YCbCr Color Space as an Effective Solution to the Problem of Low Emotion Recognition Rate of Facial Expressions In-The-Wild

Hadjer Boughanem[1], Haythem Ghazouani[1,2] and Walid Barhoumi[1,2]
[1]*Université de Tunis El Manar, Institut Supérieur d'Informatique d'El Manar, Research Team on Intelligent Systems in Imaging and Artificial Vision (SIIVA), LR16ES06 Laboratoire de recherche en Informatique, Modélisation et Traitement de l'Information et de la Connaissance (LIMTIC), 2 Rue Abou Rayhane Bayrouni, 2080 Ariana, Tunisia*
[2]*Université de Carthage, Ecole Nationale d'Ingénieurs de Carthage, 45 Rue des Entrepreneurs, 2035 Tunis-Carthage, Tunisia*

Keywords: In-The-Wild FER, Deep Features, YCbCr Color Space, CNN, Features Extraction, Deep Learning.

Abstract: Facial expressions are natural and universal reactions for persons facing any situation, while being extremely associated with human intentions and emotional states. In this framework, Facial Emotion Recognition (FER) aims to analyze and classify a given facial image into one of several emotion states. With the recent progress in computer vision, machine learning and deep learning techniques, it is possible to effectively recognize emotions from facial images. Nevertheless, FER in a wild situation is still a challenging task due to several circumstances and various challenging factors such as heterogeneous head poses, head motion, movement blur, age, gender, occlusions, skin color, and lighting condition changes. In this work, we propose a deep learning-based facial expression recognition method, using the complementarity between deep features extracted from three pre-trained convolutional neural networks. The proposed method focuses on the quality of features offered by the YCbCr color space and demonstrates that using this color space permits to enhance the emotion recognition accuracy when dealing with images taken under challenging conditions. The obtained results, on the SFEW_2.0 dataset captured in wild environment as well as on two other facial expression benchmark which are the CK+ and the JAFFE datasets, show better performance compared to state-of-the-art methods.

1 INTRODUCTION

Nowadays, along with the excess in computer performance and the anticipation increase of human computer interaction, FER has attracted rising attention from researchers in different fields. In addition to the FER studies in computer science field (Ghazouani, 2021), (Bejaoui et al., 2019), (Sidhom et al., 2023), (Bejaoui et al., 2017) the emotion recognition is present in psychology (Banskota et al., 2022), neuroscience (Yamada et al., 2022) and other related disciplines. Despite the numerous studies in the FER, recognizing an emotion in uncontrolled circumstances remains a real challenge. The complexity of backgrounds and other circumstances in real-world conditions hinders the correct detection of faces from the backgrounds and subsequently affects the emotion recognition rate. However, regardless of the conditions in which facial expressions images have been taken, the process leading to recognizing emotions is the same. A typical FER system is mainly composed of three core steps, starting with face detection, then features extraction and finishing by emotion classification (Boughanem et al., 2021). Accurate results of face detection enable features extraction to be performed on well-focused image regions and certainly to have a high recognition rate. Several methods were proposed for face detection. We denote methods that use classic machine learning techniques (Hu et al., 2022), CNNs (Billah et al., 2022), classification techniques (Hosgurmath et al., 2022) and those that use skin color detection using different color spaces (Khanam et al., 2022), (Ittahir et al., 2022). This factor plays an integral role to separate the skin parts from the non-skin ones and provides an important cue for face detection. Several color spaces have been investigated, and the most used ones for are RGB, HSV and YCbCr (Al-Tairi et al., 2014), (Rahman et al., 2014). Among these spaces, the YCbCr is the most recommended when dealing with facial images. In fact, the skin color range is well defined in this space (Terrillon et al., 2000), (Yan et al., 2021).

Boughanem, H., Ghazouani, H. and Barhoumi, W.
YCbCr Color Space as an Effective Solution to the Problem of Low Emotion Recognition Rate of Facial Expressions In-The-Wild.
DOI: 10.5220/0011795300003417
In *Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2023) - Volume 5: VISAPP*, pages 822-829
ISBN: 978-989-758-634-7; ISSN: 2184-4321

Recent FER methods include CNN and Deep Learning (DL) techniques for feature extraction and emotion classification. They are widely used due to their satisfactory results obtained even when dealing with resolution issues. DL methods and CNN models are used with different space colors. The YCbCr is suitable for image classification applications where the lightness conditions change drastically and especially for applications involving skin color. That is because, the YCbCr color space does not contain the effects of light which can change the characteristics of the skin color. Therefore, many feature information can be obtained robustly even in-the-wild conditions. This motivated us to use CNN extraction methods along with the YCbCr space for FER in-the-wild. In fact, in order to improve the emotion recognition rate of wild facial expressions, we deal in this work with facial images converted into the YCbCr space color. Indeed, deep features are extracted from three pre-trained CNNs. Then, they are combined to be fed to a Support Vector Machine (SVM) in order to classify the facial features into one emotional class.

The remainder of this paper is organized as follows. Section 2 details the related works dealing with in-the-wild expressions and especially those using the YCbCr color space. In section 3, a description of the proposed method is given. In section 4, we discuss the experimental results. Finally, section 5 summarizes the proposed method and highlights future scope.

2 RELATED WORK

Face detection and feature extraction in different backgrounds are technically difficult, especially when dealing with complex backgrounds of the unconstrained environments. The major challenge in face detection is to cope with different variations in the human face caused by several factors such as face orientation, face size, facial expression, people ethnicity, age and lighting changes. Therefore the face detection step is a crucial step, because it determines the quality of the features extracted and then classified to recognize the emotions. Several techniques change the default space color into YCbCr space, to detect the faces based on the skin color region which are easier to distinguish from the non-skin parts in this color space. In (Nugroho et al., 2021), the highest accuracy for face detection is obtained in YCbCr color space reaching 96.13%. Indeed, the authors used a segmentation step with thresholding and morphological operation. The authors in (Yan et al., 2021) deal also with images in YCbCr color space. They used the elliptic skin color model and logistic regression analysis to determine the skin color probability while using a genetic algorithm to segment the face region. The obtained results show an improvement of face detection and a good robustness to posture and expression changes. The work of (Li, 2022), proposed a method based on skin color segmentation, particle swarm search and curve approximation aiming to improve the accuracy of expression recognition in facial images converted into YCbCr color space. The results show that the method can eliminate the interference factor and improve the facial recognition rate. (Ahmady et al., 2022), used two different types of features, including fuzzified Pseudo Zernike Moments features and structural features like teeth existence, eye and mouth-opening, and eyebrow constriction). The feature extraction was based on images converted into YCbCr color space to localize facial components. The experimental results of this method demonstrate the robustness of the method in terms of age, ethnicity, and gender changes, as well as to increase the recognition rate of facial expression. The research of (Vansh et al., 2020) improved the face detection using the YCbCr space and Adaboost. It involves pre-processing of input images to extract skin tone in YCbCr color space, followed by face detection using Haar cascade classifiers. The approach in this paper provides the ability to detect the occluded faces or side faces in the input image. The test results in (Putra et al., 2020) illustrate that the YCbCr color space has obtained maximum accuracy when recognizing skin diseases among all color spaces. Results obtained by the aforementioned image processing applications involving skin color are promising. This motivated us to use the YCbCr color space in order to deal with issues of FER in-the-wild environment. Subsequently, deep relevant facial features extracted using fine-tuned CNN architetures from images converted to the YCbCr color space are fed to an SVM classifier to recognize facial emotions in unconstrained environments. For the purpose of fulfilling the need to deal with FER in-the-wild in many applications, the suggested method proposes an enhanced deep learning-based method to recognize spontaneous emotions captured in unconstrained environments. The method relies on the complementarity between the deep features extracted from different CNN models (Boughanem et al., 2022).

3 PROPOSED METHOD

The proposed method is structured on three main components: Pre-processing and face detection, feature extraction and selection and emotion classification. A flowchart of the proposed method is provided

in the Figure 1. It is based on the deep feature extraction from facial expression images converted into YCbCr color space. It uses three pre-trained models. The features are extracted from each model separately. The most relevant ones are then selected and concatenated into one final feature vector. The feature selection mechanism used in this work ensures the quality of the final feature vector. Moreover, the complementarity of deep features extracted in the YCbCr space and selected from specific layers, ensures the enhancement of the overall emotion recognition rate.

3.1 Image Pre-Processing and Face Detection

Image pre-processing is the first step in the FER. The quality of input images and facial features selection are critical to obtain good classification results. However, challenging environments and bad acquisition conditions can lead to poor quality images. Additionally, movement, noise, luminosity, face orientation, and face position offset can make the feature extraction a complicated step (Deng et al., 2021). Moreover, the presence of complex background or extra facial features such as glasses, beard and moustache can increase significantly the FER task. Consequently, pre-processing is an essential step to deal with noise caused by image acquisition and digitization. In this step, the input facial images are aligned and normalized to shorten the neural network learning time and to obtain a better inference generalization in order to ensure lighting change robustness. Subsequently, the input images are converted to YCbCr color space as illustrated in Figure 2.

The images in YCbCr space are stored as three dimensional matrix, according to the three components Y, Cb and Cr. Finally, in order to keep only useful regions and simultaneously to eliminate the maximum of the non-facial parts, the image have been cropped by detecting face over the entire image. In this work, the simple and robust face detection algorithm of Viola & Jones (Viola and Jones, 2001) is applied.

YCbCr Color Space: YCbCr is the standard for digital television and image compression, where Y represents the luminance component (luma) which is more sensitive to the human eye, whereas, the Cb and Cr represent the chrominance component (chroma), which refer to the blue and the red color respectively (Rahman et al., 2014). The Luma component is calculated by a weighted sum of the components of Red, Green and Blue as indicated in (1).

$$Y = 0.299 \times Red + 0.587 \times Green + 0.114 \times Blue \quad (1)$$

The chroma components are calculated from the Luma as illustrated respectively in (2) and (3) (Khanam et al., 2022):

$$Cb = Blue - Y \quad (2)$$

$$Cb = Red - Y \quad (3)$$

The difference between YCbCr and RGB is that RGB represents colors as combinations of red, green and blue signals, while YCbCr represents colors as combinations of a brightness signal and two chroma signals. In YCbCr, Y is luma (brightness), Cb is blue minus luma (B-Y) and Cr is red minus luma (R-Y). The luma channel, typically denoted Y approximates the monochrome picture content. The two chroma channels, Cb and Cr, are color difference channels.

After applying the face detection on images transformed in YCbCr space, we perform data augmentation (DA) to feed sufficient training images to finetune the CNN models. Indeed, DL based FER methods are mostly driven by the availability of large samples of training data. It is not always possible, even unfeasible, to obtain enough training samples, furthermore sufficient samples for each category of emotion, especially when concerning facial images in-the-wild. In order to tackle this issue, geometric DA techniques are applied to generate sufficient number of training samples. We have applied four geometric DA techniques to generate new training images from each cropped image, which are: horizontal and vertical translations, horizontal reflection and random image rotations with a rotation angle within [-10°, 10°].

3.2 Feature Extraction and Selection

Once the face detection is completed, the images are resized into $224 \times 224 \times 3$. Then, the facial expression information is extracted from the facial images in YCbCr color space, using the feature extraction methods. After that, the emotions are classified according to the extracted features. Wherefore, facial feature extraction is considered the key step in FER process. It determines the final emotion recognition result and also affects the recognition rate. In this stage, we implement three well-known powerful pre-trained CNN models. The choice of the ResNet101, VGG19, and GoogleNet models were argued in our previous published work (Boughanem et al., 2022). Moreover, these models have been applied on datasets taken in controlled environments in several works (Siam et al., 2022) (Saurav et al., 2022), and similarly, they proved their effectiveness. CNNs are the most popular path of processing and analyzing images. Their hidden layers called convolutional layers are exploited to extract valuable deep features.

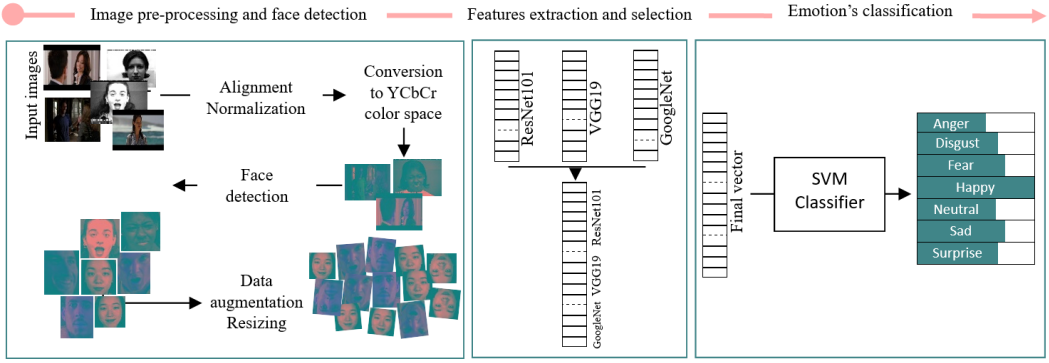

Figure 1: Proposed method's layout.

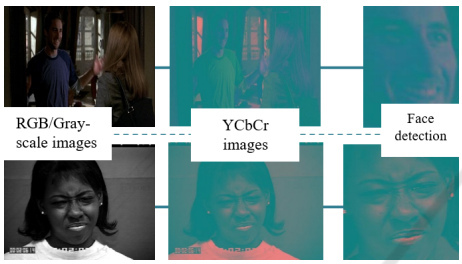

Figure 2: Color space conversion and face detection.

The ResNet101 (He et al., 2016), VGG19 (Simonyan and Zisserman, 2014) and GoogleNet (Szegedy et al., 2015) models are used for deep feature extraction. Thereafter, the selected features from the three pre-trained models are combined to form one feature vector containing all the facial expression features.
In order to extract the foremost facial features, for a first step, we perform a transfer learning on the three neural networks. The parameters of the transfer learning have been fixed according to the declared parameters in (Boughanem et al., 2022) paper. We use a learning rate of 1.e-4 and we optimize the GoogleNet and the ResNet101 models by ADAM optimizer, however we use the SIGMOID optimizer for the VGG19 model. The results obtained on the three datasets are shown in the following table (Table 1).

Table 1: Transfer learning results on three datasets.

	JAFFE	CK+	SFEW_2.0
ResNet101	90.48%	91.04%	59.69%
GoogleNet	92.86%	89.90%	54.53%
VGG19	90.48%	87.62%	54.90%

The second part examines the facial feature extraction from each model used models to be combined in one facial feature vector. The final feature vector is fed to an SVM classifier in order to determine the emotional state of the input face. The most suitable features are selected from top block layers of each used DL model. We retained the two combinations that gave the highest recognition rates for in-the-wild environments. The first combination is composed of two pooling layers and one fully connected layer. The second one is composed of two fully connected layers and one pooling layer. These two combinations have been tested on the YCbCr datasets and gave better recognition rates compared to those obtained on the RGB color space when applied on the three datasets, while outperforming also those of relevant state-of-the-art methods. The results, using the retained combinations, on the three datasets are listed in Table 2.

Table 2: Recognition rates using the YCbCr color space.

	SFEW_2.0	JAFFE	CK+
1st combination	**92.28%**	**100%**	98.90%
2nd combination	91.46%	**100%**	**99.17%**

3.3 Emotion Classification

The performance of the emotion classification is closely related to the pre-treatment step. The output of the feature extraction step, is a single feature vector gathering relevant facial features from three pre-trained neural networks. A supervised SVM classifier is trained to classify the extracted feature into right emotion categories. The test images are different of the training ones. Their number is reduced compared to the training images, since the test images enfold only 20% of the total of each dataset.

4 EXPERIMENTS AND DISCUSSION

In this section, the datasets used in this work are firstly described. Then we present extensive quantitative re-

sults and comparison between the proposed method and the existing works. Finally, we analyze and discuss the results.

4.1 Datasets

In this work, we deal with spontaneous emotions as well as posed ones in two different environmental conditions. The focus was on emotions in-the-wild environments, considering the complexity of their context which is the closest to reality. In order to ensure the effectiveness of the proposed method, we used two other datasets conceived under controlled laboratory conditions for the experimental results. We conduct experiments on three FER datasets (Table 3), namely SFEW_2.0 (Dhall et al., 2012), CK+ (Kanade et al., 2000) and JAFFE (Lyons et al.,) .

- **The Static Facial Expressions in the Wild (SFEW_2.0):** It is a static version (Dhall et al., 2014) collected by extracting images from the videos of the Acted Facial Expressions in the Wild (AFEW) dataset. This version of SFEW dataset was updated in 2018. It is composed of three sets, the training set contains 958 images, the validation one contains 436 and the test sets includes 372 images. All the sets are distributed into seven classes of emotion (Angry, Disgust, Fear, Happy, Neutral, Sad, Surprise).
- **The Extended Cohn-Kanade Dataset (CK+):** The CK+ is an extended version of the CK dataset. It is partitioned into six basic emotions (Anger, Disgust, Fear, Happiness, Sadness, Surprise) and a "Contempt" emotion, containing posed and spontaneous emotions. The dataset is conceived in constrain laboratory conditions. It is comprised of male and female subjects belonging to different ethnic groups (Lucey et al., 2010).
- **The Japanese Female Facial Expression Dataset (JAFFE):** This dataset is also conceived in laboratory-controlled conditions. It contains 213 facial expression images of 10 Japanese women. The dataset is composed only of posed emotions. The facial expression images are in grayscale sized 256×256 pixels, encompassing the seven universal emotions.

Table 3: Datasets samples distribution.

	Training set (80%)	Test set (20%)
SFEW_2.0	984	236
CK+	4331	1083
JAFFE	170	43

4.2 Facial Emotion Recognition Results

The findings of each method step have been illustrated in Tables 1 and 2. The first step results are three feature vectors corresponding to the three CNN models. After the combination, we obtain one feature vector containing all facial features selected from each model. We summarize the results of the two steps in Table 4. We report the confusion matrices corresponding to each dataset in the Figures 3,4 and 5.

Table 4: Facial emotion recognition results.

	JAFFE	CK+	SFEW_2.0
ResNet101	90.48%	91.04%	59.69%
GoogleNet	92.86%	89.90%	54.53%
VGG19	90.48%	87.62%	54.90%
Overall accuracy	**100%**	**99.17%**	**92.28%**

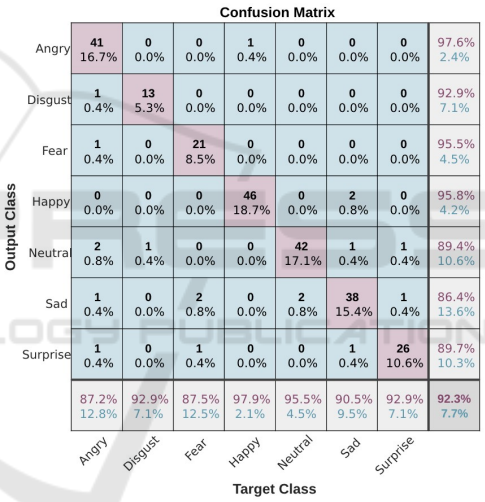

Figure 3: Confusion matrix of the proposed method on the SFEW_2.0 dataset.

4.3 Discussion

The experiments on the in-the-wild dataset (SFEW_2.0) have shown very satisfactory results. The overall recognition rate after combining different features selected from the three neural networks reached 92.3%. The recognition rates obtained by the three CNNs individually for this in-the-wild dataset are: 59.69%, 54.90% and 54.53% for ResNet101, VGG19 and GoogleNet, respectively. The three recognition rates are close to each other. However, the facial features conveyed from the three models are complementary. This fact explains the higher overall recognition rate reached after feature combination. For the second dataset CK+ containing spontaneous

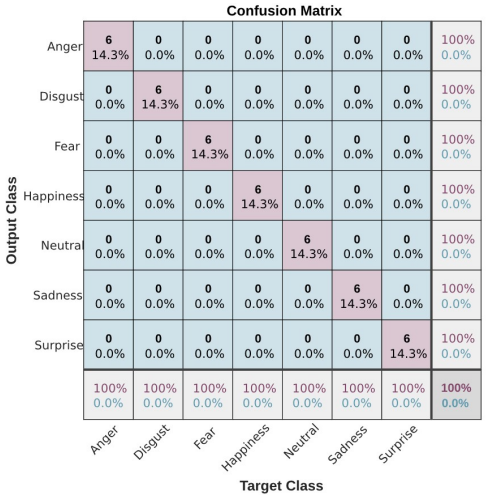

Figure 4: Confusion matrix of the proposed method on the JAFFE dataset.

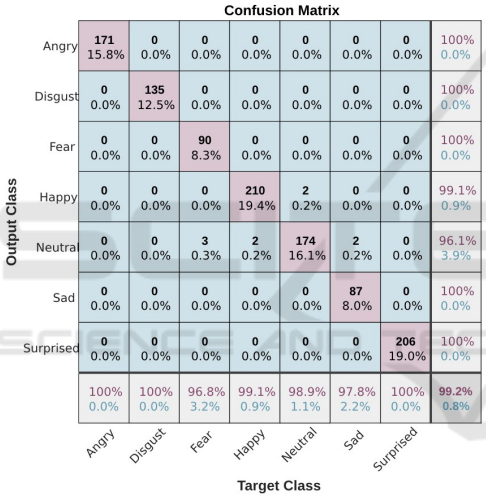

Figure 5: Confusion matrix of the proposed method on the CK+ dataset.

and posed emotions in laboratory conditions, the final recognition rate reaches 99.2% which is as well a high rate. It was achieved by combining the selected deep and relevant facial features extracted from different CNN layers. It is noteworthy that the misclassified images are only nine images. The majority of these images are confused with neutral emotion. This is because the "Neutral" class does not exist in this dataset and was designed manually by collecting the three first sequences of each person's facial expressions from the six emotions. Regarding the JAFFE dataset, all the emotions have been well recognized. The obtained recognition rate of 100% is a proof of the complementarity between facial features assembled from the facial features of each pre-trained model. Comparing the results of the two tested layers' combinations presented in Table 2, we can notice that the two combinations provide similar values for the CK+ and the SFEW_2.0 datasets. A certain margin of difference of 0.27% and 0.82% have been scored respectively. Similarly, the two combinations impart identical values for the case of the JAFFE dataset. The highest ranked combination (two pooling layers and one fully connected layer) for wild environments in the benchmark work, remains in the first position for this dataset. The conversion to the YCbCr color space brought more relevant facial features leading to improve the overall recognition rate. In the case of the CK+ dataset, the highest recognition rate was obtained by applying the second combination (two fully connected layers and one pooling layer) with a minimum percentage gap of 0.27%. In Table 5 and Table 6, we evaluated the efficiency of the proposed method by comparing its results with some relevant state-of-the-art methods, including the work of (Boughanem et al., 2022) using RGB color space. Table 5 presents an expanded comparison on the SFEW_2.0 dataset. The outcomes of the proposed method applied using the YCbCr color space outperform all the state-of-the-art methods, even the work of (Boughanem et al., 2022) which deals with the same problems and datasets while using the RGB color space. The recognition rates obtained on the YCbCr space reached an increase of 4.1% compared to (Boughanem et al., 2022) and 29.4% compared to the second best recognition rate (Cai et al., 2022) cited in the table. With regard to the two controlled-laboratory conditions datasets (Table 6), the recognition rates obtained using the original color space (RGB or Grayscale) of JAFFE and CK+ datasets are almost similar, except in (Lakshmi and Ponnusamy, 2021), which shows an average difference of 7.86% on the JAFFE dataset, and 1% on the CK+ dataset. Nevertheless, the results achieved on the datasets in the YCbCr color space are still better than several recent works and attain 100% of recognition rate on the JAFFE dataset. We notice that the second combination tested in YCbCr space on all datasets, presents better results than the top one layers' combination used in the RGB color space in (Boughanem et al., 2022). This fact can be attributed to the robustness of the facial features based on the skin color driven by the YCbCr color space.

5 CONCLUSION

This work presents a relevant deep feature extraction-based method for in-the-wild FER. We implemented it from facial images in the YCbCr color space using

deep CNNs, where three CNN models have been used as feature extractors. The outcomes of the emotion recognition from facial images in the YCbCr color space prove that the extracted features contain more relevant facial expression features comparing to the RGB color space. The fact that the luminance component (Y) is separated from the two chrominance components (Cb and Cr) confirms that it does not affect the facial expressions features, what allows many feature information to be acquired robustly in-the-wild conditions as well as in controlled conditions. Therefore the YCbCr is appropriate for emotion recognition through facial images. Experiments have been conducted on three datasets: SFEW_2.0, CK+ and JAFFE, and obtained results show that the combination of deep features from different neural networks achieve a global rewarding and satisfactory recognition rates under in-the-wild and controlled environments. The findings marks recognition rates that have not been achieved before, especially for the static facial expression in the wild dataset. In future work, we will use skin color detection-based techniques for face detection, from the same color space YCbCr, while extending the method for real-time recognition.

Table 5: Comparaison of the recognition rate (%) with state-of-the-art methods on the SFEW_2.0 dataset.

Studies	SFEW_2.0	Color space
(Boughanem et al., 2022)	88.20%	RGB
(Cai et al., 2022)	62.90%	RGB
(Ruan et al., 2022)	62.16%	RGB
(Sadeghi and Raie, 2022)	61.01%	RGB
(Nan et al., 2022)	55.14%	RGB
(Zhu et al., 2022)	54.87%	RGB
(Nan et al., 2022)	54.56%	RGB
The proposed method	**92.30%**	YCbCr

Table 6: Comparison of the recognition rate (%) with state-of-the-art methods on the JAFFE and the CK+ datasets.

Studies	CK+	JAFFE
(Kar et al., 2022)	98.81%	99.30%
(Boughanem et al., 2022)	98.80%	97.62%
(Chen et al., 2022)	98.38%	99.17%
(Lakshmi and Ponnusamy, 2021)	97.66%	90.83%
The proposed method	**99.20%**	**100%**

REFERENCES

Ahmady, M., Mirkamali, S. S., Pahlevanzadeh, B., Pashaei, E., Hosseinabadi, A. A. R., and Slowik, A. (2022). Facial expression recognition using fuzzified Pseudo Zernike Moments and structural features. *Fuzzy Sets and Systems*, 443:155–172.

Al-Tairi, Z. H., Rahmat, R. W., Saripan, M. I., and Sulaiman, P. S. (2014). Skin segmentation using yuv and rgb color spaces. *Journal of information processing systems*, 10(2):283–299.

Banskota, N., Alsadoon, A., Prasad, P. W. C., Dawoud, A., Rashid, T. A., and Alsadoon, O. H. (2022). A novel enhanced convolution neural network with extreme learning machine: facial emotional recognition in psychology practices.

Bejaoui, H., Ghazouani, H., and Barhoumi, W. (2017). Fully automated facial expression recognition using 3d morphable model and mesh-local binary pattern. In *Advanced Concepts for Intelligent Vision Systems*, pages 39–50.

Bejaoui, H., Ghazouani, H., and Barhoumi, W. (2019). Sparse coding-based representation of lbp difference for 3d/4d facial expression recognition. *Multimedia Tools and Applications*, 78(16):22773–22796.

Billah, M., Wang, X., Yu, J., and Jiang, Y. (2022). Real-time goat face recognition using convolutional neural network. *Computers and Electronics in Agriculture*, 194:106730.

Boughanem, H., Ghazouani, H., and Barhoumi, W. (2021). Towards a deep neural method based on freezing layers for in-the-wild facial emotion recognition. In *2021 IEEE/ACS 18th Int Conference on Computer Systems and Applications (AICCSA)*, pages 1–8.

Boughanem, H., Ghazouani, H., and Barhoumi, W. (2022). Multichannel convolutional neural network for human emotion recognition from in-the-wild facial expressions. *The Visual Computer*, pages 1–26.

Cai, J., Meng, Z., Khan, A. S., Li, Z., O'Reilly, J., and Tong, Y. (2022). Probabilistic attribute tree structured convolutional neural networks for facial expression recognition in the wild. *IEEE Transactions on Affective Computing*.

Chen, Q., Jing, X., Zhang, F., and Mu, J. (2022). Facial expression recognition based on a lightweight cnn model. In *2022 IEEE Int Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, pages 1–5.

Deng, J., Wang, X., and Zhang, H. (2021). Online environment abnormal expression detection based on improved autoencoder. In *2021 IEEE DASC/PiCom/CBDCom/CyberSciTech*, pages 554–559.

Dhall, A., Goecke, R., Joshi, J., Sikka, K., and Gedeon, T. (2014). Emotion recognition in the wild challenge 2014: Baseline, data and protocol. In *Int Conference on Multimodal Interaction*, pages 461–466.

Dhall, A., Goecke, R., Lucey, S., and Gedeon, T. (2012). Collecting large, richly annotated facial-expression databases from movies. *IEEE Multimedia*, 19(3):34–41.

Ghazouani, H. (2021). A genetic programming-based feature selection and fusion for facial expression recognition. *Applied Soft Computing*, 103:107173.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of*

the IEEE conference on computer vision and pattern recognition, pages 770–778.

Hosgurmath, S., Mallappa, V. V., Patil, N. B., and Petli, V. (2022). Effective face recognition using dual linear collaborative discriminant regression classification algorithm. *Multimedia Tools and Applications*, 81(5):6899–6922.

Hu, Y., Xu, Y., Zhuang, H., Weng, Z., and Lin, Z. (2022). Machine learning techniques and systems for mask-face detection—survey and a new ood-mask approach. *Applied Sciences*, 12(18).

Ittahir, S., Idbeaa, T., and Ogorban, H. (2022). The system for estimating the number of people in digital images based on skin color face detection algorithm. *AlQalam Journal of Medical and Applied Sciences*, pages 215–225.

Kanade, T., Cohn, J. F., and Tian, Y. (2000). Comprehensive database for facial expression analysis. In *IEEE Int Conference on Automatic Face and Gesture Recognition*, pages 46–53.

Kar, N. B., Babu, K. S., and Bakshi, S. (2022). Facial expression recognition system based on variational mode decomposition and whale optimized kelm. *Image and Vision Computing*, page 104445.

Khanam, R., Johri, P., and Diván, M. J. (2022). *Human Skin Color Detection Technique Using Different Color Models*, pages 261–279.

Lakshmi, D. and Ponnusamy, R. (2021). Facial emotion recognition using modified hog and lbp features with deep stacked autoencoders. *Microprocessors and Microsystems*, 82:103834.

Li, Z.-J. (2022). A method of improving accuracy in expression recognition. *European Journal of Electrical Engineering and Computer Science*, 6(3):27–30.

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., and Matthews, I. (2010). The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *2010 IEEE conference on computer vision and pattern recognition*, pages 94–101.

Lyons, M., Kamachi, M., and Gyoba, J. The japanese female facial expression (jaffe) dataset.

Nan, F., Jing, W., Tian, F., Zhang, J., Chao, K.-M., Hong, Z., and Zheng, Q. (2022). Feature super-resolution based Facial Expression Recognition for multi-scale low-resolution images. *Knowledge-Based Systems*, 236:107678.

Nugroho, H. A., Goratama, R. D., and Frannita, E. L. (2021). Face recognition in four types of colour space: a performance analysis. In *Materials Science and Engineering*, volume 1088, page 012010.

Putra, I., Wiastini, N., Wibawa, K. S., and Putra, I. M. S. (2020). Identification of skin disease using k-means clustering, discrete wavelet transform, color moments and support vector machine. *Int J. Mach. Learn. Comput*, 10(5):700–706.

Rahman, M. A., Purnama, I. K. E., and Purnomo, M. H. (2014). Simple method of human skin detection using hsv and ycbcr color spaces. In *2014 Int Conference on Intelligent Autonomous Agents, Networks and Systems*, pages 58–61.

Ruan, D., Mo, R., Yan, Y., Chen, S., Xue, J.-H., and Wang, H. (2022). Adaptive deep disturbance-disentangled learning for facial expression recognition. *Int Journal of Computer Vision*, 130(2):455–477.

Sadeghi, H. and Raie, A.-A. (2022). Histnet: Histogram-based convolutional neural network with chi-squared deep metric learning for facial expression recognition. *Information Sciences*, 608:472–488.

Saurav, S., Gidde, P., Saini, R., and Singh, S. (2022). Dual integrated convolutional neural network for real-time facial expression recognition in the wild. *The Visual Computer*, 38(3):1083–1096.

Siam, A. I., Soliman, N. F., Algarni, A. D., El-Samie, A., Fathi, E., and Sedik, A. (2022). Deploying machine learning techniques for human emotion detection. *Computational Intelligence and Neuroscience*, 2022:8032673.

Sidhom, O., Ghazouani, H., and Barhoumi, W. (2023). Subject-dependent selection of geometrical features for spontaneous emotion recognition. *Multimedia Tools and Applications*, 82(2):2635–2661.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9.

Terrillon, J.-C., Shirazi, M. N., Fukamachi, H., and Akamatsu, S. (2000). Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images. In *4th IEEE Int Conference on Automatic Face and Gesture Recognition*, pages 54–61.

Vansh, V., Chandrasekhar, K., Anil, C. R., and Sahu, S. S. (2020). Improved face detection using ycbcr and adaboost. In Behera, H. S., Nayak, J., Naik, B., and Pelusi, D., editors, *Computational Intelligence in Data Mining*.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 conference on computer vision and pattern recognition. CVPR 2001*, volume 1, pages I–I.

Yamada, Y., Inagawa, T., Hirabayashi, N., and Sumiyoshi, T. (2022). Emotion recognition deficits in psychiatric disorders as a target of non-invasive neuromodulation: A systematic review. *Clinical EEG and Neuroscience*, 53(6):506–512.

Yan, H., Liu, Y., Wang, X., Li, M., and Li, H. (2021). A face detection method based on skin color features and adaboost algorithm. In *Journal of Physics: Conference Series*, volume 1748, page 042015.

Zhu, Q., Mao, Q., Jia, H., Noi, O. E. N., and Tu, J. (2022). Convolutional relation network for facial expression recognition in the wild with few-shot learning. *Expert Systems with Applications*, 189:116046.