Improving Lightweight Convolutional Neural Network for Facial Expression Recognition via Transfer Learning

By Anggit Wikanningrum

Improving Lightweight Convolutional Neural Network for Facial Expression Recognition via Transfer Learning

Anggit Wikanningrum*, Reza Fuad Rachmadi† and Kohichi Ogata‡
*Department of Informatics, Universitas Dr. Soetomo, Surabaya, Indonesia
†Department of Computer Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
‡Department of Computer Science and Electrical Engineering, Kumamoto University, Kumamoto, Japan
anggit.wikaningrum@unitomo.ac.id, fuad@its.ac.id
ogata@cs.kumamoto-u.ac.jp

***Abstract*—Image-based facial expression recognition is an important problem especially for analyzing the human emotion or feeling under a specific condition, such as while watching movie scene or playing a computer game. Furthermore, the convolutional neural network (CNN) is one of the underlying technology proven to be applicable to image-based facial expression recognition problem. Unfortunately, the available CNN architecture that applied for image-based facial expression recognition problem only focuses on the accuracy instead of other factors such as the number of parameters and the execution time. In this paper, we investigated whether transfer learning from a medium-size and large-size dataset is feasible to improve the performance of lightweight CNN architecture on image-based facial expression recognition problem. We use lightweight residual-based CNN architecture originally used for CIFAR dataset to analyze the effect of the transfer learning from five different datasets, including CIFAR10, CIFAR100, ImageNet32, CINC-10, and CASIA-WebFace. The FER+ (Facial Expression Recognition Plus) dataset is used to evaluate the lightweight CNN architecture performance. Experiments show that our lightweight CNN classifier can also be improved even when the transfer learning performing from middle-size dataset comparing when training the classifier from scratch.**

***Index Terms*—facial expression recognition, lightweight deep convolutional neural network, transfer learning**

I. Introduction

Image-based facial expression recognition is one of the important problems and can be applied for a lot of applications, including movie scene analysis, human-robot interaction, and human physiology understanding. The demand of solving image-based facial expression recognition makes researchers to propose several different image-based human facial expression recognition datasets, including FER (Facial Expression Recognition) [1], FER+ [2], CK (Cohn-Kanade) [3], and CK+ [4]. With the availability of the facial expression recognition dataset, the researcher proposed some state-of-the-art method for image-based facial expression recognition problem, including deep-learned Tandem Facial Expression (TPR) [5], adaptive deep metric learning [6], joint fine tuning [7], VGG19 [2] and inception [8]. As the popularity of deep learning, all state-of-the-art approaches on image-based facial expression recognition problem use deep CNN as their main method. One of the advantages of deep learning method is the transfer learning mechanism capability. There are two different things that need to be noted when using deep learning, the computation power needed for the model and the size of the model. Usually, the deep learning method for image-based facial expression recognition problem will have more parameters by extending the network to improve the performance of the classifier [5]–[7]. As far as we concern, no one tries to perform image-based facial expression recognition problem under limited resource assumption and design the classifier with a minimum number of parameters.

In this paper, we investigated the effect of transfer learning mechanism from a middle-size dataset and huge-size dataset on image-based facial expression recognition problem. Firstly, we utilize lightweight residual-based CNN architecture [9] originally used for CIFAR dataset on the medium-size and large-size dataset and secondly we fine-tuning the classifier using FER+ dataset [2]. We assume that the lightweight CNN model will have the maximum number of parameters roughly around 1 million. Our contributions can be listed as follows

- We have investigated the effect of transfer learning mechanism for residual-based lightweight CNN architecture [9] on image-based facial expression recognition problem using FER+ dataset [2].
- We use several different datasets which categorize as a medium-size dataset (CIFAR [10] and CINC-10 [11]), and large-size dataset (ImageNet32 [12]) as the source of the weights for the fine-tuning process. We also investigated the transfer learning from the relatively same domain by using CASIA-WebFace dataset as the source of the weights for the fine-tuning process.
- We proved that transfer learning from a medium-size dataset can also improve the performance of the classifier on image-based facial expression recognition problem.

The rest of the paper organizes as follows. Section II described the related work on image-based facial expression recognition problem and transfer learning. The experiments setup and the results described in section III and IV respectively. Lastly, we conclude the experiments in section V.

978-1-7281-2965-5/19/$31.00 ©2019 IEEE

II. Related Work

In this section, we described the related work on image-based facial expression recognition and the transfer learning method.

A. *Image-based facial expression recognition*

Image-based facial expression recognition task has been one of the long-term research topics on computer vision field. In the era of deep learning, almost all of the approaches on image-based facial expression recognition are based on deep CNN classifier, including deep-learned Tandem Facial Expression (TPR) [5], adaptive deep metric learning [6], joint fine-tuning [7], VGG19 [2] and inception [8]. Barsoum et. al. [2] and Mollahoseseini et. al. [8] use state-of-the-art CNN architecture to solve image-based facial expression recognition problem. Barsoum et. al. [2] utilize VGG19 CNN architecture with additional label disrupted method to improve the accuracy of the classifier on FER+ dataset. Similar to Barsoum et al. [2] but different CNN architecture, Mollahoseseini et. al. [8] use Inception network to solve the facial expression recognition problem. Li et. al. [5] proposed joint fine-tuning of classifier (called Tandem Facial Expression-Joint Learning or TFE-JL) for face recognition and facial expression recognition by concatenating the high-level features of the classifier and add a fully-connected layer for computing the final classification score. Experiments on FER+ and CK+ dataset show that TFE-JL achieves state-of-the-art performance on CK+ and FER+ dataset. Liu et. al. [6] proposed a similar method but with a different strategy. They use Inception CNN classifier as a basis of their classifier, train the classifier on FER2013 dataset, and fine-tuning the network with additional triplet loss function attached at the end of the classifier.

B. *Transfer learning*

Transfer learning is a popular method to solve a lot of applications, including methods described in [13]–[16]. Yosinski et al. [13] investigate the transferability of deep features in the transfer learning scenario. To conduct the experiments, they create several scenarios by freezing weights on several layers and learning weights on other layers. Experiments conducted by Yosinski et al. [13] show that initializing using transferred features can improve the performance of the classifier. Ng et al. [14] perform transfer learning from ImageNet weights for emotion recognition. They use VGG-based CNN architecture to perform the experiments. Experiments on FER-2013 and EmotiW dataset show that CNN classifier with transferred features improves the performance of the classifier. Research conducted by Huh et al. [15] try to investigate what makes ImageNet good for transfer learning. Experiments on several problems, including image classification, object detection, and action recognition, show that pre-trained weights from full or selected ImageNet data will produce around the same performance which concluded that the CNN classifier is not required very large dataset as excepted before. Han et al. [16] investigate the transfer learning method with additional data augmentation taken from the web. Three classifiers are used in the experiments, including AlexNet, VGG16, and ResNet-152, and seven different datasets are used as the target of transfer learning, including Dogs dataset, Flower-102, Caltech-101, Event-8, 15 Scene, and 67 Indoor scenes. The experiments show that the classifier with pre-trained ImageNet weights produces higher accuracy comparing with the classifier trained from scratch and the data augmentation from web improve the performance of the classifier around 2%.

C. *Remarks*

As discussed before, all approaches for facial expression recognition are made for accuracy and the investigation about the lightweight CNN architecture for facial expression recognition problem is not well studied. In our opinion, lightweight CNN architecture is also important because if the facial expression recognition problem can be solved with around the same accuracy as the deeper CNN architecture, the lightweight CNN can be one of the choices to implementing the facial expression recognition system in real-world applications.

TABLE I
LIGHTWEIGHT CNN ARCHITECTURE BASED ON RESIDUAL NETWORK (ADOPTED FROM [9]).

	Num of Kernels		
	$k = 16$	$k = 32$	$k = 64$
Residual Network	$2n + 1$	$2n$	$2n$

*) n denote the number of residual module.

III. Experiments Setup

All of the experiments conducted using Caffe deep learning framework [17] and FER+ dataset [2] as the main data for evaluating the performance of lightweight CNN.

A. *Lightweight CNN*

In this paper, we use lightweight CNN architecture for facial expression recognition. The lightweight CNN is used to investigate whether the lightweight CNN architecture is feasible for facial expression recognition application. One advantage of lightweight CNN is that the classifier can easily implement in the embedded system. Follows the success of residual network CNN architecture [9], we adopted two residual networks originally used for CIFAR dataset, including ResNet-20 and ResNet32 CNN classifier. Table I shows the configuration protocol for constructing the lightweight CNN architecture. We use $n = 3$ (ResNet-20) and $n = 5$ (ResNet-32) to constructing the classifier and attaching a final fully-connected layer at the end of the classifier. To performs the transfer learning, we use network-based transfer learning method by pretraining weights of source domain dataset (e.g. CIFAR10 weights or CIFAR100 weights) on FER+ dataset directly and only changes the final layer of the CNN architecture.

TABLE II
SUMMARY OF OUR EXPERIMENTS ON FER+ DATASET (AVERAGING FROM FIVE ITERATIONS).

No	Method	Num Params	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Average
1	ResNet-20	280K	79.32	78.06	78.46	78.86	78.49	78.64
2	ResNet-20 Finetuned from CIFAR10		82.43	81.64	81.76	82.34	81.85	82.00
3	ResNet-20 Finetuned from CIFAR100		82.55	81.79	82.04	82.04	81.82	82.05
4	ResNet-20 Finetuned from ImageNet32		79.96	79.19	79.29	79.19	80.17	79.56
5	ResNet-20 Finetuned from CINIC-10		82.40	82.52	82.00	82.68	82.71	82.46
6	ResNet-20 Finetuned from CASIA-WebFace		80.97	81.33	81.27	80.60	82.10	81.25
7	ResNet-32	474K	79.68	79.93	80.11	79.99	80.11	79.96
8	ResNet-32 Finetuned from CIFAR10		83.81	84.05	84.14	84.05	83.53	83.92
9	ResNet-32 Finetuned from CIFAR100		83.50	83.32	83.26	82.86	82.55	83.09
10	ResNet-32 Finetuned from ImageNet32		78.50	78.22	79.41	79.16	78.95	78.85
11	ResNet-32 Finetuned from CINIC-10		83.75	83.17	83.99	82.55	83.62	83.42
12	ResNet-32 Finetuned from CASIA-WebFace		81.45	82.89	82.28	81.42	82.04	82.02

Fig. 1. Examples of face images on FER+ dataset.

B. FER+ Dataset

To evaluate the lightweight CNN classifier, we use FER+ facial expression recognition dataset. FER+ dataset is an improvement version of FER2013 dataset. The FER2013 dataset was created by Pierre Luc Carrier and Aaron Courville by searching face images on the internet based on emotion-related keywords. The reports provided by Goodfellow et al. [1] show that even the label on FER2013 dataset was filtered by human labelers, the label accuracy is not very high. Barsoum et al. [2] proposed dataset called FER+ dataset by re-label the FER2013 dataset using crowdsourcing and adding some several additional images to the dataset. The FER+ dataset consists of 32,615 face images with 8 different emotion types, including neutral, happiness, surprise, sadness, anger, disgust, fear, and contempt. The dataset is divided into three subsets, 26,029 face images for the training process, 3,274 face images for the validation process, and the rest of the face images are used for the testing process. Figure 1 shows face images examples on FER+ dataset along with the emotion label.

C. Training Process

The training process is done using two steps, training the model using domain source dataset (including CIFAR10, CIFAR100, ImageNet32, CINIC-10, and CASIA-WebFace) and fine-tuning the weights on FER+ dataset. Follow the strategy of residual network training process on CIFAR dataset, we do not apply data augmentation for the training process on domain source dataset except the random region cropping. For CIFAR10 and CIFAR100 dataset, we padded the image using zero paddings for 4 pixels and performs a random 32×32 region cropping. Instead of using zero paddings, we resize the image into 36×36 and performs a 32×32 random cropping for other domain source datasets. The training process on domain source dataset is done around 20-60 epochs depending on the size of the dataset.

The fine-tuning process (the second steps) is done for around 8 epochs using NAG (Nesterov Accelerated Gradient) with learning rate initialized at 0.01 and decreased by a factor of 0.1 at epoch 4 and epoch 6. The dataset for training process was balanced such that each class will have the same amount of examples by performing data augmentation, including random rotation, noise, and random translation. Unlike Barsoum et al. [2] approaches, we only use the majority voting label of the human labelers on FER+ dataset.

D. Testing Process

The testing process performs by resizing the input face into 36×36, cropping the image into ten 32×32 cropped region, and substracted the region by the training data mean value. All cropped regions were classified using the lightweight CNN classifier and the final classification score was calculated by averaging the prediction score of all cropped regions. The same process is used for the experiments using an ensemble of lightweight CNN classifier.

IV. RESULTS

We divide this section into two subsections, the results of experiments using single classifier and ensemble classifier. The

TABLE III
SUMMARY OF OUR EXPERIMENTS USING ENSEMBLE LIGHTWEIGHT CNN CLASSIFIER ON FER+ DATASET (AVERAGING FROM FIVE ITERATIONS). THE NUMBER PROVIDED IN THE ENSEMBLE CONFIGURATION IS BASED ON TABLE II.

No	Method	Num Params	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Average
1	Ensemble (2) + (3)	560K	83.26	82.65	83.17	82.83	82.46	82.87
2	Ensemble (3) + (5)		83.41	83.38	83.04	83.62	82.71	83.23
3	Ensemble (2) + (5)		83.62	83.01	83.04	83.47	83.38	83.304
4	Ensemble (2) + (3) + (5)	840K	83.68	83.32	83.29	83.32	83.35	83.39
5	Ensemble (8) + (9)	948K	84.17	84.48	84.45	84.3	84.05	84.29
6	Ensemble (8) + (11)		84.2	84.11	84.51	84.33	84.36	**84.30**
7	Ensemble (9) + (11)		84.17	83.84	84.33	83.59	84.02	83.99
8	Ensemble (2) + (3) + (8)	1034K	84.02	84.14	84.14	84.02	83.93	84.05
9	Ensemble (3) + (5) + (8)		83.96	84.3	83.87	84.2	83.65	83.99
10	Ensemble (2) + (5) + (8)		84.11	84.14	84.05	84.17	84.05	84.10
11	Ensemble (2) + (3) + (9)		83.96	83.96	83.87	83.78	83.29	83.77
12	Ensemble (3) + (5) + (9)		83.96	83.96	83.87	83.78	83.29	83.77
13	Ensemble (2) + (5) + (9)		83.68	83.56	83.81	83.9	83.35	83.66

summary of experiments using single classifier and ensemble of classifiers can be viewed in Table II and III.

A. *Single Classifier*

As shown in Table II, two lightweight CNN architectures produce mean accuracy near 80% which is considered very good accuracy on FER+ dataset. The ResNet-32 is superior comparing with ResNet-20 which is very reasonable due to more number of parameters existed in the classifier. The transfer learning method from several domain source datasets improved the performance of the classifier except when the transfer learning performed from ImageNet32 weights. One possible reason is that the ImageNet32 weights are not generalized the dataset (indicated by low training accuracy) and it affected the transfer learning process. The best accuracy of ResNet-20 classifier is achieved when trained using transfer learning from CINIC-10 weights while the best accuracy of ResNet-32 classifier is achieved when trained using transfer learning from CIFAR10 weights. The highest accuracy of single classifier (83.92%) is just around 1% lower than the accuracy reported by Barsoum et al. [2] which using deep VGG13 classifier.

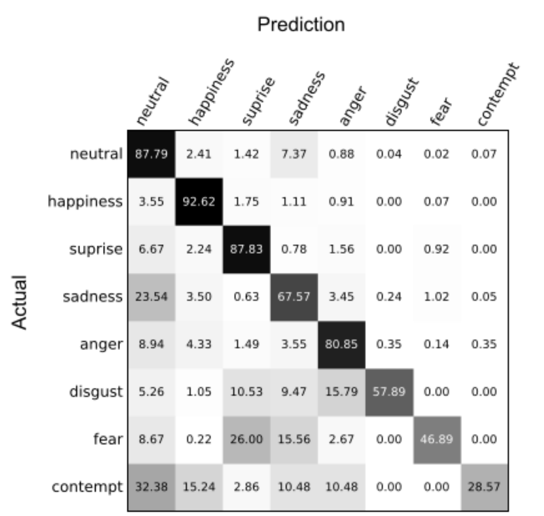

Fig. 2. Confusion matrix for Ensemble (8) + (11) configuration with accuracy of 84.30%.

B. *Ensemble Classifiers*

To improve the performance of the classifier, we also conducted experiments using 13 different ensemble configurations with the assumption that the maximum number of parameters is roughly around 1 million parameters. Table III shows the summary of the experiments using 13 ensemble configuration with the number in the ensemble configuration is the classifier order in Table II. As shown in Table III, the ensemble configuration consistently increases the performance of the classifier around 1% compared with single classifier configuration. Although some ensemble classifier has a higher number of parameter compared with other ensemble configurations, the best accuracy produces by Ensemble (8) + (11) configuration with an accuracy of 84.30%. The Ensemble (8) + (11) configuration combined two same CNN architecture but trained using different domain source dataset. Figure 2 shows the confusion matrix of Ensemble (8) + (11) classifier with a global accuracy of 84.30%. The average class accuracy of Ensemble (8) + (11) is 68.75% which is lower compared with global accuracy due to the unbalance testing dataset.

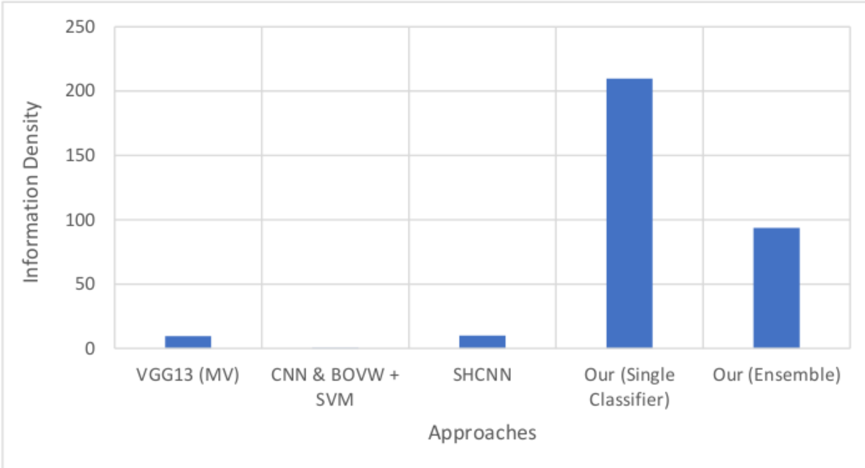

Fig. 3. Information density comparison for several state-of-the-art methods on FER+ dataset (higher is better). Only methods that reported the number of parameters is shown.

C. Comparison

To get a better understanding about the performance of lightweight CNN architecture, we compare the results with several state-of-the-art methods, including VGG13 [2], TFE-JL [5], CNN and BOVW + global SVM [18], SHCNN [19], and STL + Laplacian RTNN [20]. Table IV shows the comparison between our lightweight CNN classifier with five state-of-the-art methods on FER+ dataset along with the information about the number of parameters. As shown in Table IV, our lightweight CNN classifier can outperform VGG13 with majority voting label training reported by Barsoum et al. [2], produces the same accuracy as TFE-JL method, but still lower from other methods.

To further see the effects of the number of parameters in the classifier, we included information (if reported) about the number of parameters in the classifier. As shown in Table IV, our lightweight CNN classifier has the lowest number of parameters compared with other approaches. For comparing the effectiveness of the classifier, we compute the information density of the classifier. The information density is a metric described as a ratio between the performance of the classifier (%) and the number of parameters (in million) which can be written as follows

$$D = \frac{p_c}{n_p} \quad (1)$$

with D is the information density, p_c is the performance of the classifier in percentage, and n_p is the number of parameters in the classifier (in million). The metric is also used in [21]–[23] for evaluating the effectiveness of the classifier. Figure 3 shows the information density value comparison for several state-of-the-art methods on FER+ dataset. As shown in Figure 3, our lightweight CNN classifier has way more information density value comparing with several other state-of-the-art methods. Unfortunately, not all state-of-the-art methods on FER+ dataset reported the number of parameters in their classifier.

TABLE IV
COMPARISON OF ENSEMBLE OF LIGHTWEIGHT CNN ARCHITECTURE WITH SEVERAL STATE-OF-THE-ART METHOD ON FER+ DATASET.

Method	#Params	Acc.
VGG13 (MV) [2]	8.75M	83.8%
TFE-JL [5]	n/a	84.30%
CNN and BOVW + global SVM [18]	300M+	87.76%
SHCNN [19]	8.7M	86.45%
STL + Laplacian RTNN [20]	n/a	88.16%
Our Single Classifier (Best)	0.4M	83.92%
Our Ensemble Classifier (Best)	0.9M	84.30%

V. CONCLUSION

In this paper, we present our investigation of lightweight CNN architecture for facial expression recognition problem. We utilize two lightweight CNN architectures, ResNet-20 and ResNet-32, and improve the performance of classifier via transfer learning and ensemble configuration. Several domain source datasets, including CIFAR, CINIC-10, ImageNet32, and CASIA-WebFace are used as domain source dataset on the transfer learning process. Experiments on FER+ dataset show that the lightweight CNN architecture can produce a very good accuracy and by ensembling the classifier, the accuracy can further be improved and the results are comparable with several state-of-the-art methods. The domain source dataset is not very demanding for transfer learning and from experiments we can see that the relatively same domain problems dataset (CASIA-WebFace) produces lower accuracy compared with other domain source dataset. Although our classifier does not produce the highest accuracy on FER+ dataset, the information density of our lightweight CNN classifier is very high compared with several other state-of-the-art methods.

Joint transfer learning training and weighted ensemble configuration is our concern for future work of this research to improve the performance of the classifier. Several other facial expression datasets are also demanding to analyze using transfer learning method and lightweight CNN architecture.

References

[1] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee *et al.*, "Challenges in representation learning: A report on three machine learning contests," in *International Conference on Neural Information Processing*. Springer, 2013, pp. 117–124.

[2] E. Barsoum, C. Zhang, C. Canton Ferrer, and Z. Zhang, "Training deep networks for facial expression recognition with crowd-sourced label distribution," in *ACM International Conference on Multimodal Interaction (ICMI)*, 2016.

[3] T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*. IEEE, 2000, pp. 46–53.

[4] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*. IEEE, 2010, pp. 94–101.

[5] M. Li, H. Xu, X. Huang, Z. Song, X. Liu, and X. Li, "Facial expression recognition with identity and emotion joint learning," *IEEE Transactions on Affective Computing*, 2018.

[6] X. Liu, B. Vijaya Kumar, J. You, and P. Jia, "Adaptive deep metric learning for identity-aware facial expression recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 20–29.

[7] H. Jung, S. Lee, J. Yim, S. Park, and J. Kim, "Joint fine-tuning in deep neural networks for facial expression recognition," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 2983–2991.

[8] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in *2016 IEEE Winter conference on applications of computer vision (WACV)*. IEEE, 2016, pp. 1–10.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[10] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," Citeseer, Tech. Rep., 2009.

[11] L. N. Darlow, E. J. Crowley, A. Antoniou, and A. J. Storkey, "CINIC-10 is not imagenet or CIFAR-10," *CoRR*, vol. abs/1810.03505, 2018.

[12] P. Chrabaszcz, I. Loshchilov, and F. Hutter, "A downsampled variant of imagenet as an alternative to the cifar datasets," *arXiv preprint arXiv:1707.08819*, 2017.

[13] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Advances in neural information processing systems*, 2014, pp. 3320–3328.

[14] H.-W. Ng, V. D. Nguyen, V. Vonikakis, and S. Winkler, "Deep learning for emotion recognition on small datasets using transfer learning," in *Proceedings of the 2015 ACM on international conference on multimodal interaction*. ACM, 2015, pp. 443–449.

[15] M. Huh, P. Agrawal, and A. A. Efros, "What makes imagenet good for transfer learning?" *arXiv preprint arXiv:1608.08614*, 2016.

[16] D. Han, Q. Liu, and W. Fan, "A new image classification method using cnn transfer learning and web data augmentation," *Expert Systems with Applications*, vol. 95, pp. 43–56, 2018.

[17] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22Nd ACM International Conference on Multimedia*, ser. MM '14, 2014, pp. 675–678.

[18] M.-I. Georgescu, R. T. Ionescu, and M. Popescu, "Local learning with deep and handcrafted features for facial expression recognition," *IEEE Access*, vol. 7, pp. 64 827–64 836, 2019.

[19] S. Miao, H. Xu, Z. Han, and Y. Zhu, "Recognizing facial expressions using a shallow convolutional neural network," *IEEE Access*, vol. 7, pp. 78 000–78 011, 2019.

[20] R. K. Pandey, S. Karmakar, A. Ramakrishnan, and N. Saha, "Improving facial emotion recognition systems using gradient and laplacian images," *arXiv preprint arXiv:1902.05411*, 2019.

[21] A. Canziani, A. Paszke, and E. Culurciello, "An analysis of deep neural network models for practical applications," *arXiv preprint arXiv:1605.07678*, 2016.

[22] A. Wong, M. J. Shafiee, and M. S. Jules, "Micronnet: A highly compact deep convolutional neural network architecture for real-time embedded traffic sign classification," *IEEE Access*, vol. 6, pp. 59 803–59 810, 2018.

[23] R. F. Rachmadi, G. Koutaki, and K. Ogata, "Lightweight spatial pyramid convolutional neural network for traffic sign classification," in *2018 Indonesian Association for Pattern Recognition International Conference (INAPR)*. IEEE, 2018, pp. 23–28.

Improving Lightweight Convolutional Neural Network for Facial Expression Recognition via Transfer Learning

ORIGINALITY REPORT

14%

SIMILARITY INDEX

PRIMARY SOURCES

1	web.ee.its.ac.id Internet	114 words — 3%
2	etheses.uinsgd.ac.id Internet	85 words — 2%
3	s3-ap-southeast-1.amazonaws.com Internet	44 words — 1%
4	Diana Purwitasari, Chastine Fatichah, Surya Sumpeno, Christian Steglich, Mauridhi Hery Purnomo. "Identifying collaboration dynamics of bipartite author-topic networks with the influences of interest changes", Scientometrics, 2020 Crossref	42 words — 1%
5	Reza Fuad Rachmadi, Gou Koutaki, Kohichi Ogata. "Lightweight Spatial Pyramid Convolutional Neural Network for Traffic Sign Classification", 2018 Indonesian Association for Pattern Recognition International Conference (INAPR), 2018 Crossref	38 words — 1%
6	XiaoKang Xu, Ran Tao, Xiangyang Feng, Ming Zhu. "Chapter 27 A Lightweight Facial Expression Recognition Network Based on Dense Connections", Springer Science and Business Media LLC, 2022	38 words — 1%

Crossref

7	www.sciencebuddies.org Internet	29 words — 1%
8	"Computer Vision – ECCV 2016 Workshops", Springer Nature, 2016 Crossref	28 words — 1%
9	"Computer Vision – ACCV 2016", Springer Science and Business Media LLC, 2017 Crossref	23 words — 1%
10	Xuemei Wu, Jie He, Changqin Huang, Qionghao Huang, Jia Zhu, Xiaodi Huang, Hamido Fujita. "FER-CHC: Facial expression recognition with cross-hierarchy contrast", Applied Soft Computing, 2023 Crossref	23 words — 1%
11	"Table of contents", 2019 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM), 2019 Crossref	21 words — 1%
12	Emad Barsoum, Cha Zhang, Cristian Canton Ferrer, Zhengyou Zhang. "Training deep networks for facial expression recognition with crowd-sourced label distribution", Proceedings of the 18th ACM International Conference on Multimodal Interaction - ICMI 2016, 2016 Crossref	21 words — 1%
13	paperswithcode.com Internet	20 words — 1%

EXCLUDE QUOTES OFF EXCLUDE SOURCES < 1%