© DEC 2023 | IRE Journals | Volume 7 Issue 6 | ISSN: 2456-8880

Facial Emotion Recognition of Cat Breeds by Using Convolution Neural Network

POONAM JAIN[1], AMIT KUMAR PANDEY[2], VIKAS MANOJ KUMAR PANDEY[3], BIPIN YADAV[4]

[1] *Assistant Professor, Department of IT, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai, Maharashtra, India*

[2, 3, 4] *PG Student, Department of IT, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai, Maharashtra, India*

Abstract- Facial expressions play a pivotal role in decoding the emotional states of animals, and this research delves into the domain of cat facial emotion recognition utilizing Convolutional Neural Network (CNN) algorithms. Drawing inspiration from analogous studies on dog facial expressions, our investigation focuses on the intricate task of detecting and classifying emotional expressions in cats. Leveraging a diverse dataset encompassing various cat breeds and emotional states, the CNN model demonstrates its efficacy with minimal preprocessing. The training process involves careful dataset augmentation and optimizer fine-tuning, ensuring the model's ability to generalize effectively. The results exhibit the model's proficiency in distinguishing between different emotional states in cats, presenting a promising avenue for further exploration in feline behavior analysis. This research contributes to the growing field of animal emotion recognition, shedding light on the subtleties of cat facial expressions through the lens of advanced machine learning techniques.

Indexed Terms- Cat Facial Expressions, Convolution Neural Network, Kerasmodule, Emotion Detection.

I. INTRODUCTION

Understanding the emotional states of animals is a fundamental aspect of fostering meaningful human-animal interactions. Among the vast array of non-verbal communication cues exhibited by animals, facial expressions stand out as crucial indicators of their emotional well-being. While much attention has been dedicated to decoding human facial expressions, the realm of animal facial expressions, particularly in domesticated companions like cats, remains a relatively unexplored frontier.

This research endeavors to bridge this gap by leveraging the advancements in computer vision and machine learning, specifically Convolutional Neural Networks (CNNs), to unravel the intricate language embedded in the facial expressions of cats. Our exploration is inspired by the success of similar studies in understanding dog facial expressions, acknowledging the nuanced and diverse ways in which different species communicate their emotions.
Cats, known for their enigmatic and independent nature, exhibit a rich tapestry of facial expressions that convey a spectrum of emotions. From contentment to curiosity, and from affection to displeasure, decoding these expressions can provide profound insights into the emotional well-being of our feline companions. The goal of this research is to develop a robust facial emotion recognition system tailored to the unique features of cat faces, thereby contributing to the burgeoning field of animal behavior analysis.

In this introductory section, we will outline the motivation behind the study, provide an overview of relevant literature, and delineate the methodology employed in our pursuit of unveiling the emotional landscape encapsulated in cat facial expressions. Through this endeavor, we aspire to enhance our understanding of the intricate non-verbal communication exhibited by cats, fostering improved connections between humans and their feline counterparts.

II. LITERATURE REVIEW

Fuzail Khan in his research suggested a framework to classify a person's facial ex-pressions. Any one of the six universal emotions or the neutral feelings may be represented by these classifiable phrases. Following the first face localization, the fiducial features of the eyebrows, eyes, nose, and lips are identified using facial landmark detection and feature extraction techniques. Modern face landmark detection algorithms, as well as more conventional edge and corner point recognition techniques, such as Sobel filters and Shi Tomasi corner point detection methods, are generally used to do this. In order to categorize, this results in the development of input feature vectors that are formed using Euclidean distances and trained into a Multi-Layer Perceptron (MLP) neural network. The outcomes also addressed the more uniform display of some emotions and the fundamentally subjective character of expression [1]. Results showed that cats were able to cross-modally match footage of emotional faces with their connected vocalizations', notably for emotions of high intensity. You may have read many researchers' papers related to facial expressions, so most people have done more research on things like human or animal behaviours, emotions, breeds, and so on. The difference in my research is that, along with facial expression, I will try to detect the emotions and behaviours of animals as well, using convolution neural networks (CNN). Facial expression projects are the internal emotions of dogs. The present study is aimed toward investigating cats' spontaneous ability to match acoustic and visual signals for the popularity of each conspecific and human emotions. Completely different conspecific (cat "purr" and "hiss") and hetero specific (human "happiness" and "anger") emotional stimuli were given to the tested population employing a sentience paradigm.[2]

Facial Expression Recognition is currently a very active research topic in the fields of computer vision, pattern recognition, artificial intelligence, and human-computer interaction, including human emotion analysis and image indexing, etc. There could also be some redundant or irrelevant options in feature sets so as to get rid of those redundant/irrelevant options that don't have any vital impact on the classification method, we have a tendency to propose a feature choice (FS) technique referred to as the supervised filter harmony search formula (SFHSA) supported trigonometric function similarity and minimal-redundancy maximal-relevance (MRMR) Trigonometric function similarity aims to get rid of similar options from feature vectors, whereas MRMR was accustomed verify the feasibleness of the optimum feature subsets through Pearson's parametric statistics(PCC). This favours' the options that have 1 lower correlation values with alternative features as well as higher correlation values with the facial feature categories. The formula was evaluated on 2 benchmark FER datasets, particularly the Radboud Face Information (Ra FD) and also the Japanese Feminine Facial Feature (JAFFE). 5 totally different progressive feature descriptors as well as uniform native binary pattern (ULBP), horizontal–vertical neighbourhood native binary pat-tern (hvnLBP), Dennis Gabor filters, bar chart of orientated gradients (HOG) and pointed HOG (PHOG) were thought of for FS.[3]

Facial recognition could be a major challenge within the field of laptop vision. Here we've enforced numerous biometric identification algorithms like LBPH, Eigenface, and Fisher Face. The Haar cascade has been utilized for face recognition. We have a tendency to train the algorithms with identical knowledge sets and have gotten some insights, from that. We've tried to spot that the formula offers the U.S. the most effective results. Different algorithms are compared and their workings are mentioned. At the end, tabular comparisons are provided, so it might be easier to grasp the distinction between algorithms.[4]

The ability to acknowledge facial expressions mechanically permits novel applications in human-computer interaction and different areas. Consequently, there has been active analysis in this field, with many recent works utilizing convolutional neural networks (CNNs) for feature extraction and abstract thought. These two works differ considerably in terms of CNN architecture and different factors. supported the reportable results alone, the performance impact of those factors is unclear. In this paper, we tend to review the state of the art in image-based countenance recognition exploitation of CNNs and highlight recursive variations and their performance impact. On this basis, we tend to determine existing bottlenecks and consequently directions for advancing this analysis field. Moreover, we tend to demonstrate that over-coming one of these bottlenecks—the relatively basic architectures of the CNNs used in this field-results in a considerable performance increase. By forming an Associate in Nursing ensemble of recent deep CNNs, we tend to get a FER2013 check accuracy of seventy-five.2%, outperforming previous works while not requiring auxiliary coaching information or facing registration.[5]

Some researchers' papers propose a method of neural network-based drowsiness detection with eyes open using power spectrum analysis and auto-regressive model-ling. The transition of facial expression recognition from laboratory-controlled to in-wild conditions and the recent success of deep learning in various fields, deep neural networks have increasingly been used to learn discriminative representations for automatic facial expression recognition. The speedy advances in machine learning (ML) and data fusion have created the potential to endow machines with the power of feeling, understanding, recognition, and analysis. Feeling recognition has attracted progressively intense interest from researchers from various fields. Human emotions are recognized through facial expressions, speech, behaviour (gesture/posture), or physiological signals. However, the primary 3 strategies are ineffective since humans might involuntarily or deliberately conceal their real emotions (so-called social masking). The utilization of physiological signals will result in additional objective and reliable feelings recognized. [6]

In that, the researchers provide a comprehensive review on deep facial expression recognition, including datasets and algorithms that provide insights into this problem. Then they describe the standard pipeline of a deep facial expression recognition system and their evaluation principles. Competitive performance and experimental comparisons on widely used benchmarks are also summarized. To achieve optimum results, the system employs a ranked recognition strategy. In these settings, ex-pressions are divided into 3 classes: supporting elements of the face that contribute most toward associate-degree expression. At the primary level, SWLDA and HCRF are utilized to acknowledge the expression category; whereas, at the second level, the label for the expression class is employed, employing a separate set of SWLDA and HCRF, trained only for that class. Four publicly available data sets were used in four separate experiments to validate the system. The weighted average recognition rate for the intended FER method across the four independent data sets was 96.37%, which is a significant improvement over current FER methods. [7]

We created just one descriptor and a framework for automated and dependable countenance identification. The framework is based on early re-search of human eyesight and works well on exposure as well as spontaneous expressions. The following are the main conclusions of the study:

1. Face expressions are frequently mechanically evaluated by simulating human sensory perceptions, i.e., collecting possibilities primarily from prominent facial areas.
2. Extracted features from the planned pointed native binary pattern (PLBP) operator.[8]

This facial expression integrates the study of this behaviour with the anthropological study of communication and sociality in general. Studies of facial expression are available, but results are not typically framed in an evolutionary study of facial expression, which includes the facial expression co-ordinate, the unique context and function, human smiling is used as an ex-ample of adaptation of adaption, and testable hypotheses concerning the human smile. The persecution holistic illustration results are compared to the outcomes of facial expression recognition victimization alternatives from domain-specific areas. The intended face characteristics recognition system performed well on publicly available expanded Cohn Kanade (CK+) facial feature information sets.[9]

Deep-learning-based FER techniques were later awarded, with deep net-works sanctioning "end-to-end" learning. This review also focuses on a con-temporary hybrid deep-learning strategy that combines a convolutional neural network (CNN) for spatial options of a single frame with long-term immediate memory (LSTM) for temporal options of successive frames. In the next section of this article, a brief overview of publicly available analysis metrics is provided, followed by a comparison with benchmark outcomes, which provide a typical for a quantitative comparison of FER investigations.[10]

Researchers are thinking about the emotions and behaviours of the face in social interaction and social intelligence is widely recognized in anthropology.

III. CNN

A feed-forward neural network called a convolutional network analyzes visual pictures by processing data in a grid-like architecture. A "ConvNet" is sometimes referred to as a "ConvNet." A convolutional neural network is employed to find and categorize items in a picture. The process of removing functional elements from an image begins with this. Multiple filters work together to execute the convolution action in a convolution layer. Each image may be thought of as a matrix of pixel values. A filter matrix with a 3x3 dimension is also included. moving the filter matrix across the picture to obtain the convolved feature matrix and compute the dot product. The rectified linear unit is referred to as the ReLU. The next step is to transfer the feature maps to a ReLU layer after they have been retrieved. ReLU does an operation element-by-element, setting all the negative pixels to 0. The result is a corrected feature map, and it gives the network non-linearity. The down-sampling process of pooling lowers the feature map's dimensionality. To create a pooled feature map, the corrected feature map is now passed through a layer of pooling. predict the classes with greater accuracy. At this step, the error will be calculated and then backpropagated. The weights and detector area unit were adjusted to optimize the performance of the model. In this way, the network trains on the data. To distinguish distinct portions of the picture, such as edges, corners, bodies, feathers, eyes, and beaks, the pooling layer employs a variety of filters. Flattening is the procedure's next phase. The generated 2-dimensional arrays from pooled feature maps are flattened into a single, lengthy continuous linear vector. To categorize the picture, the flattened matrix is provided as an input to the fully linked layer.

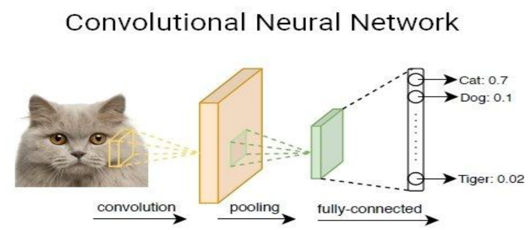
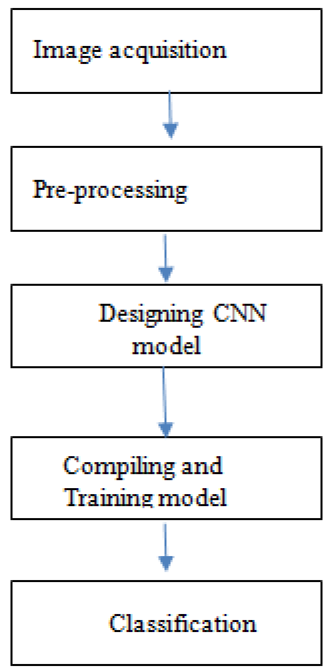


Fig. 1. Illustration of the proposed Convolutional Neural Network (CNN)

IV. FLOWCHART



V. METHODOLOGY

- Dataset

The dataset used in this work includes many images with attributes that may be used to identify real footwear. A total of 1000 photos were utilized for training, and 300 images were used for validation, with the distribution of training and validation data being 70% and 30%, respectively. There are hidden layers, an output layer, and an input layer in CNN. Convolutional, ReLU, pooling, and fully linked layers make up the hidden layers most often. Pictures are first input into the model as RGB images, and these colours combine to create a three-dimensional matrix. Next, the images are rescaled for binary classification. Since multiple-sized matrix images cannot be provided to the neural network, they are reduced in size to 200x200x3 pixels. Images are then covered in convolutional layers with maximum pooling. A 200x200x3 picture with 16 3x3 ReLu-activated filters and a 2x2 max pooling layer are applied at the first convolutional layer in order to extract the maximum number of pixels possible. Applying additional convolutional layers with 32 and 64 filters of size 3x3 with ReLu activation function over the image of size 200x200x3 and a max pooling layer of size 2x2 which increases the number of channels in the network, which improves model accuracy. The next step is flattening afterwards, which creates a single, lengthy continuous linear vector from the two-dimensional arrays of the pooled feature maps. Following the application of two dense layers with sigmoid activation functions, pictures are classified based on the results of convolutional layers.

VI. RESULTS

In this work, 1000 training photographs and 300 validation images were used to train the model. These photos were trained using the CNN model with a variety of optimizer techniques, including loss categorical cross-entropy, RMS, and an optimizer with a learning rate of 0.001. The comparison of the proposed model's accuracy and loss performance for each optimizer employed after training with 15 iterations (epochs) is shown in Table 1. The best accuracy and loss performance is provided by the RMS optimizer. In comparison to the conventional, untrainable approaches, the supervised learning model is better. Table 1 and figure 2. It contains the results of epochs and shows how the model identifies an image.

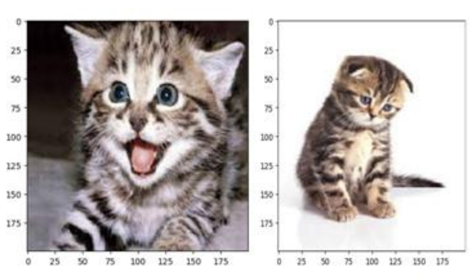


Table 1. Results depicting accuracy gained by Supervised Learning Model.

Epochs	Accuracy	Validation Accuracy
Epoch 1/10	0.5556	0.4
Epoch 1/10	0.4286	0.538
Epoch 1/10	0.7778	0.461
Epoch 1/10	0.5714	0.538
Epoch 1/10	0.4444	0.538
Epoch 1/10	0.5714	0.769
Epoch 1/10	0.8571	0.9231
Epoch 1/10	0.7143	0.5385
Epoch 1/10	0.8571	0.769
Epoch 1/10	0.9571	1.000

CONCLUSION

In wrapping up our exploration into Facial Emotion Recognition, we took a close look at various research papers, drawing insights from studies on human expressions and even cat detection. However, our primary focus has been on deciphering the facial expressions of cats using Convolution Neural Networks (CNNs), mimicking the success observed in dog breeds.

Utilizing the CNN method, coupled with data augmentation, has proven to be notably efficient, especially in the realm of image processing. Our proposed model showcases a superior approach, consistently achieving higher validation accuracy compared to existing models. Similar to our canine-focused research, we've delineated two distinct emotional classes for cats—happy and sad—as the basis for emotion classification.

The fundamental aim of this research has been to craft a robust facial expression recognition system tailored specifically for cats, employing the power of

CNNs and data augmentation. This journey not only enhances our understanding of feline emotions but also sets the stage for further refinements and applications. As we conclude this phase, the door remains wide open for future endeavors, beckoning further strides in the intricate realm of Facial Emotion Recognition for our beloved feline companions.

REFERENCES

[1] Fuzail Khan Facial Expression Recognition using Facial Landmark Detection and Feature Extraction via Neural Networks.2020 Jul 15.

[2] Quaranta A, d'Ingeo S, Amoruso R, Siniscalchi M. Emotion recognition in cats. Animals. 2020 Jun 28;10(7):1107

[3] Saha S, Ghosh M, Ghosh S, Sen S, Singh PK, Geem ZW, Sarkar R. Feature selection for facial emotion recognition using cosine similarity-based harmony search algorithm. Applied Sciences. 2020 Apr 19;10(8): 2816.S.L Happy, A. Routray, Automatic Facial Expression Recognition Using Features of Salient Facial Patches in IEEE Transactions on Affective Computing - May 2015, DOI 10.1109.

[4] Jagtap AM, Kangale V, Unune K, Gosavi P. A Study of LBPH, Eigenface, Fisherface and Haar-like features for Face recognition using OpenCV. In2019 International Conference on Intelligent Sustainable Systems (ICISS) 2019 Feb 21 (pp. 219-224). IEEE.

[5] Pramerdorfer C, Kampel M. Facial expression recognition using convolutional neural networks: state of the art. arXiv preprint arXiv:1612.02903. 2016 Dec 9.

[6] Zhang J, Yin Z, Chen P, Nichele S. Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review. Information Fusion. 2020 Jul 1; 59:103-26.

[7] Siddiqi MH, Ali R, Khan AM, Park YT, Lee S. Human facial expression recognition using stepwise linear discriminant analysis and hidden conditional random fields. IEEE Transactions on Image Processing. 2015 Feb 24;24(4):1386-98.

[8] R.A Khan, A. Meyer, H. Konik, S. Bouakaz, Framework for reliable, real-time facial expression recognition for low resolution images. Pattern Recognit, Lett. 2013, 34, 11591168.

[9] D. Ghimire; S. Jeong; J. Lee; S.H Park, Facial expression recognition based on local region-specific features and support vector machines. Multimed. Tools Appl. 2017, 76, 78037821.

[10] Ko BC. A brief review of facial emotion recognition based on visual information. Sensors. 2018 Jan 30; 18(2):401.