¹Akhmetkan A.,
Master's student, ORCID ID: 0009-0009-9353-7416,
e-mail: az_akhmetkan@kbtu.kz
²*Mutaliyev Ye.,
PhD student, ORCID ID: 0000-0002-1755-8161,
*e-mail: emutaliev11@gmail.com

¹Kazakh-British Technical University,
Almaty, Kazakhstan
²SDU University,
Kaskelen, Kazakhstan

DEVELOPMENT OF RESIDUAL CNN ARCHITECTURE FOR FACIAL EXPRESSION RECOGNITION

Abstract

The research introduces a deep neural network system which achieves multi-class emotion classification through its development process. The system identifies seven emotional states through its classification system which includes angry, disgust, fear, happy, neutral, sad and surprise. The researchers divided their dataset into training and testing parts after preprocessing and they used precision and recall and F1-score and confusion matrix and ROC-AUC curves to evaluate their results. The model achieves its highest accuracy when detecting happy emotions at 89% followed by surprise at 68% and disgust at 49% according to the confusion matrix. The model achieves good to excellent classification results for most emotions yet it struggles with "fear and neutral emotions because their features overlap or their class distributions are unbalanced. The researchers computed Receiver Operating Characteristic (ROC) curves and Area Under Curve (AUC) values for each class in the study. The model produced its best AUC results for happy and surprise emotions at 0.92 and 0.90 respectively followed by disgust at 0.84. The lowest AUC score of 0.71 appeared in the fear category because this emotion showed weak discriminative properties. The model achieved a macro-averaged AUC score of 0.82 when evaluating all classes together. The proposed neural network shows strong performance in emotion recognition tasks through its ability to detect intense emotions such as happiness and surprise.

Keywords: emotion recognition, deep learning, residual convolutional neural network, multi-class classification, facial expression analysis.

Introduction

Facial Expression Recognition (FER) systems play a key role in affective computing and psychological assessment, and Convolutional Neural Networks (CNNs) have become the standard approach for automated facial expression recognition, outperforming traditional methods in applications such as driver monitoring and surveillance [1, 2]. The widespread adoption of deep learning technology has transformed traditional handcrafted approaches [6]. The identification of facial expressions through FER remains difficult because of the wide range of expressions within each class and the occurrence of occlusions and illumination variations and small expression variations [7].

The majority of CNN-based architectures face two main problems because they either have too few layers which results in underfitting or they have too many layers which causes overfitting and poor computational performance [9, 10]. The existing models that use general image classification

architectures fail to extract the specific spatial details which are vital for distinguishing facial expressions in low-resolution grayscale FER2013 images [11–14].

Our research presents a new deep residual convolutional neural network architecture which specializes in facial expression recognition tasks. The proposed model uses lightweight components with residual learning blocks to enable better gradient flow and feature reuse. The residual block structure includes two convolutional layers with batch normalization and ReLU activation followed by a projection shortcut that maintains compatibility when channel dimensions differ. The model uses global average pooling to reduce spatial dimensions and prevent overfitting and dropout regularization runs throughout the model to boost generalization performance.

The main aim of this research involves creating and assessing a new CNN system which unites deep residual learning with dropout regularization and global average pooling to achieve efficient facial emotion recognition from grayscale images. The proposed model aims to enhance both classification precision and model adaptability while minimizing the chance of model overfitting and computational resource usage.

Objectives:

1. The research aims to create a new residual convolutional neural network structure which combines identity shortcuts with batch normalization and dropout and global average pooling for improved facial emotion recognition efficiency and robustness.

2. The proposed model requires implementation and training with the dataset through proper data preprocessing and optimization methods to maximize learning results.

3. The evaluation process assesses the proposed architecture through performance tests which measure classification precision and generalization strength and computational speed.

The main contribution of this research involves creating and testing an efficient residual CNN model for FER applications. Our architecture differs from other methods because it starts from scratch to optimize performance for low-resolution grayscale facial expression datasets. The architecture achieves its unique value through its carefully designed residual block structure which strikes an optimal balance between model depth and complexity for both limited resource environments and real-time usage.

The research presents a complete assessment of the architecture's performance on recognized benchmarks which shows better results than standard models. The proposed model strikes an optimal balance between performance quality and computational requirements which establishes it as a new reference point for FER researchers.

The rising need for emotion-based systems makes robust facial emotion recognition (FER) an essential research area within computer vision and artificial intelligence. The development of architectures which balance low complexity with effective facial expression feature extraction remains an essential unmet need. Current research either fails to optimize architectures or depends on pre-trained models which do not match the specific characteristics of facial data. The research fills this knowledge gap through its domain-focused architecture which delivers top performance while using a small number of model parameters.

The field of computer vision faces a major problem with FER because human facial expressions show wide variability and lighting conditions and occlusions and low inter-class variance. The current research direction in CNN development focuses on architectural improvements and regularization methods which work to boost generalization and reduce overfitting while maintaining computational efficiency.

The recent FER research field adopted deep residual learning as a major approach through identity shortcut connections which enable deep network gradient propagation without performance loss. The authors Pham et al. develop an innovative method to boost CNN performance for FER applications through an attention-based masking system [15]. The Residual Masking Network (RMN) combines the benefits of Deep Residual Networks with U-Net–like architecture design. The model uses a segmentation-based masking module to enhance feature maps and steer the model toward expression-specific facial areas while eliminating unimportant features. The network achieves better decision-making through its ability to focus on discriminative face areas. The RMN achieves top performance

on both FER2013 and VEMO private dataset through its attention and segmentation mechanisms which enhance FER accuracy. The research demonstrates how attention-based feature refinement through architectural innovations leads to better performance in FER systems than traditional CNN and residual network approaches.

The development of facial expression recognition has advanced through architectural improvements which focus on extracting meaningful features while eliminating useless information. The ResEmoteNet presents a deep learning system which unites Convolutional and Residual and Squeeze-and-Excitation (SE) blocks to boost facial emotion recognition through better feature extraction [16]. The SE mechanism functions as a vital component because it adjusts channel-wise feature responses adaptively to highlight essential facial areas while eliminating unnecessary information. The model achieves deep feature extraction through residual blocks which prevents performance deterioration while learning complex facial patterns across different emotional expressions. The model achieved 79.79% accuracy on FER2013 and 94.76% on RAF-DB and 72.39% on AffectNet-7 and 75.67% on ExpW during its evaluation on these four benchmark datasets. The results prove ResEmoteNet stands as a top performer which outperforms multiple current FER models through its combination of residual learning with attention-based feature recalibration.

Facial expressions contain delicate emotional signals which prove difficult to detect because of environmental changes and differences within each class and hidden facial parts. The authors in introduce an end-to-end system which employs residual blocks to achieve robust facial emotion recognition [17]. The system directly transforms facial images into emotional outputs through residual learning to maintain vital information and prevent network degradation at deeper levels. The model achieved 75% accuracy when tested on FERGIT which extends FER2013 with 49,300 images while handling difficult data conditions including unbalanced data and hidden facial parts. The approach reached 97% accuracy when tested on the CK+ dataset which shows its ability to perform well in controlled environments with properly labeled expressions. The research proves that residual architectures excel at processing both extensive complex datasets and smaller benchmark collections within the FER domain.

The majority of previous research dedicated to emotion recognition has concentrated on achieving maximum classification accuracy but disregarded the specific performance differences between individual emotions. The classification of fear and neutral emotions proves difficult because their features overlap and the data distribution is unbalanced and the expression signals are faint. The recognition gap between low-salience and ambiguous emotions presents a major challenge because it affects the practical applications of FER systems in mental health tracking and human-computer interface adaptation.

We present deep neural network solution for seven-class emotion recognition that targets the universal emotions of angry, disgust, fear, happy, neutral, sad and surprise. The research uses precision and recall and F1-score and ROC curves and confusion matrices as evaluation metrics to provide detailed performance analysis for each emotion category. The evaluation method enables researchers to understand both the model's successful areas and its weak points in different emotional categories. The research demonstrates high performance in detecting happy and surprise emotions but reveals difficulties in identifying fear and neutral expressions which adds new evidence about FER model weaknesses. The study demonstrates deep networks' effectiveness for FER while pinpointing essential research paths that need development for dealing with class imbalances and overlapping features and faint expression detection.

Materials and methods

The section provides details about preprocessing operations together with applied methods and dataset organization. The FER (Facial Expression Recognition) dataset from Kaggle contains 35,887 grayscale images [18]. The dataset contains seven emotional categories which include anger and contempt and fear and joy and neutral and sadness and surprise. The dataset contains randomly selected faces of ten different people from the dataset.

Figure 1 – Sample of dataset

The research used deep learning techniques to create and test a CNN model with residual connections for identifying multiple emotional states. The research process included data preparation followed by model design and training and final performance assessment.

A. Dataset preparation

The dataset included grayscale facial pictures which researchers grouped into seven emotional categories: angry, disgust, fear, happy, neutral, sad and surprise. The researchers standardized all images to 48×48 pixels while converting them to single-channel grayscale format for better computational efficiency. The dataset received a training and validation split for accurate performance assessment. The Keras library provided ImageDataGenerator to handle image processing and batch management. The training generator used random shuffling to boost generalization but the validation generator operated deterministically for consistent evaluation. The batch size remained constant at 256.

B. Model architecture

The Keras functional API enabled the development of a custom deep residual CNN architecture. The network started with a convolutional layer before adding three residual blocks which used filter sizes that grew progressively from 64 to 128 to 256. The residual blocks contained two convolutional layers with batch normalization and ReLU activation which used shortcut connections to enable gradient flow and prevent vanishing gradients. The network included max pooling and dropout layers after each block to prevent overfitting. The feature maps underwent global average pooling before passing through a dense layer with 128 units and dropout activation. The output layer consisted of seven neurons with softmax activation which represented the different emotion classes. The model used Adam optimizer with 0.0001 learning rate and categorical cross-entropy as its loss function.

C. Training procedure

The model underwent 15 epochs of training with early checkpointing to save the best validation accuracy weights. The training process used mini-batches to perform steps per epoch and validation steps based on dataset size and batch size.

D. Evaluation Metrics

The evaluation of model performance included multiple assessment metrics which worked together as a system. The evaluation used standard classification metrics to calculate overall accuracy and precision and recall and F1-score at both micro and weighted average levels. The normalized confusion matrix enabled researchers to study specific classification errors which occurred in each class. The evaluation of discriminative ability for each emotion class used Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) scores through one-vs-rest strategies. The macro-averaged AUC score served as an overall performance metric for classification tasks.

Results and discussion

The proposed lightweight convolutional neural network with residual connections (Residual CNN) achieved evaluation for facial emotion recognition tasks. The training and validation

performance of the model appears in Figure 2 through loss function and classification accuracy graphs spanning 15 training epochs.

The model learned successfully because the training loss function decreased progressively throughout the training process. The validation loss values demonstrated unstable behavior because they first decreased before starting to fluctuate which indicates possible overfitting.

The classification accuracy follows the same pattern as the loss function. The training accuracy reached 51% at the end of training but the validation accuracy demonstrated unpredictable behavior with major changes during the last training epochs. The model fails to achieve good generalization results because it received training from a restricted dataset.

The model achieves high performance on training data but demonstrates weak ability to predict validation set results. The upcoming research will concentrate on three main objectives which include implementing Dropout and L2 regularization techniques and implementing early stopping and data augmentation for training dataset expansion.
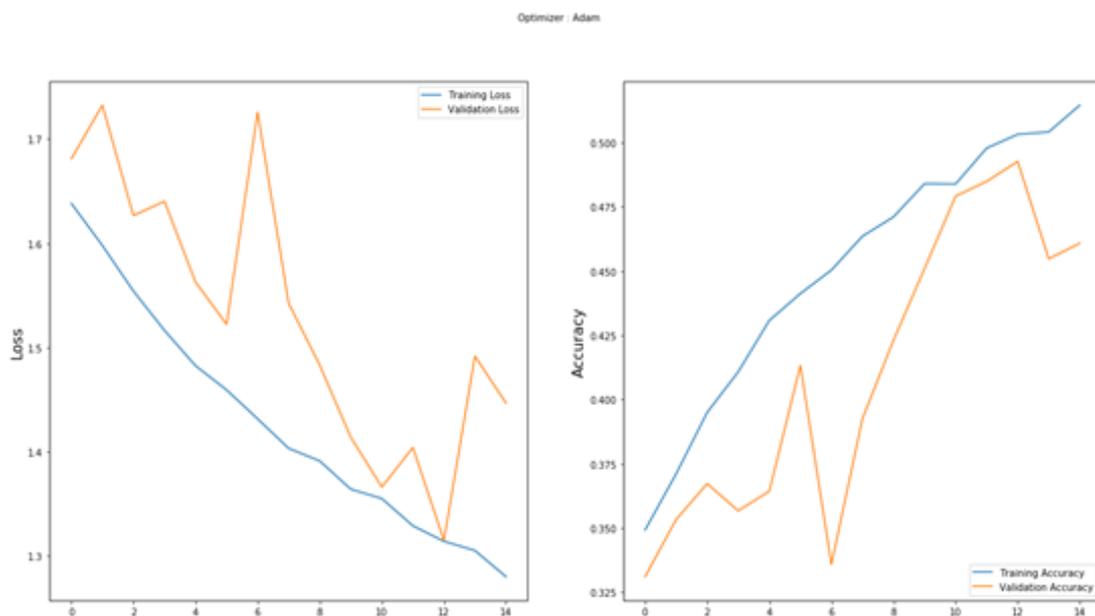

Figure 2 – Training and validation loss and accuracy curves over 15 epochs

The researchers used a Normalized confusion matrix to evaluate the classification results for each emotion category. The results appear in Figure 3. The model shows its best performance in detecting happy emotions because it correctly identifies 89% of these samples. The model demonstrates strong performance in detecting sad and surprise emotions with recognition rates of 68% and 40% respectively. The model learns effectively from these classes because they contain more pronounced visual characteristics. The model faces challenges when trying to identify disgust emotions together with fear and neutral emotions. The model identifies disgust emotions incorrectly as sad in 49% of cases and as happy in 32% of cases. The fear emotion gets misclassified as sad in 46% of instances and as neutral in 24% of instances. The neutral class shows a similar pattern of misclassification because it gets identified as happy in 35% of cases and as sad in 37% of cases. The model correctly identifies anger in only in 14% of cases but incorrectly identifies it as sadness in 47% of instances. The model demonstrates superior performance in detecting strong emotions but struggles with identifying emotions that have weak visual characteristics or share similar visual features. The results demonstrate a need to enhance the model through class balancing techniques and additional training data for underrepresented emotions and through the implementation of more sensitive architectures and attention mechanisms.
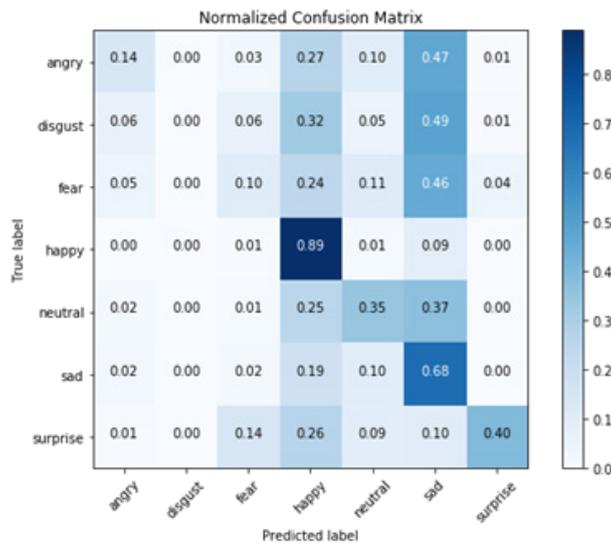
Figure 3 – Confusion matrix for each class

The model received an additional evaluation through one-vs-all ROC curve construction for each emotion class to determine its performance in distinguishing between specific emotional states. The AUC values from Figure 4 show the best performance for happy emotions with 0.92 while surprise and disgust emotions achieved high values of 0.90 and 0.84 respectively. The model demonstrates strong confidence in identifying surprise and disgust expressions because of their high AUC values. The results from the confusion matrix analysis match the findings of this study. The model achieves moderate success in identifying angry and neutral and sad emotions through AUC values between 0.78 and 0.80. The model shows its lowest performance on fear emotion detection because its AUC score reaches only 0.71. The model shows a tendency to mistake fear expressions for sad or neutral expressions. The model achieves its best performance with clear and expressive emotions yet it struggles with identifying facial expressions that are ambiguous or share visual similarities. The model's performance suffers from two main factors: the unclear nature of facial expressions and the restricted availability of training data for specific emotions. The upcoming research will concentrate on developing better discriminative abilities through improved training of challenging emotions and the implementation of more sensitive architectural elements.
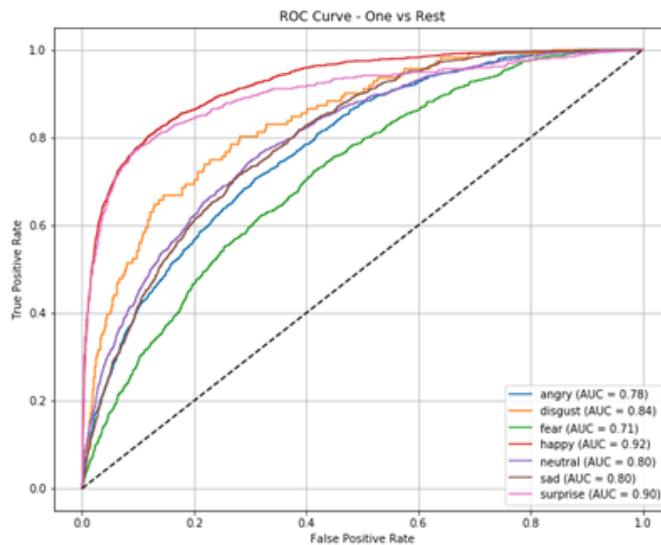

Figure 4 – ROC curve for each class

The classification report in Table 1 demonstrates that the happy class achieved the highest performance through its 0.56 precision and 0.89 recall and 0.69 F1-score while using the most examples from the dataset (1825). The results confirm that the model performs best when detecting emotions that are clearly positive. The results for all emotions except happy show unsatisfactory performance. The angry class demonstrates high precision at 0.53 yet its recall rate remains at 0.14 which indicates that the model frequently misses many instances. The fear and disgust classes demonstrate the worst performance in all evaluation metrics because the model completely failed to identify any disgust instances. The model failed to recognize disgust instances because the training data for this class consisted of only 111 examples which created a severe class imbalance. The F1-scores for sad and neutral emotions reach 0.43 and 0.35 respectively. The model demonstrates high precision for surprise classification at 0.84 but its recall rate of 0.40 indicates it only labels instances when confident while missing various less intense cases.

The macro-average results demonstrate that the model performs differently for each class because precision equals 0.44 and recall equals 0.36 and F1-score equals 0.35. The weighted average scores which consider class sample sizes result in precision at 0.50 and recall at 0.48 and F1-score at 0.43.

Table 1 – Classification report for each class

Emotion	Precision	Recall	F1-Score	Support
Angry	0.53	0.14	0.22	960
Fear	0.35	0.10	0.16	1018
Happy	0.56	0.89	0.69	1825
Neutral	0.50	0.35	0.41	1216
Sad	0.32	0.68	0.43	1139
Surprise	0.84	0.40	0.54	797

The model shows better performance at detecting happy and surprise emotions because these classes receive proper representation and visual distinction yet it fails to identify fear and disgust emotions which are less common in the dataset. The results confirm that the model requires better class distribution and additional training data and architectural improvements. Our model achieves moderate results but fails to match the Residual Masking Network (RMN) performance on FER2013 because it lacks the attention mechanism's ability to precisely locate facial expressions [15]. The ResEmoteNet architecture achieved better results than baseline models on various datasets through Squeeze-and-Excitation (SE) blocks which demonstrated how attention-based modules enhance the detection of rare emotions [16]. The results indicate that residual learning provides strong resistance because it maintains stable performance between sad (F1 = 0.43) and neutral (F1 = 0.41) yet the model fails to achieve optimal results when dealing with uncertain emotions like fear.

Conclusion

This study showcased implemented and tested deep residual convolutional neural network to identify multiple emotions from seven categories which included angry, disgust, fear, happy, neutral, sad and surprise. The proposed model shows strong performance in emotion recognition tasks because it achieves high accuracy rates for detecting happy and surprise emotions and moderate accuracy for other emotions. The quantitative evaluation through precision and recall measurements and F1-score calculations and confusion matrix analysis and ROC-AUC results showed the model's stability with a macro-averaged AUC score of 0.82.

The model shows promising performance but it faces difficulties when identifying fear and neutral emotions because their facial characteristics overlap and their class distribution might be unbalanced. The model successfully detects various emotional patterns yet its performance could improve through data augmentation methods and class balancing techniques and attention mechanism integration. The model shows reliable detection of fundamental emotions which makes it suitable for human-computer interaction and affective computing and mental health monitoring systems that require precise emotion detection.

REFERENCES

1 Ekundayo, O.S., and Viriri, S. Facial expression recognition: A review of trends and techniques. IEEE Access, 9, 136944–136973 (2021). https://doi.org/10.1109/ACCESS.2021.3113464

2 Sajjad, M., Ullah, F.U.M., Ullah, M., Christodoulou, G., Cheikh, F.A., Hijji, M., and Rodrigues, J.J. A comprehensive survey on deep facial expression recognition: challenges, applications, and future guidelines. Alexandria Engineering Journal, 68, 817–840 (2023). https://doi.org/10.1016/j.aej.2023.01.017

3 Zholshiyeva, L., Zhukabayeba, T., Serek, A.G., Duisenbek, R., Berdieva, M., and Shapay, N. Deep learning-based continuous sign language recognition. Journal of Robotics and Control (JRC), 6 (3), 1106–1118 (2025). https://doi.org/10.18196/jrc.v6i3.25881

4 Zholshiyeva, L., Zhukabayeva, T., Baumuratova, D., and Serek, A.G. Design of QazSL sign language recognition system for physically impaired individuals. Journal of Robotics and Control (JRC), 6 (1), 191–201 (2025). https://doi.org/10.18196/jrc.v6i1.23879

5 Serek, A., Amirgaliyev, B., Li, R.Y.M., Zhumadillayeva, A., and Yedilkhan, D. Crowd density estimation using enhanced multi-column convolutional neural network and adaptive collation. IEEE Access (2025). https://doi.org/10.1109/ACCESS.2025.3597393

6 Singh, R., Saurav, S., Kumar, T., Saini, R., Vohra, A., and Singh, S. Facial expression recognition in videos using hybrid CNN & ConvLSTM. International Journal of Information Technology, 15 (4), 1819–1830 (2023). https://doi.org/10.1007/s41870-023-01183-0

7 Pise, A.A., Alqahtani, M.A., Verma, P., K., P., Karras, D.A., S., P., and Halifa, A. Methods for facial expression recognition with applications in challenging situations. Computational Intelligence and Neuroscience, 2022 (1), 9261438 (2022). https://doi.org/10.1155/2022/9261438

8 Ge, H., Zhu, Z., Dai, Y., Wang, B., and Wu, X. Facial expression recognition based on deep learning. Computer Methods and Programs in Biomedicine, 215, 106621 (2022). https://doi.org/10.1016/j.cmpb.2022.106621

9 Dishar, H.K., and Muhammed, L.A. A review of the overfitting problem in convolution neural network and remedy approaches. Journal of Al-Qadisiyah for Computer Science and Mathematics, 15 (2), 155 (2023). https://doi.org/10.29304/jqcm.2023.15.2.1240

10 Bejani, M.M., and Ghatee, M. A systematic review on overfitting control in shallow and deep neural networks. Artificial Intelligence Review, 54 (8), 6391–6438 (2021).

11 Salehi, A.W., Khan, S., Gupta, G., Alabduallah, B.I., Almjally, A., Alsolai, H., et al. A study of CNN and transfer learning in medical imaging: advantages, challenges, future scope. Sustainability, 15 (7), 5930 (2023). https://doi.org/10.3390/su15075930

12 Gupta, J., Pathak, S., and Kumar, G. Deep learning (CNN) and transfer learning: a review. Journal of Physics: Conference Series, 2273 (1), 012029 (2022).

13 Nan, F., Jing, W., Tian, F., Zhang, J., Chao, K.M., Hong, Z., and Zheng, Q. Feature super-resolution based facial expression recognition for multi-scale low-resolution images. Knowledge-Based Systems, 236, 107678 (2022). https://doi.org/10.1016/j.knosys.2021.107678

14 Lo, L., Ruan, B.K., Shuai, H.H., and Cheng, W.H. Modeling uncertainty for low-resolution facial expression recognition. IEEE Transactions on Affective Computing, 15 (1), 198–209 (2023). https://doi.org/10.1109/TAFFC.2023.3264719

15 Pham, L., Vu, T.H., and Tran, T.A. Facial expression recognition using residual masking network. Proc. 25th Int. Conf. Pattern Recognition (ICPR), 4513–4519 (2021). https://doi.org/10.1109/ICPR48806.2021.9411919

16 Roy, A.K., Kathania, H.K., Sharma, A., Dey, A., and Ansari, M.S.A. ResEmoteNet: bridging accuracy and loss reduction in facial emotion recognition. IEEE Signal Processing Letters (2024). https://doi.org/10.1109/LSP.2024.3521321

17 Bah, I., and Xue, Y. Facial expression recognition using adapted residual based deep neural network. Intelligence & Robotics, 2 (1), 72–88 (2022). https://doi.org/10.20517/ir.2021.16

18 msambare. fer2013 (Facial Expression Recognition 2013 Dataset). Kaggle. URL: https://www.kaggle.com/datasets/msambare/fer2013 (accessed: 26.09. 2025).

¹Ахметжан А.,
магистрант, ORCID ID: 0009-0009-9353-7416,
e-mail: az_akhmetkan@kbtu.kz
²*Муталиев Е.,
докторант, ORCID ID: 0000-0002-1755-8161,
*e-mail: emutaliev11@gmail.com

¹Қазақстан-Британ техникалық университеті,
Алматы қ., Қазақстан
²SDU University,
Қаскелең қ., Қазақстан

БЕТ МИМИКАЛАРЫН ТАНУҒА АРНАЛҒАН RESIDUAL CNN АРХИТЕКТУРАСЫН ЖЕТІЛДІРУ

Аңдатпа

Бұл зерттеуде эмоцияларды көпкластық жіктеуді жүзеге асыратын терең нейрондық желі жүйесі ұсынылады. Жүйе жеті түрлі эмоциялық күйді анықтайды: ашу (angry), жиіркену (disgust), қорқыныш (fear), қуаныш (happy), бейтарап күй (neutral), мұң (sad) және таңдану (surprise). Зерттеушілер деректер жиынтығын алдын ала өңдеуден өткізгеннен кейін оны оқыту және тестілеу бөліктеріне бөліп, нәтижелерді бағалау үшін дәлдік (precision), толықтық (recall), F1-нәтижесі, қателер матрицасы (confusion matrix) және ROC-AUC қисықтарын қолданды. Қателер матрицасының нәтижелері бойынша модель ең жоғары дәлдікті «қуаныш» эмоциясын тануда көрсетті (89%). Одан кейін «таңдану» (68%) және «жиіркену» (49%) эмоциялары орналасқан. Жалпы алғанда, модель эмоциялардың басым бөлігі бойынша жақсы және өте жақсы нәтижелер көрсетті. Алайда «қорқыныш» және «бейтарап күй» эмоцияларын тануда белгілі бір қиындықтар байқалды, себебі олардың белгілері бір-бірімен ұқсас немесе деректердегі сыныптардың таралуы теңгерімсіз болады. Зерттеушілер әрбір сынып үшін қабылдағыштың операциялық сипаттамаларының (ROC) қисықтарын және қисық астындағы ауданды (AUC) есептеді. Модель ең жоғары AUC нәтижелерін «қуаныш» (0,92) және «таңдану» (0,90) эмоциялары үшін көрсетті, одан кейін «жиіркену» (0,84) эмоциясы орналасқан. Ең төменгі AUC көрсеткіші – 0,71 «қорқыныш» санаты үшін тіркелді, бұл эмоцияның айқын белгілерінің әлсіздігімен түсіндіріледі. Барлық сыныптарды бірге бағалағанда модель макроорташа AUC көрсеткіші 0,82 нәтижесіне қол жеткізді. Ұсынылған нейрондық желі бет-әлпет мимикасын талдау негізінде эмоцияларды тану міндеттерінде жоғары тиімділікті көрсетті, әсіресе қуаныш және таңдану сияқты айқын эмоцияларды анықтауда нәтижелі болды.

Тірек сөздер: эмоцияларды тану, терең оқыту, қалдықты конволюциялық нейрондық желі, көпсанатты жіктеу, бет әлпетін талдау.

¹Ахметжан А.,
магистрант, ORCID ID: 0009-0009-9353-7416,
e-mail: az_akhmetkan@kbtu.kz
²*Муталиев Е.,
докторант, ORCID ID: 0000-0002-1755-8161,
*e-mail: emutaliev11@gmail.com

¹Казахстанско-Британский технический университет, г. Алматы, Казахстан
²Университет СДУ, г. Каскелен, Казахстан

РАЗРАБОТКА RESIDUAL CNN-АРХИТЕКТУРЫ
ДЛЯ РАСПОЗНАВАНИЯ ВЫРАЖЕНИЙ ЛИЦА

Аннотация

Данное исследование представляет систему глубокой нейронной сети, которая обеспечивает многоклассовую классификацию эмоций в процессе разработки. Система определяет семь эмоциональных состояний: гнев (angry), отвращение (disgust), страх (fear), радость (happy), нейтральное состояние (neutral), грусть (sad) и удивление (surprise). Исследователи разделили свой набор данных на обучающую и тестовую части после предварительной обработки и использовали точность (precision), полноту (recall), F1-меру, матрицу ошибок (confusion matrix) и кривые ROC-AUC для оценки результатов. Согласно матрице ошибок, модель достигает наибольшей точности при распознавании радости (89%), далее следуют удивление (68%) и отвращение (49%). Модель показала хорошие и отличные результаты по большинству эмоций, однако испытывает трудности с эмоциями «страх» и «нейтральное состояние», поскольку их признаки пересекаются или распределение классов несбалансированно. Исследователи вычислили кривые операционных характеристик приемника (ROC) и площади под кривой (AUC) для каждого класса. Модель показала лучшие AUC-результаты для радости (0,92) и удивления (0,90), далее для отвращения (0,84). Наименьший AUC-показатель – 0,71 был зафиксирован для категории «страх» из-за слабой различимости этой эмоции. При оценке всех классов вместе модель достигла макроусредненного AUC-показателя 0,82. Предложенная нейронная сеть демонстрирует высокую эффективность в задачах распознавания эмоций благодаря своей способности выявлять интенсивные эмоции, такие как радость и удивление.

Ключевые слова: распознавание эмоций, глубокое обучение, остаточная сверточная нейронная сеть, многоклассовая классификация, анализ мимики лица.