender, eastern and western, young and mature.
	CAFR	-		
	MORPH Album 2	-		

C. Evaluation Indicators

We employ two common evaluation indicators accuracy and F1-score to evaluate the performance of our proposed model. Fig. 5 shows that the calculation method is based on the following four basic definitions of the confusion matrix:

TP: For a positive (P) sample, if the prediction is also true (T).

TN: For a negative (N) sample, if the prediction is also true (T).

FP: For a positive (P) sample, if the prediction is false (F).

FN: For a negative (N) sample, if the prediction is false (F).

		Predicted class	
		Positive	Negative
True class	Positive	TP (True positive)	FN (False negative)
	Negative	FP (False positive)	TN (True negative)

Fig. 5. The confusion matrix structure for binary classification.

Therefore, the accuracy is the ratio of the number of correct recognitions to the total number of samples.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \qquad (8)$$

In addition to using accuracy to evaluate overall effectiveness, we also use precision, recall, and F1-score. The precision is the ratio of correctly recognized positive values to the total number of recognized positive values, and the recall (same as sensitivity or True Positive Rate) is the ratio of correctly recognized positive values to the number of actual positive values, whereas F1-score might be a better measure to use if we need to seek a balance between precision and recall. The above three indexes can be described using the following equation:

$$\text{Precision} = \frac{TP}{TP + FP} \qquad (9)$$

$$\text{Recall} = \frac{TP}{TP + FN} \qquad (10)$$

$$\text{F1-score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \qquad (11)$$

D. *Validation*

The purpose of validation is to confirm the performance of the proposed model, and the number of samples determines the validation method. Since this study has used data augmentation to increase the number of face images in the database, the experiment will employ hold-out as a validation method. So 70% of the samples is used for training and the remaining 30% is used for testing.

E. *Experimental Results*

Performance Comparison of Different Face Recognition Models

To evaluate the effectiveness of the proposed method, we compared the performance of GoogLeNet with different classifiers based on transfer learning. In addition to the six well-known deep neural network classifiers, AlexNet, VGGNet16, VGGNet19, ResNet18, ResNet50, and ResNet101, six state-of-the-art neural network classifiers were also used, including Inception-v3 [29], Inception-ResNet-v2 [30], Xception [31], DenseNet201 [32], MobileNetV2 [33], and ShuffleNet [34]. In Table V, the numerical results of the model's verification performance led us to draw the following conclusions:

1) GoogLeNet showed the best accuracy (87.14%) among the 11 selected methods. The accuracy results indicated the low-dimensional discriminative features learned by the GoogLeNet model, suitable for the facial recognition of drug addicts. Subsequently, transfer learning was performed to obtain a more refined classification model, thereby obtaining better recognition performance.
2) The other three classification performances of the GoogLeNet model were also high, with the precision of 86.11%, the recall of 88.57%, and the F1-score of 87.32%. This demonstrated that GoogLeNet model showed the best overall classification performance among all selected models, considering the accuracy and the recall of the classification model. The final comparison is shown in Fig. 6.
3) Neural network deepening can also improve the model performance. Since deeper models have better nonlinear expression capabilities, they can learn more complex transformations to fit intricate feature inputs. However, if a neural network is too deep, it may overfit and reduce a model's performance, such as Inception-ResNet-v2, DenseNet201 model.

TABLE V
PERFORMANCE OF FACE RECOGNITION FOR DRUG ADDICTION USERS AND NON-DRUG USERS

Model	#Layers	Measure (%)			
		Accuracy	Precision	Recall	F1-Score
AlexNet	8	74.42	72.41	60.00	65.63
VGGNet16	16	74.42	69.70	65.71	67.65
VGGNet19	19	73.26	80.00	45.71	58.18
ResNet18	18	80.23	78.13	71.43	74.63
ResNet50	50	77.91	71.05	77.14	73.97
ResNet101	101	83.72	81.82	77.14	79.41
GoogLeNet	22	**87.14**	**86.11**	**88.57**	**87.32**
Inception-v3	48	76.74	72.73	68.57	70.59
Inception-ResNet-v2	164	70.93	64.71	62.86	63.77
Xception	71	75.58	69.44	71.43	70.42
DenseNet201	201	79.07	79.31	65.71	71.88
MobileNetV2	28	77.91	76.67	65.71	70.77
ShuffleNet	50	74.42	65.85	77.14	71.05

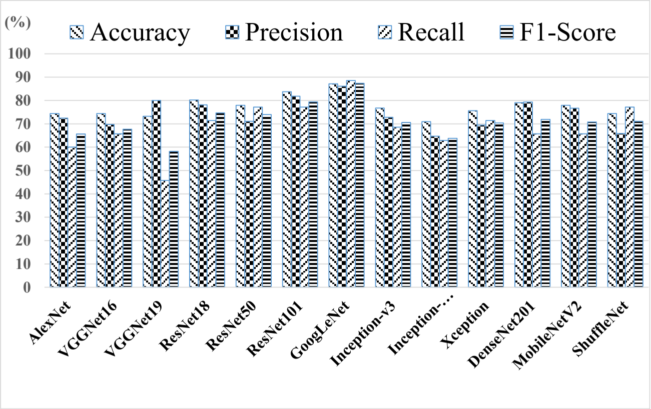
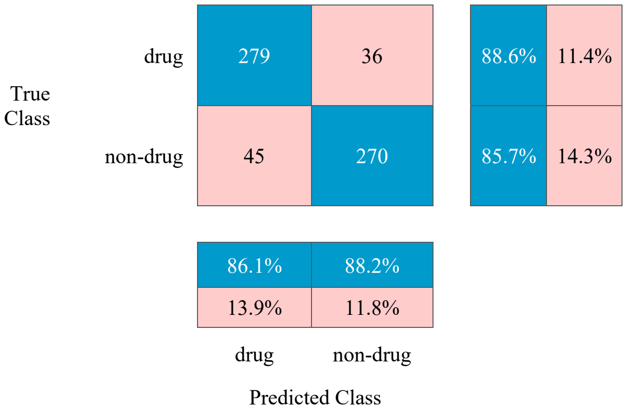


Fig. 6. Comparison results.

In this study, a total of 630 face images were used for testing, and the confusion matrix obtained for the test face images is shown in Fig. 7.



True Class

drug	279	36	88.6%	11.4%
non-drug	45	270	85.7%	14.3%
	86.1%	88.2%		
	13.9%	11.8%		
	drug	non-drug		

Predicted Class

Fig. 7. Confusion matrix of GoogLeNet on the Drug Addiction Dataset.

As shown in the first column of the confusion matrix, the GoogLeNet with transfer learning model classifies 279 of the 324 drug addiction users images as True (True Positive, TP), and 45 images as False (False Positive, FP). In second column, the model classifies 270 images from non-drug users' pictures as True Negative (TN), and classifies 36 as false (False Negative, FN).

In addition, three better performance ROC (Receiver Operating Characteristic) curves are shown in Fig. 8, the vertical axis is True Positive Rate (=TP/(TP+FN)), the horizontal axis is False Positive Rate (=FP/(FP+TN)), i.e. the proportion of incorrect predictions in positive class), and the curve appears to be closer to the upper left. And the AUC (Area Under the Curve) value is the highest at 0.87, which means that the GoogLeNet has good discrimination for face recognition of drug addicts.

At the same time, we also observed testing samples with incorrect judgments and found that most of these testing samples belonged to non-methamphetamine addicts. It means that the long-term abuse of different drugs has different effects on the appearance of the face. Methamphetamine drugs have the most serious damage to the appearance of drug users.

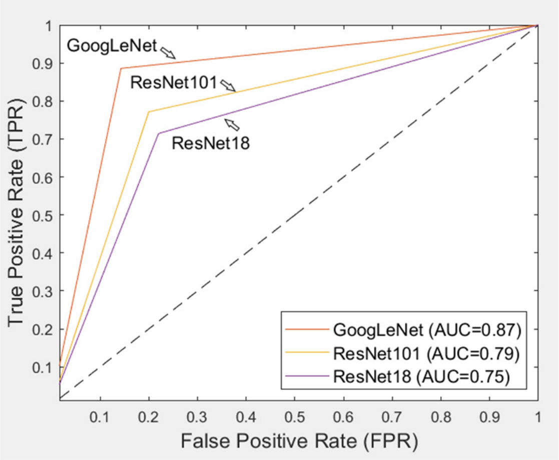
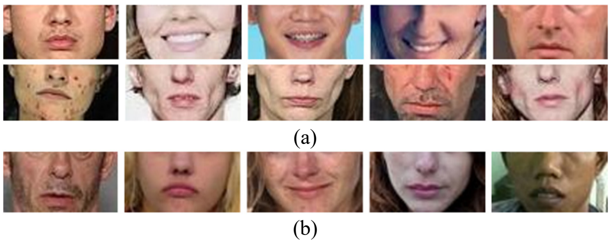


Fig. 8. ROC curve of the top 3 highest AUC values on the Drug Addiction Dataset.

Visualization of the Classification Results

The classification results of the best GoogLeNet are visualized (see Fig. 9). Fig. 9(a) Row 1 shows that the samples were actually non-drug users, were predicted to be non-drug users, and that all test samples of non-drug users were correctly classified. The samples in Row 2 are actually drug addicts and are predicted to be drug addicts as well. Fig. 9(b) also shows some misclassified examples, which are actually not easily judged by experts.


(a)

(b)

Fig. 9. Examples of classification results: (a) Correct classification, (b) Misclassification.

The reason for this is that the drug users in the photos are not long-term drug users, so there are no obvious wrinkles, loss of elasticity, sunken cheeks, redness, rashes in appearance. Or someone who isn't a drug addict but is misjudged because of facial scars, pimples, dirt.

Performance Comparison of Different Activation Functions

The actiation function can introduce nonlinearity into the deep neural network model, and will significantly affect the performance of the model. However, saturated activation functions such as Sigmoid and Tanh have the vanishing gradient problem, so they have been gradually abandoned. Currently, most studies tend to use non-saturated activation functions. This work uses the GoogLeNet model with the best classification performance as the experimental object. The default activation function of the GoogLeNet model is ReLU, and then it is replaced by LeakyReLU, ELU, and Clipped ReLU to test separately. Table VI shows the experimental results of the above four activation functions, including accuracy and macro-average F1-Score. Among them, the ReLU activation function still performed best, with accuracy of 87.14% and F1-score of 87.32%. It is undeniable that although Leaky ReLU and Clipped ReLU can manually adjust parameter values to achieve better results, their performance still cannot surpass ReLU.

TABLE VI
CLASSIFICATION RESULTS OF GOOGLENET WITH DIFFERENT ACTIVATION FUNCTIONS.

Activation	Parameter	Accuracy (%)	F1-Sore (%)
ReLU	-	87.14	87.32
Leaky ReLU	*a*=0.1	77.91	76.54
	a=0.01	80.23	76.06
	a=0.05	80.23	78.48
ELU	-	76.74	71.43
Clipped ReLU	*ceiling*=10	76.74	71.43
	ceiling=15	74.42	71.79
	ceiling=20	80.23	77.92

Fine-tune the Probability of Dropout to Improve Performance

Dropout was first proposed in the AlexNet neural network. It randomly turns off (does not update the weights) some hidden neuron connections at the specified ratio, which will effectively reduce the overfitting problem [35]. GoogLeNet also introduces a dropout layer before the fully connected layer, and turns off 40% of hidden neuron connections by default to improve performance. Therefore, this study fine-tuned the dropout probability and found that the better accuracy is at a probability of 0.45, which is a slight increase of 0.88% compared to the default dropout probability of 0.4, as shown in Table VII.

TABLE VII
CLASSIFICATION PERFORMANCE OF GOOGLENET WITH DIFFERENT DROPOUT PROBABILITY.

	Without dropout	Dropout probability			
		0.3	0.35	0.4	0.45
Accuracy (%)	79.07	75.58	80.23	86.05	**87.91**
		0.5	0.55	0.6	
		81.40	79.07	77.91	

V. Conclusion

Deep convolutional neural networks have excellent performance in solving image classification and recognition problems. In this paper, we combine face detection, GoogLeNet, data augmentation, and transfer learning technology to establish a special facial image recognition framework that can differentiate drug users from non-drug users to expand the applications of CCTV in regulating the overall drug situation. The results show that the model in this study has a high classification accuracy of 87.14%, and its overall classification performance reached 87.32%. In addition, adjusting the activation function and the dropout probability can further optimize the model's performance. In the next stage of research, our work will focus on fusing the facial blob features of drug addicts into this model to further improve its recognition accuracy.

References

[1] M. Jacquet, C. Champod, "Automated face recognition in forensic science: Review and perspectives," Forensic Science International, vol. 307, Feb. 2020.

[2] Y. Jang, D. Kim, J. Park, D. Kim, "Conditional effects of open-street closed-circuit television (CCTV) on crime: A case from Korea," International Journal of Law, Crime and Justice, vol. 53, Jun. 2018, pp. 9-24.

[3] J. L. S. Gonzáleza, C. Zaccaroa, J. A. Álvarez-García, et al., "Real-time gun detection in CCTV: An open problem," Neural Networks, vol. 132, Dec. 2020, pp. 297-308.

[4] National Rehabs Directory, "Before and After Alcohol and Drug Addiction Photos," 2014. https://www.rehabs.com/explore/faces-of-drug-arrests/ [last accessed 14.02.2021]

[5] ImageNet Large Scale Visual Recognition Challenge (ILSVRC), http://image-net.org/challenges/LSVRC/ (accessed Sep. 2020)

[6] A. Krizhevsky, I. Sutskever, G. Hinton, "ImageNet classification with deep convolutional neural networks," International Conference on Neural Information Processing Systems, vol. 1, Dec. 2012, pp. 1097-1105.

[7] M. D. Zeiler, R. Fergus, "Visualizing and understanding convolutional networks," European Conference on Computer Vision, 2014, pp. 818-833.

[8] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015 International Conference on Learning Representations (ICRL), 2015, pp. 1-14.

[9] C. Szegedy, W. Liu, Y. Jia, et al., "Going deeper with convolutions," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, USA, 7-12 June, 2015, pp. 1-9.

[10] J. Hu, L. Shen, G. Sun, "Squeeze-and-excitation networks," 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 7132–7141.

[11] Q. Wang, G. Guo, "Benchmarking deep learning techniques for face recognition," Journal of Visual Communication and Image Representation, vol. 65, Dec. 2019.

[12] S. Qian, H. Liu, C. Liu, S. Wu, H. S. Wong, "Adaptive activation functions in convolutional neural networks," Neurocomputing, vol. 272, Jan. 2018, pp. 204-212.

[13] J. Han, C. Morag, "The influence of the sigmoid function parameters on the speed of backpropagation learning," IWANN '96: Proceedings of the International Workshop on Artificial Neural Networks: From Natural to Artificial Neural Computation, vol. 930, 1995, pp. 195–201.

[14] V. Nair, G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," International Conference on International Conference on Machine Learning, Jun. 2010, pp. 807-814.

[15] A. L. Maas, A. Y. Hannun, A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models", ICML Workshop on Deep Learning for Audio Speech and Language Processing (WDLASL 2013), vol. 28, 2013.

[16] D.-A. Clevert, T. Unterthiner, S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUs)," arXiv, arXiv:1511.07289, 2015.

[17] Z. Cai, X. He, J. Sun, N. Vasconcelos, "Deep learning with low precision by half-wave Gaussian quantization," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 5406–5414.

[18] P. Golik, P. Doetsch, H. Ney, "Cross-entropy vs. squared error training: A theoretical and experimental comparison," 14th Annual Conference of the International Speech Communication Association, Lyon, France, Aug. 2013, pp. 1756-1760.

[19] P. Viola, M. Jones, "Rapid object detection using a boosted cascade of simple features," Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), USA, 2001, pp. 511-518.

[20] M. Kubat, S. Matwin, "Addressing the curse of imbalanced training sets: One sided selection," In Proceedings of the Fourteenth International Conference on Machine Learning, 1997, pp. 179-186.

[21] Md. Y. Arafat, S. Hoque, S. Xu, D. M. Farid, "Machine learning for mining imbalanced data," IAENG International Journal of Computer Science, vol. 46, issue 2, 2019, pp332-348.

[22] X. Wang, K. Wang, S. Lian, "A survey on face data augmentation," arXiv, arXiv:1904.11685, Apr. 2019.

[23] S. J. Pan, Q. Yang, "A survey on transfer learning," IEEE Transactions on Knowledge and Data Engineering, vol. 22, issue 10, 2009, pp. 1345-1359.

[24] K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2016.

[25] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, A. Zisserman, "VGGFace2: A dataset for recognising face across pose and age," International Conference on Automatic Face and Gesture Recognition, arXiv, arXiv:1710.08092v2, May 2018.

[26] J. Zhao, Y. Cheng, Y. Cheng, Y. Yang, H. Lan, F. Zhao et al., "Look across elapse: Disentangled representation learning and photorealistic cross-age face synthesis for age-invariant face recognition," arXiv, arXiv:1809.00338, Oct. 2018.

[27] K. Ricanek, T. Tesafaye, "Morph: A longitudinal image database of normal adult age-progression," In: Automatic Face and Gesture Recognition, Intl. Conf. on. IEEE, 2006, pp. 341-345.

[28] S. S. Mullick, S. Datta, S. G. Dhekane, S. Das, "Appropriateness of performance indices for imbalanced data classification: An analysis," Pattern Recognition, Vol. 102, Jun. 2020.

[29] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, "Rethinking the inception architecture for computer vision," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jul. 2016, pp. 2818-2826.

[30] C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi, "Inception-V4 inception-resnet and the impact of residual connections on learning," Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17), Feb. 2017.

[31] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Nov. 2017, pp. 1251-1258.

[32] G. Huang, Z. Liu, L.V.D. Maaten, K. Weinberger, "Densely connected convolutional networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Nov. 2017, pp. 2261-2269.

[33] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2018, pp. 4510-4520.

[34] X. Zhang, X. Zhou, M. Lin, J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2018, pp. 6848-6856.

[35] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," The Journal of Machine Learning Research, Vol. 15, No. 1, 2014, pp. 1929-1958.

Chi-Hung Chuang is an Associate Professor, Department of Information and Computer Engineering, Chung Yuan Christian University. Received his Ph.D. degree from Department of Computer Science and Information Engineering, National Central University, Taiwan, in 2009. His research interest includes Artificial Intelligence & Pattern Recognition, Image Processing, Computer Vision Virtual Reality and Augmented Reality.

Cheng-Tan Tung is an Assistant Professor, Department of Information Management, Central Police University. Received his Master degree from Department of Computer Science and Information Engineering, National Central University, Taiwan. His research interest includes Digital Forensics, Data Visualization, and Machine Learning.

Yuan-Song Chang is an Assistant Professor, Department of Foreign Affairs Police, Central Police University. Received his Ph.D. degree from Department of Administration Police, Central Police University, Taiwan, in 2013. His research interest includes Cross-Domain Drug Law Enforcement and Governance, Research Methods, Human Trafficking Research, Policing Policy and Practice.

Edward Lin is a graduate student at the Department of Electrical Engineering, Stanford University. Received his BA degree from the Department of Electrical Engineering and Computer Science, UC Berkeley, USA, in 2022. His research includes Optimization, Artificial Intelligence, Machine Learning, Embedded Systems, and Network Reliability.

Chih-Ping Yen is an Associate Professor, Department of Information Management, Central Police University. Received his Ph.D. degree from Department of Computer Science and Information Engineering, National Central University, Taiwan, in 2014. His research interest includes Artificial Intelligence & Pattern Recognition, Image Processing, Management Science, Inventory Problem, and Management Information Systems.