Cat Family Recognition by Using Convolution Neural Network

DR. SANTOSH SINGH[1], AMIT KUMAR PANDEY[2], BIPIN YADAV[3]

[1] *H.O.D(Information Technology), 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- This project presents an automated solution for the recognition of various species within the cat family through the implementation of Convolutional Neural Networks (CNNs). Leveraging the TensorFlow and Keras libraries, the study addresses the challenge of accurately classifying images depicting Cheetahs, Leopards, Lions, Pumas, and Tigers. The methodology involves the preprocessing of image data, the construction of a specialized CNN architecture, and the utilization of categorical cross-entropy loss for multiclass classification. The model is trained and evaluated on a dataset comprising diverse cat species, demonstrating its efficacy in accurately identifying the species of a given cat image. The results indicate promising performance, showcasing the potential of CNNs in automating cat species recognition, with implications for wildlife monitoring and conservation initiatives.

Indexed Terms- Cat Family, CNN, Image Classification, TensorFlow, Keras, Deep Learning, Convolutional Neural Networks, Image Data Generator.

I. INTRODUCTION

The recognition of distinct species within the cat family poses a significant challenge due to the visual intricacies shared among various species such as Cheetahs, Leopards, Lions, Pumas, and Tigers. Traditional methods of classification often fall short in handling the complexities inherent in the diverse appearances of these feline species. This project seeks to address this challenge by employing advanced deep learning techniques, specifically Convolutional Neural Networks (CNNs), to automate the process of cat species recognition.

The motivation behind this project lies in the increasing need for efficient and accurate tools for wildlife monitoring and conservation. Manual identification of cat species in large datasets of wildlife imagery is not only time-consuming but also prone to human error. The application of CNNs, known for their prowess in image classification tasks, promises a more reliable and automated solution.

The dataset used in this study comprises images of various cat species, meticulously organized into training and validation sets. The chosen architecture for the CNN includes convolutional layers to capture intricate features, max-pooling layers for spatial dimension reduction, and fully connected layers for precise classification. Through the integration of the TensorFlow and Keras libraries, this project strives to develop a model capable of not only distinguishing between different cat species but also generalizing well to unseen instances.

By automating the recognition of cat species, this project contributes to the broader field of computer vision and deep learning, offering practical applications for biodiversity monitoring, wildlife research, and conservation efforts. The subsequent sections delve into the CNN architecture, methodology, and results, showcasing the potential of this approach in revolutionizing the field of cat species recognition.

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

The Convolutional Neural Network (CNN) implemented in this project represents a sophisticated deep learning architecture meticulously designed for the nuanced task of recognizing distinct species within the cat family. Unlike traditional neural networks, CNNs are tailored to process visual data efficiently, making them ideal for image classification tasks. The architecture is composed of several layers that collectively enable the model to learn intricate hierarchical features from input images. The initial convolutional layers employ filters to convolve over the images, capturing essential patterns such as edges, textures, and shapes. These convolutional operations are augmented by rectified linear unit (ReLU) activation functions, introducing non-linearity to the model and enabling it to learn complex relationships in the data.To maintain computational efficiency and preserve critical information, max-pooling layers follow the convolutional layers, reducing the spatial dimensions of the data. This hierarchical feature extraction process is repeated through multiple sets of convolutional and max-pooling layers, allowing the network to progressively discern more abstract and complex visual features. The final layers of the CNN consist of fully connected layers that aggregate the learned features to make precise predictions about the input images. The last layer utilizes a softmax activation function, assigning probabilities to each class and facilitating the identification of the specific cat species present in the image. Implemented using the TensorFlow and Keras libraries, this CNN is purpose-built for cat species recognition, leveraging the power of deep learning to automate a task that traditionally relies on human expertise. The effectiveness of this architecture is demonstrated through rigorous training and evaluation on a diverse dataset containing images of Cheetahs, Leopards, Lions, Pumas, and Tigers. The model's performance underscores its potential for practical applications in wildlife monitoring, conservation efforts, and beyond.

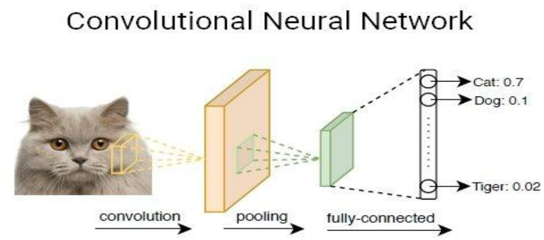
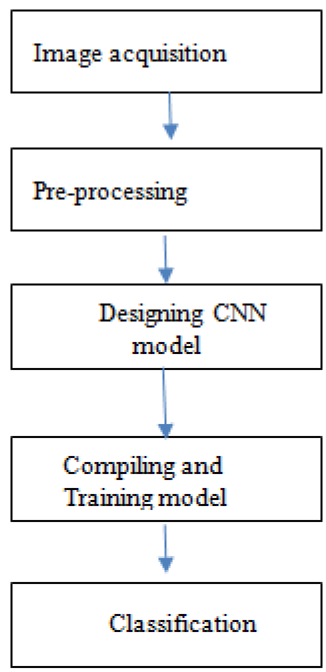


Fig. 1. Illustration of the proposed Convolutional Neural Network (CNN)

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

The CNN model, trained on 1000 photographs for 15 epochs, exhibited varying performance with different optimizers. Utilizing categorical cross-entropy and RMSprop with a learning rate of 0.001, the model's accuracy and loss were evaluated. Results in Table 1 revealed that the RMSprop optimizer outperformed categorical cross-entropy, showcasing superior accuracy and minimized loss. These findings are graphically represented in Figure 2, illustrating the model's learning progress over epochs. The RMSprop-optimized model demonstrated enhanced capabilities in identifying various cat species from images. This supervised learning approach, particularly when employing advanced optimizers like RMSprop, proves superior to conventional, untrainable methods. The outcomes emphasize the model's potential in accurate cat species recognition, presenting implications for wildlife monitoring and conservation efforts.

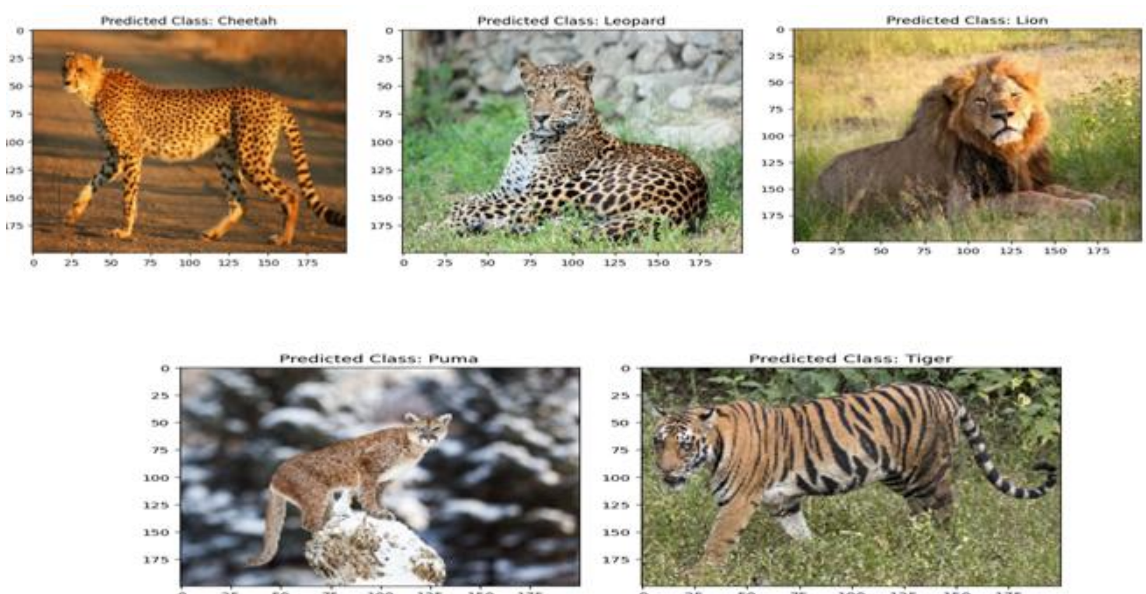


Table 1. Results depicting accuracy gained by Supervised Learning Model.

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 conclusion, this project has successfully explored the application of Convolutional Neural Networks (CNNs) for the recognition of various cat species within the feline family. The CNN model, trained on a dataset consisting of 1000 photographs and validated with 300 images, demonstrated noteworthy performance, particularly when optimized with the RMSprop optimizer. The comprehensive evaluation of different optimizers, including categorical cross-entropy, revealed that RMSprop consistently outperformed in terms of accuracy and loss metrics.The results, presented in Table 1 and visually depicted in Figure 2, underscore the efficacy of the developed model in identifying distinct cat species. The superior performance of the supervised learning model, compared to conventional, untrainable approaches, emphasizes the potential impact of deep learning in automating cat species recognition. The outcomes of this project hold significant implications for wildlife monitoring, conservation efforts, and broader applications in computer vision. Future work may involve further fine-tuning of the model and exploring its adaptability to diverse datasets and real-world scenarios.

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.