Face Blending Data Augmentation for Enhancing Deep Classification

Emna Ghorbel[a], Ghada Maddouri and Faouzi Ghorbel

CRISTAL Laboratory, GRIFT Research Group ENSI, La Manouba University 2010, La Manouba, Tunisia

Keywords: Low Size Dataset, Face Data Augmentation, InceptionV3, Face Classification.

Abstract: Facial image classification plays a vital role in computer vision applications, particularly in face recognition. Convolutional Neural Networks have excelled in this domain, however, their performance decline when dealing with small facial datasets. In that context, data augmentation methods have been proposed. In line with this, we introduce the Face Blending data augmentation method, which augments intra-class variability while preserving image semantics. By interpolating faces, we generate non-linear deformations, resulting in in-between images that maintain the original's global aspect. Results show that Face Blending significantly enhances facial classification. Comparisons with Mix-up and Random Erasing techniques reveal improved accuracy, precision, recall, and F1 score, particularly with limited datasets. This method offers promise for realistic applications contributing to more reliable and accurate facial classification systems with limited data.

1 INTRODUCTION

In the field of computer vision, Convolutional Neural Networks (CNNs) have made significant strides in the recognition and classification of facial images. In fact, CNNs have demonstrated their effectiveness in a wide range of applications within facial analysis, whether it involves 2D or 3D facial structures, grayscale, or color images. Nevertheless, the performance of CNNs can decline when confronted with the challenges of small-scale facial datasets. The learning phase of neural network models demands copious data for convergence, and such datasets, in practical applications, often fall short. To address this limitation, several data augmentation methods have been proposed (Summers and Dinneen, 2019; Inoue, 2018; Kang et al., 2017; Zhong et al., 2020; Gatys et al., 2015; Konno and Iwazume, 2018; Bowles et al., 2018; Su et al., 2019; El-Sawy et al., 2016; Patel et al., 2019; Ciregan et al., 2012; Sato et al., 2015; Patel et al., 2019; Yin et al., 2019; Paulin et al., 2014; Chatfield et al., 2014). These techniques can be classified into three distinct categories. The first one encompasses the geometric-based methods as the similarities' transformations (scale, rotation, translation, and flipping) on images (Ciregan et al., 2012; Sato et al., 2015; Simard et al., 2003), Part-based method applying linear and affine transformations on shape parts after a cut detection process (Patel et al., 2019), and the Shape morphing based technique for augmenting 2D shapes (Ghorbel et al., 2023). The second category is driven by deep learning as Neural Style Transfer data augmentation (Gatys et al., 2015), Features space based on Auto-Encoders (Konno and Iwazume, 2018), Generative Adversarial Neuronal (GANs) architectures (Bowles et al., 2018). The third category operates at the pixel level, deploying techniques such as kernel filters (Kang et al., 2017), homography (Paulin et al., 2014), mixing pixels (Summers and Dinneen, 2019; Inoue, 2018), and random erasing techniques (Yin et al., 2019; Zhong et al., 2020).

Despite their contributions, adapting these augmentation methods for facial classification can be especially challenging, given the nuanced complexities involved in this task. Often, these methods prove inadequate in capturing the nuances of intra-class variations, which may lead to the loss of significant facial details in the process.

In this paper, we introduce a novel data augmentation technique meticulously crafted to enhance CNN performance in facial classification. Our method aims to augment intra-class variability while preserving the semantic integrity of facial images. We present the Face Blending data augmentation, tailored specifically to improve facial classification tasks. This paper explores the theoretical foundations, integration of Face Blending into CNN architecture, and its application to three small-scale facial datasets. Our objective is to provide compelling evidence that this in-

[a] https://orcid.org/0000-0002-6179-1358

Ghorbel, E., Maddouri, G. and Ghorbel, F.
Face Blending Data Augmentation for Enhancing Deep Classification.
DOI: 10.5220/0012357900003654

In *Proceedings of the 13th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2024)*, pages 274-280
ISBN: 978-989-758-684-2; ISSN: 2184-4313

novative technique significantly boosts facial classification accuracy, particularly in scenarios with limited datasets. Notably, this method is a hybrid, combining geometric and pixel-based approaches.

Throughout this paper, we present a comprehensive analysis of the Face Blending data augmentation method as applied to facial classification. We compare its performance with existing techniques and substantiate its efficacy in elevating the performance of facial classification while preserving the meaningful content of facial images.

2 DATA AUGMENTATION BASED ON FACE BLENDING

In this section, the data augmentation based on face blending is proposed. Facial blending, often employed in image processing and computer vision, is a technique that combines different facial features or characteristics from multiple images to create a composite or blended face. We describe in the following the key components of our proposed Face Blending data augmentation method, which include Landmark Face Detection (Xiang and Zhu, 2017), Face Interpolation based on Warping and Delaunay Triangulation(Wolberg, 1998), and Data Cleaning through face detection.

The process begins by selecting pairs of images denoted respectively A and B from the same class of a dataset $D = \{Class_1, ...Class_k\}$. Then, the landmark face detection algorithm namely the Multi-Task Cascaded Convolutional Networks (MTCNN) (Xiang and Zhu, 2017) is applied on faces in order to extract the face reference points noted respectively $\{(x_{1_A}, y_{1_A}), \ldots, (x_{i_A}, y_{i_A}), \ldots, (x_{n_A}, y_{n_A})\}$ and $\{(x_{1_B}, y_{1_B}), \ldots, (x_{i_B}, y_{i_B}), \ldots, (x_{n_B}, y_{n_B})\}$. From Image A and Image B, and their corresponding landmarks, the Delaunay triangulation is performed on the landmarks separately for both images, resulting in two sets of triangles, namely T_A and T_B, respectively.

For each corresponding pair of triangles (t_A, t_B) (one from T_A and one from T_B), points positions are interpolated in order to create a new set of triangles T_{AB}, which consist in the morphed face. Therefore, we perform the warping phase by blending the positions of the triangles in T_A and T_B with $t \in [0, 1]$ in order to obtain the in-between images as follows,

$$P_i(t) = (1 - t) \cdot P_i^A + t \cdot P_i^B, \quad 1 \leq i \leq n$$

where each vertex $P_i(t) \in T_{AB}$ is the position of P_i at time t, P_i^A is the position of P_i in Image A, P_i^B is the position of P_i in Image B.

Thereafter, in-between triangular meshes are generated, where each triangle in T_{AB} is deformed accordingly based on the warped vertices $P_i(t)$. Then, the morphed image is obtained by mapping the pixel values from the original images (A and B) to the corresponding triangles in T_{AB} based on the barycentric coordinates of each pixel within the triangles. Thereafter, m images are generated. Figure 2 illustrates an example of the obtained face interpolation. Since in-between images are obtained, we carry out post-processing for cleaning the generated object in order to ensure the integrity and consistency of our augmented face dataset. Therefore, We reapply the robust face detection algorithm MTCNN (Xiang and Zhu, 2017) to thoroughly scan and localize faces within the augmented dataset. Any objects identified as non-facial are subsequently removed, ensuring that the dataset exclusively comprises facial samples. This meticulous cleaning methodology ensures that our face dataset remains free from non-face elements. The Face Blending data augmentation pipeline is illustrated in Figure 1, showcasing the stages of our approach.

Figure 3 presents examples of obtained results where the in-between images include faces belonging to a same class. In the following, we propose to validate the proposed method qualitatively and quantitatively through a chosen model in the case of low-size face datasets.

3 EXPERIMENTS

In this part, qualitative and quantitative results are presented in order to validate the proposed method for enhancing face classification.

3.1 Datasets

We evaluate the proposed method with the three following small benchmarks. The Face Shape dataset (Lucifierx, 2022) encompasses seven distinctive face shape categories. These categories represent a variety of facial shapes, including heart, oval, square, and more. Within this dataset, each of the seven classes comprises a carefully selected collection of facial images, with each category containing between 10 to 20 images. This dataset serves as a valuable resource for facial shape identification and differentiation. The 14 Celebrity Faces Dataset (Nelson, 2018) is composed of 14 distinct classes. Each class contains between 15 and 20 images of the respective celebrity. This dataset has been curated for the specific purpose of celebrity face recognition, ensuring a balanced distri-

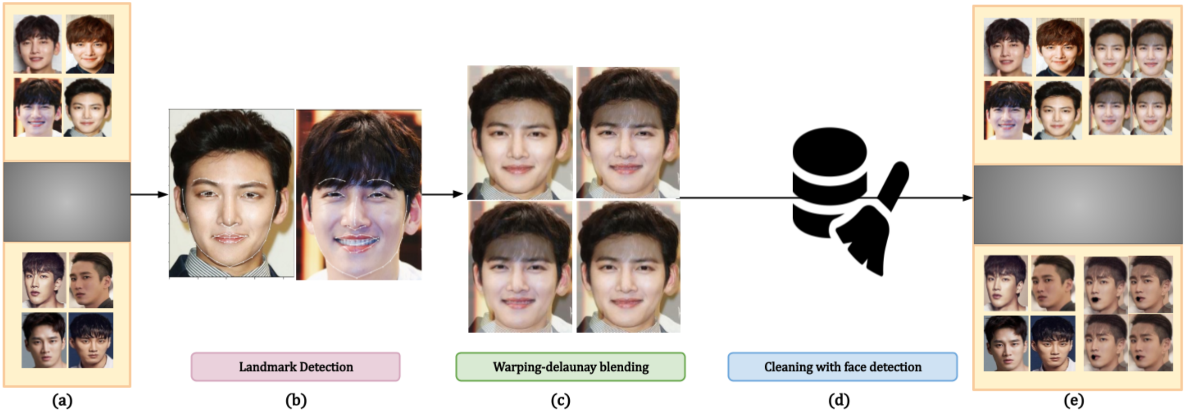

Figure 1: Blending Face Data Augmentation pipeline; (a) Original dataset, (b) MTCNN landmark detection on pair of images belonging to a same class, (c) generated in-between images with the Warping-Delaunay blending, (d) MTCNN face detection for the data cleaning postprocessing, and (e) Augmented dataset.

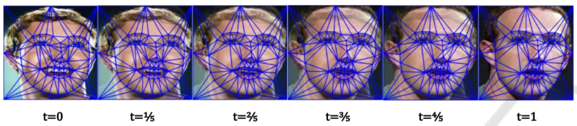

Figure 2: An example of a warping-delaunay-based blending between two faces belonging to a same class.

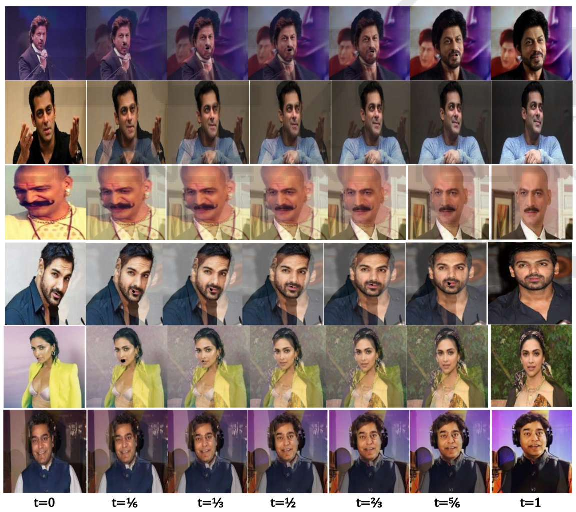

Figure 3: Examples of images generated with the Data augmentation Blending Face method applied on The Pathaan Movie Characters Dataset (Ichhadhari, 2019).

bution of images across different personalities to facilitate accurate identification and distinction. The 5 Celebrity Faces Dataset (DanB, 2017) presents a robust resource for celebrity recognition. This dataset features five distinct classes, each corresponding to an actor and containing between 15 and 20 images.

3.2 Implementation Settings

Since the blending face data augmentation method is fully automatic, we propose to add it as an augmentation layer. We use Python and the Keras framework to carry out our experimental study. All experiments are conducted on the Inception convolutional neuronal networks namely InceptionV3 (Szegedy et al., 2015) pre-trained on ImageNet. The datasets are separated into three sets: 70% are used for the training, 10% for the validation, and 20% for the testing. We use a learning rate of 0.006 and a momentum of 0.1 with an SGD optimizer.

3.3 Qualitative Results

Here, qualitative examples of data augmentation results for each dataset are presented.

Figures 4, 5 and 6 illustrate several examples of blending data augmentation applied on the Face Shape dataset, the 14 Celebrity Faces dataset, and the 5 Celebrity Faces Dataset, respectively. Results, in all cases, are visually satisfactory. In fact, this data augmentation method generates images that have semantic meaning. This will significantly improve the classification results, as observed in the next subsection.

3.4 Quantitative Results

In this part, quantitative results are reported. First, both the accuracy and loss history of the blending-face-InceptionV3 model trained on each dataset are presented. Then, comparisons with Mix-up and Random erasing data augmentation methods are conducted according to several metrics of which the Accuracy, the weighted average (w.a.) Precision, the w.a.

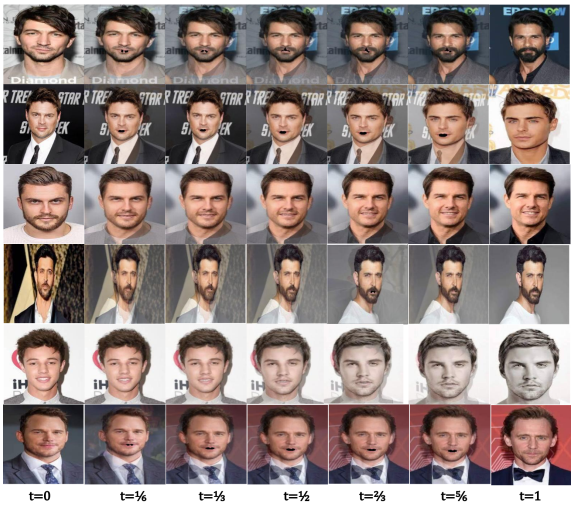

Figure 4: Examples of images generated with the Data augmentation Blending face method applied on the Face Shape Dataset.

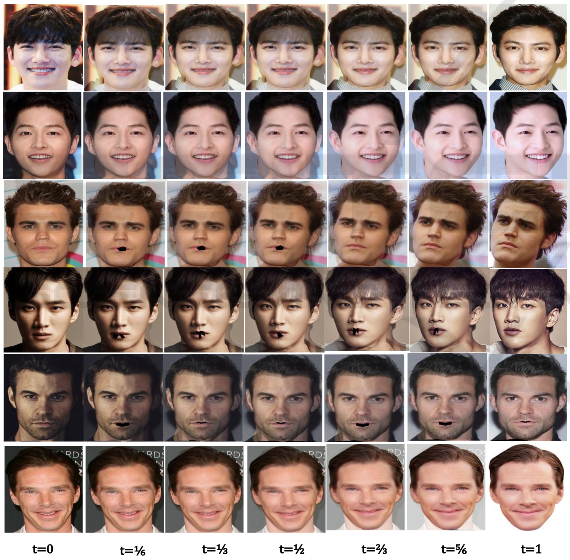

Figure 5: Examples of images generated with the Data augmentation Blending face method applied on the 14 Celebrity Faces Dataset.

Recall, the w.a. F1 score. Note that weighted average metrics are used because of the unbalanced classes.

To demonstrate the added value of our proposed method, we conduct a comparative analysis between the "Blending-Face-InceptionV3" model and the standard InceptionV3 model according to loss, validation loss, accuracy, and validation accuracy histories for each dataset. Figures 7, 8, and 9 illustrate these comparisons when models are trained on the Face Shape dataset, the 14 Celebrity Faces dataset, and 5 Celebrity Faces Dataset, respectively. In all

Figure 6: Examples of images generated with the Data augmentation Blending face method applied on the 5 Celebrity Faces Dataset.

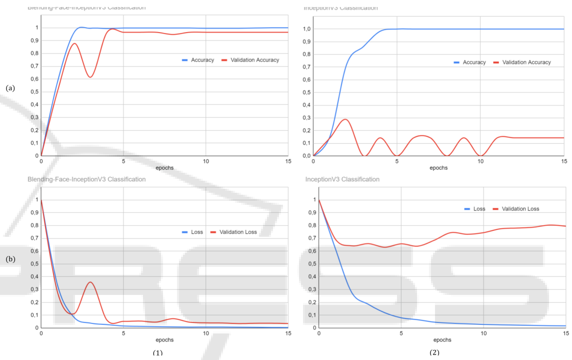

Figure 7: Comparison between Blending-Face-InceptionV3 and InceptionV3 according to the Training and Validation metrics when trained on the Face Shape dataset. (a-1) Accuracy and Accuracy validation of the proposed model. (b-1) Accuracy and Accuracy validation of InceptionV3. (a-2) Loss and Loss validation of the proposed model. (b-2) Loss and Loss validation of InceptionV3.

cases, the Blending-Face-InceptionV3 models is, relatively, performing well on the training data while providing a significative generalization.

Figure 10 presents a comparison between no augmentation (**No-aug**), **Mix-up** (Summers and Dinneen, 2019), Random Erasing (**RE**) (Zhong et al., 2020), and Blending Face (**BF**) data augmentation techniques according to the training metrics of which accuracy and loss.

In the case of the Face Shape Dataset, we observe in Table 1 that the model without augmentation achieves the lowest accuracy (23.53%). The weighted average (w.a.) precision is also relatively low (0.15), suggesting a high rate of false positives. The w.a. recall and F1 score are also relatively low, indicating that the model struggled to identify relevant instances in the data. The **Mix-up** method performed better

Table 1: Comparison of several data augmentation methods with the pre-trained InceptionV3 model trained on the Face Shape dataset according to various performance metrics.

Method	Accuracy(%)	Precision	Recall	F1 score
No-aug	23.53	0.15	0.06	0.08
Mix-up (Summers and Dinneen, 2019)	10.53	0.03	0.11	0.05
RE (Zhong et al., 2020)	10.53	0.11	0.11	0.11
BF (ours)	95.03	0.97	0.91	0.94

Table 2: Comparison of several data augmentation methods with the pre-trained InceptionV3 model trained on the 14 Face Celebrity dataset according to various performance metrics.

Method	Accuracy(%)	Precision	Recall	F1 score
No-aug	67.86	0.83	0.18	0.29
Mix-up (Summers and Dinneen, 2019)	62.96	0.15	0.11	0.12
RE (Zhong et al., 2020)	51.85	0.15	0.11	0.12
BF (ours)	96.80	1.00	0.95	0.97

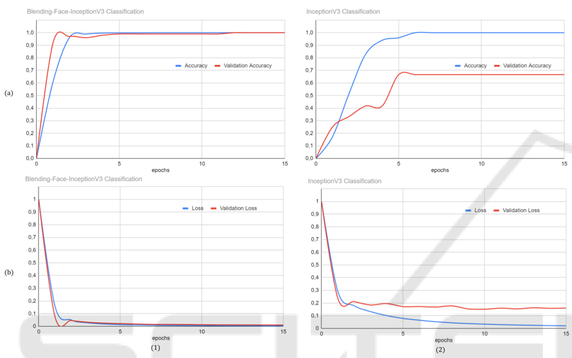

Figure 8: Comparison between Blending-Face-InceptionV3 and InceptionV3 according to the Training and Validation metrics when trained on 14 Celebrity Faces dataset. (a-1) Accuracy and Accuracy validation of the proposed model. (b-1) Accuracy and Accuracy validation of InceptionV3. (a-2) Loss and Loss validation of the proposed model. (b-2) Loss and Loss validation of InceptionV3.

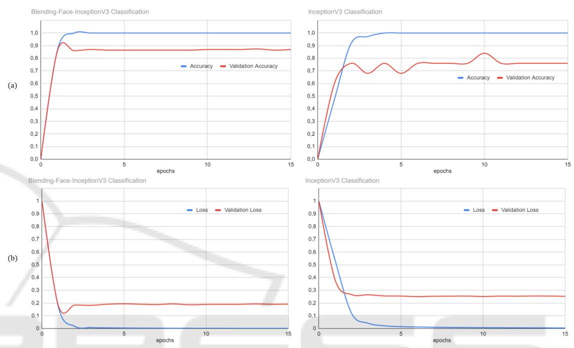

Figure 9: Comparison between Blending-Face-InceptionV3 and InceptionV3 according to the Training and Validation metrics when trained on 5 Celebrity Faces Dataset. (a-1) Accuracy and Accuracy validation of the proposed model. (b-1) Accuracy and Accuracy validation of InceptionV3. (a-2) Loss and Loss validation of the proposed model. (b-2) Loss and Loss validation of InceptionV3.

than **No-aug** in terms of accuracy (10.53%), but it's still quite low. Precision is very low (0.03), indicating a high rate of false positives. The recall is relatively high (0.11), but the F1 score is still quite low. Similar to **Mix-up**, **RE**, method achieved low accuracy (10.53%). Precision (0.11) and recall (0.11) are also quite low, resulting in a low F1 score (0.11). We observe that the **BF** method outperforms the others by a significant margin with an accuracy of 95.03%. The precision, recall, and F1 score are all relatively high, indicating that this method is effective in enhancing the model's performance for face shape classification.

When performing classification on the 14 Celebrity Faces Dataset as shown in Table 2, the original model achieves an accuracy of 67.86%. The precision is quite high (0.83), suggesting a low rate of false positives. However, the recall is relatively low (0.18), resulting in an F1 score of 0.29 announcing overfitting. The **Mix-up** method achieved a relative lower accuracy as well as **RE**. Similar to the first table, the **BF** method stands out with high accuracy (96.8%) and high precision (1.00). The recall is also quite high (0.95).

The results presented in Table 3, based on the 5 Celebrity Faces Dataset, offer a comprehensive comparison of these methods. First, the performance without data augmentation attains an accuracy of 73.68%, accompanied by a precision of 0.84, a recall of 0.53, and an F1 score of 0.62. The **Mix-up** method enhances accuracy to 81.25%, it comes with lower precision, recall, and F1 score values, standing at 0.18, 0.19, and 0.18, respectively. The **RE** method has obtained an accuracy of 68.75%. The precision, recall, and F1 score exhibit values of 0.32, 0.38, and 0.34, respectively. **BF** presents relatively exceptional results with an accuracy of 98.50% while all three metrics achieve 0.98, underscoring the efficacy of our approach.

Blending Face data augmentation consistently

Table 3: Comparison of several data augmentation methods with the pre-trained InceptionV3 model trained on 5 Celebrity Faces Dataset according to various performance metrics.

Method	Accuracy(%)	Precision	Recall	F1 score
No-aug	73.68	0.84	0.53	0.62
Mix-up (Summers and Dinneen, 2019)	81.25	0.18	0.19	0.18
RE (Zhong et al., 2020)	68.75	0.32	0.38	0.34
BF (ours)	98.50	0.98	0.98	0.98

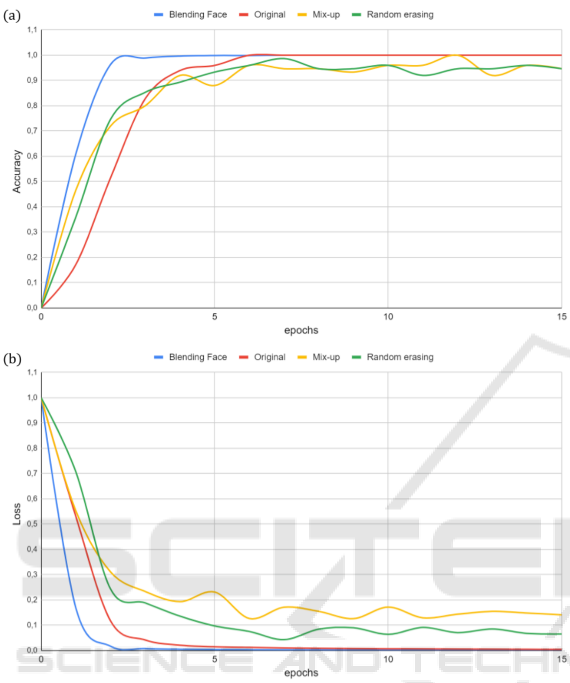

Figure 10: Comparison between various data augmentation techniques with Pre-trained InceptionV3 according to the Training metrics when trained on 5 Celebrity Faces Dataset. (a) Accuracy comparisons. (b) Loss comparisons.

outperforms the other augmentation methods in terms of accuracy, w.a. precision, w.a. recall, and w.a. F1 score. This suggests that, at least for low-size datasets, the proposed data augmentation technique is highly effective in improving model performance for face classification. However, it is important to note a limitation, which is the presence of minor artifacts in the generated images, as can be observed in three, fourth and fifth rows in Figure 5 above. Specifically, small black pixels may appear near the mouth area due to the high gradient between the teeth and the lips. Nonetheless, it is essential to emphasize that this limitation does not adversely affect the performance of the Inception V3 classification, as demonstrated by our results. Additionally, the individual's facial identity remains recognizable despite this minor occlusion.

4 CONCLUSION

In this paper, we introduced the Face Blending data augmentation method, specifically designed to enhance the performance of Convolutional Neural Networks (CNNs) in facial classification, particularly when dealing with small-scale datasets. This innovative approach aims to augment intra-class variability while preserving the meaning of images. Therefore, We have demonstrated that the proposed method significantly improves facial classification accuracy, as evidenced by a series of comprehensive experiments on three distinct small-scale facial datasets. In fact, it outperformed conventional augmentation methods, resulting in significantly improved InceptionV3 accuracy, precision, recall, and F1 score. Future work should further refine the Face Blending data augmentation for specific facial analysis tasks as improving real-world recognition systems with limited data.

REFERENCES

Bowles, C., Chen, L., Guerrero, R., Bentley, P., Gunn, R., Hammers, A., Dickie, D. A., Hernández, M. V., Wardlaw, J., and Rueckert, D. (2018). Gan augmentation: Augmenting training data using generative adversarial networks. *arXiv preprint arXiv:1810.10863*.

Chatfield, K., Simonyan, K., Vedaldi, A., and Zisserman, A. (2014). Return of the devil in the details: Delving deep into convolutional nets. *arXiv preprint arXiv:1405.3531*.

Ciregan, D., Meier, U., and Schmidhuber, J. (2012). Multi-column deep neural networks for image classification. In *2012 IEEE conference on computer vision and pattern recognition*, pages 3642–3649. IEEE.

DanB (2017). 5 celebrity faces dataset.

El-Sawy, A., Hazem, E.-B., and Loey, M. (2016). Cnn for handwritten arabic digits recognition based on lenet-5. In *International conference on advanced intelligent systems and informatics*, pages 566–575. Springer.

Gatys, L. A., Ecker, A. S., and Bethge, M. (2015). A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576*.

Ghorbel, E., Ghorbel, M., and Mhiri, S. (2023). Data augmentation based on invariant shape blending for deep learning classification. In *ICASSP 2023 - 2023 IEEE*

International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 1–5.

Ichhadhari (2019). Pathaan movie characters dataset.

Inoue, H. (2018). Data augmentation by pairing samples for images classification. *arXiv preprint arXiv:1801.02929*.

Kang, G., Dong, X., Zheng, L., and Yang, Y. (2017). Patchshuffle regularization. *arXiv preprint arXiv:1707.07103*.

Konno, T. and Iwazume, M. (2018). Icing on the cake: An easy and quick post-learnig method you can try after deep learning. *arXiv preprint arXiv:1807.06540*.

Lucifierx (2022). Face shape classification dataset.

Nelson, D. (2018). 14 celebrity faces dataset.

Patel, V., Mujumdar, N., Balasubramanian, P., Marvaniya, S., and Mittal, A. (2019). Data augmentation using part analysis for shape classification. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1223–1232.

Paulin, M., Revaud, J., Harchaoui, Z., Perronnin, F., and Schmid, C. (2014). Transformation pursuit for image classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3646–3653.

Sato, I., Nishimura, H., and Yokoi, K. (2015). Apac: Augmented pattern classification with neural networks. *arXiv preprint arXiv:1505.03229*.

Simard, P. Y., Steinkraus, D., Platt, J. C., et al. (2003). Best practices for convolutional neural networks applied to visual document analysis. In *Icdar*, volume 3, pages –.

Su, J., Vargas, D. V., and Sakurai, K. (2019). One pixel attack for fooling deep neural networks. *IEEE Transactions on Evolutionary Computation*, 23(5):828–841.

Summers, C. and Dinneen, M. J. (2019). Improved mixed-example data augmentation. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1262–1270. IEEE.

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.

Wolberg, G. (1998). Image morphing: a survey. *The visual computer*, 14(8-9):360–372.

Xiang, J. and Zhu, G. (2017). Joint face detection and facial expression recognition with mtcnn. In *2017 4th international conference on information science and control engineering (ICISCE)*, pages 424–427. IEEE.

Yin, D., Lopes, R. G., Shlens, J., Cubuk, E. D., and Gilmer, J. (2019). A fourier perspective on model robustness in computer vision. *arXiv preprint arXiv:1906.08988*.

Zhong, Z., Zheng, L., Kang, G., Li, S., and Yang, Y. (2020). Random erasing data augmentation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 13001–13008.