The Recognition and Analysis of Pet Facial Expression Using DenseNet-Based Model

Zhengchen Wu
School of Artificial Intelligence, China University of Mining & Technology, Beijing, China

Keywords: Pet Facial Expression Recognition, DenseNet, Data Augmentation, Fine-Tune.

Abstract: The research topic revolves around the recognition of pet facial expressions, a subject of growing importance in understanding the emotional states of animals, particularly pets. The primary objective of this study is to develop an effective model for pet facial expression recognition. This paper proposes a novel approach, leveraging the dense network (DenseNet) architecture, to address this challenge. Specifically, method involves the utilization of DenseNet's dense connectivity patterns to capture intricate features in pet facial expressions. This paper employs pre-trained weights from DenseNet121, implement data augmentation techniques, and fine-tune the model to improve its adaptability and recognition capabilities. This study is conducted on the Pet's Facial Expression Image Dataset, encompassing facial expressions of emotions in various categories of pets. The experimental results demonstrate the efficacy of the proposed approach. The model exhibits substantial progress in recognizing pet emotions, as indicated by impressive training accuracy. In conclusion, this research marks a significant step forward in pet facial expression recognition, with potential applications in veterinary care and enhancing pet-owner interactions. Understanding pet emotions through facial expressions has practical implications for animal welfare and human-animal communication. This research contributes to the development of tools and methods that can aid in improving the well-being of pets and strengthening the bond between pets and their owners.

1 INTRODUCTION

Facial expression recognition plays a crucial role in understanding the emotional states of humans and animals alike. While extensive research has been conducted on human facial expression recognition, the domain of pet facial expression recognition remains relatively unexplored. Pet facial expressions serve as a vital channel for emotional communication, contributing to a deeper understanding of their feelings. In this context, the emergence of deep learning has facilitated the automated extraction of image features. These advancements have introduced innovative avenues for research (Sreenivas et al 2021 & Liao et al 2021). Such as dense network (DenseNet) based deep learning, holds significant promise for recognizing and interpreting pet facial expressions and emotions.

As widely acknowledged, the Convolutional Neural Network (CNN) has established itself as a highly effective model architecture in the domain of image classification (Fukushima 1980). Numerous CNN-based network architectures have been introduced in the literature. Some, like AlexNet and ZFNet, employ fewer convolutional layers, while others, such as the Visual Geometry Group Network (VGG), Google Inception, and ResNet, delve into deeper layer configurations (Chen et al 2020, Lin et al 2017 & Yao et al 2019). The pivotal role of CNN in advancing facial expression recognition research is evident, enabling numerous researchers to attain significant achievements in image recognition. For instance, Lopes et al. explored facial feature extraction through preprocessing techniques, subsequently feeding them into a 5-layer CNN for facial expression recognition. Their approach resulted in an impressive recognition accuracy of 97.95% on the CK+ dataset (Lopes et al 2017). Li et al. introduced a CNN model for facial expression recognition that incorporates an attention mechanism. Their model achieved remarkable accuracy, reaching 75.82% on the Facial Emotion Recognition2013 (FER2013) dataset and an impressive 98.68% on the CK+ dataset (Li et al 2020). In the realm of facial expression recognition, Qin et al. introduced a novel approach that combines Gabor wavelet transform with a 2-channel CNN. Their method demonstrated significant promise, achieving an accuracy rate of 96.81% on the CK+ dataset (Qin

Wu, Z.
The Recognition and Analysis of Pet Facial Expression Using DenseNet-Based Model.
DOI: 10.5220/0012800000003885

In *Proceedings of the 1st International Conference on Data Analysis and Machine Learning (DAML 2023)*, pages 427-432
ISBN: 978-989-758-705-4

et al 2020). Li et al. introduced a novel approach for celebrity face recognition, utilizing a combination of Local Binary Patterns (LBP) and deep-CNN. Their method achieved significant results, attaining an accuracy of 80.35% on the CelebFaces Attribute (CelebA) dataset and an impressive 99.56% accuracy on the Labeled Faces in the Wild (LFW) dataset (Li and Niu 2020). Consequently, DenseNet emerges as a distinct architecture in this landscape, harnessing its dense connectivity characteristics to enhance the feature learning and propagation processes. In this study, the potential of DenseNet is harnessed to construct an innovative model for more accurate and robust recognition of pet facial expressions and emotions.

The primary aim of this research is to create and showcase the effectiveness of a DenseNet-based model for training and recognizing pet facial expressions to infer their emotions. Specifically, first, the inherent dense connectivity of DenseNet is leveraged to address the challenge of effectively capturing intricate features present in diverse pet facial expressions. The dense connections facilitate the seamless flow of information throughout the network, aiding in the extraction of both global and local features, crucial for accurate emotion recognition in pet faces. Second, the model is designed to accommodate the variability in pet breeds, sizes, and facial structures, addressing the issue of generalization across different types of pets. By learning from the shared features across different pet categories, the DenseNet model is expected to exhibit improved adaptability and recognition capabilities. Third, the predictive performance of various models, including the proposed DenseNet model, is extensively evaluated and compared. This comparative analysis serves to validate the efficacy of the DenseNet approach in capturing nuanced emotional cues from pet facial expressions. Additionally, to enhance the model's real-world application, data augmentation techniques are employed. This augmentation involves artificially introducing variations in lighting conditions, angles, and backgrounds to increase the model's robustness to real-world scenarios where such variations are common. Concurrently, transfer learning is explored, utilizing pre-trained DenseNet weights from general image datasets to kickstart model training on pet facial expressions, thereby potentially reducing the demand for extensive pet-specific training data. In conclusion, experimental results indicate that DenseNet's dense connectivity significantly enhances the model's ability to identify crucial features in pet facial expressions, leading to improved emotion recognition performance.

2 METHODOLOGY

2.1 Dataset Description and Preprocessing

The dataset used in this study, called Pet's Facial Expression Image Dataset, is sourced from Kaggle, encompassing various categories of image data (Dataset). This dataset serves the purpose of training and evaluating the classification performance of convolutional neural network models DenseNet. To cater to distinct input size requirements of these models, images are resized accordingly. During preprocessing, grayscale images are replicated into RGB channels to comply with the models' three-channel input. Furthermore, image data is normalized, scaling pixel values to the range of 0 to 1, enhancing training stability. These preprocessing steps ensure data consistency and suitability, providing substantial support for model training and evaluation. Figure 1 showcases some instances from the dataset.

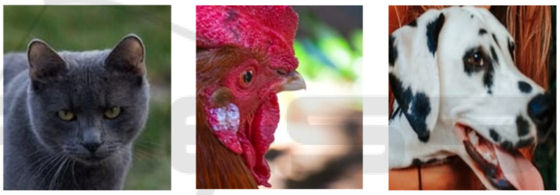

Figure 1: Images from the Pet's Facial Expression Image Dataset (Picture credit: Original).

2.2 Proposed Approach

The proposed approach for emotion classification in this study is centered around the adoption of the DenseNet model. DenseNet, a renowned convolutional neural network architecture, serves as the core of the method employed. It is well-regarded for its dense connectivity patterns, which enable effective feature extraction and utilization from pet facial images. By leveraging the capabilities of DenseNet, the model aims to capture both global and local features crucial for accurate emotion recognition in pet faces. The utilization of DenseNet's pre-trained weights, potentially reducing the demand for extensive pet-specific training data, enhances the adaptability and recognition capabilities of the model. Figure 2 below provides an illustration of the system's architecture.

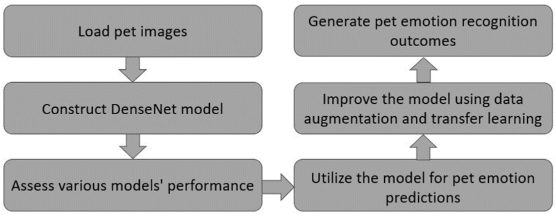
Load pet images
Construct DenseNet model
Assess various models' performance
Utilize the model for pet emotion predictions
Improve the model using data augmentation and transfer learning
Generate pet emotion recognition outcomes

Figure 2: The pipeline of the model (Picture credit: Original).

2.2.1 DenseNet

DenseNet plays a pivotal role in the emotion classification task due to its dense connectivity.Dense connectivity means that each layer is directly connected to all subsequent layers, fostering extensive feature reuse, which proves beneficial in capturing emotional information within images. DenseNet121 model leverages transfer learning by loading a pre-trained and fine-tuning it for emotion classification. This approach is advantageous as it allows the utilization of pre-trained model weights, originally trained on extensive image data, to expedite model training and improve overall performance. Densenet is a deep learning neural network architecture for image classification and target detection tasks. It was proposed by a team of researchers at Stanford University and its concepts and features were described in detail in a 2017 paper. The core concept of Densenet is dense connectivity, which connects all the feature maps of the previous layer to the inputs of the current layer by making connections on them. This means that there is a direct, dense flow of information between each layer, allowing for more efficient feature reuse in the network while reducing the number of parameters. The output of each layer is connected to the feature maps of all previous layers, facilitating the transfer and reuse of features. Compared to other network structures, Densenet has a smaller number of parameters because it shares the feature map between each layer. Due to the dense connections, gradients can be propagated more easily through the network, helping to mitigate the problem of gradient vanishing. With dense connections, Densenet allows the network to capture features at different levels from multiple layers, improving feature representation.

The DenseNet structure consists of multiple dense blocks, containing densely connected convolutional layers followed by optional 1x1 convolutional layers for channel reduction. Transition layers between dense blocks manage spatial dimensions and channel counts. Finally, global average pooling layers and dense output layers facilitate emotion classification. The implementation process unfolds with the loading of image data, its division into training and test sets, and the application of data augmentation techniques. DenseNet models is created and loaded with pre-trained weights and appended with custom classification heads. Model training occurs with the validation set, monitored by early stopping and learning rate scheduling mechanisms to prevent overfitting. Evaluation on the test data follows, with accuracy reported for model. Learning curves are visualized using Matplotlib to gain insights into the training and validation processes for each model. DenseNet's dense connectivity, transfer learning, data augmentation, and evaluation processes collectively contribute to the effective use of deep learning models for emotion classification in the provided code. DenseNet plays a pivotal role in the emotion classification task due to its dense connectivity. Dense connectivity means that each layer is directly connected to all subsequent layers, fostering extensive feature reuse, which proves beneficial in capturing emotional information within images.

2.2.2 Loss Function

The selection of an appropriate loss function holds great significance in the training of deep learning models. In the context of emotion classification tasks, the utilization of the Categorical Cross-entropy loss function is advantageous, given its proven effectiveness in addressing multi-class classification challenges. The Categorical Cross-entropy loss function calculates the discrepancy between the predicted probabilities and the true class probabilities. It penalizes the model when the predicted probabilities diverge from the target values, thus incentivizing the model to accurately assign higher probabilities to the correct emotion labels. This loss function is specifically designed for multi-class classification scenarios, where each input instance belongs to one and only one class or emotion category. When training a deep learning model for emotion classification, the optimization algorithm seeks to minimize the loss function. By minimizing the loss, the model aims to continually improve its ability to accurately classify emotions based on the given input features.

The Categorical Cross-entropy loss function incorporates the concept of logarithmic loss, which plays a vital role in capturing the discrepancies between predicted and true probabilities. It aims to minimize the difference between predicted and actual class distributions, promoting accurate classification during training, as follows：

$$\mathrm{L}(\mathrm{y},\hat{\mathrm{y}}) = -\Sigma_{\mathrm{i}}\mathrm{y}_{\mathrm{i}}\log(\hat{\mathrm{y}}_{\mathrm{i}}) \tag{1}$$

where $\mathrm{L}(\mathrm{y},\hat{\mathrm{y}})$ represents the loss function, assessing the dissimilarity between model's predictions $(\hat{y})$ and true labels (y). y_i represents ground truth probability of class, $\hat{y}_i$ represents the predicted probability of class i. The core concept of this formula is to calculate, for each class i, the product of the actual label probability (y_i) and the negative logarithm of the predicted probability $(-\log(\hat{y}_i))$. When the predicted probability closely matches the actual label probability, the loss tends towards zero; conversely, when there's a significant difference between them, the loss increases. This mechanism encourages the model to refine its predictions for each class, as higher loss values drive parameter adjustments aimed at minimizing the loss.

2.3 Implementation Details

In the execution of the suggested model, several important aspects are underscored. The pretrained DenseNet-121 models is used for image classification. Key steps include freezing the pretrained model's weights to retain valuable features, adding a global average pooling layer, and a custom classification layer. Data preprocessing involves resizing, grayscale conversion, and pixel normalization. Labels are typically one-hot encoded for classification tasks. These steps ensure that images are processed and fed into the model correctly. Data augmentation techniques, such as random rotation, horizontal flipping, and adjustments to brightness and contrast, diversifies the training data, reducing overfitting potential and enhancing the model's generalization capabilities.

3 RESULTS AND DISCUSSION

The In the conducted study, pretrained DenseNet-121 models is utilized to perform facial emotion recognition from a collection of over 1,000 images, each labelled with a specific emotion. Fig. 3 illustrates the model's learning curve.

As shown in Figure 3, epoch 1 starts with a loss of 1.2225. On the validation data, the loss is 1.2415. The initial results from Epoch 1 show relatively high loss values, which is expected at the beginning of training when the model's weights are randomly initialized. However, as training progresses, both training and validation loss decrease steadily. This is a promising sign, indicating that the model is effectively acquiring pertinent features from the pet facial expression images and enhancing its capability to classify emotions. As training advances, the model gradually acquires more information from the data, thereby improving its performance in classifying emotions. The decrease in loss values suggests that the model is progressively fitting the training data, implying that it can more accurately recognize and interpret the facial expressions of pets. This is crucial for emotion recognition as it demands the model to capture subtle emotional variations and expressions. By the end of Epoch 100, the model demonstrates remarkable improvements. The training loss decreases to 0.2142, signifying that the model is fitting the training data very well. While the validation loss (0.8686) is not as high as the training metrics, it's common to see a performance gap between training and validation datasets. This discrepancy may indicate that the model's capacity to generalize to unseen, new data is still evolving. Further training iterations or fine-tuning could potentially yield improvements in its performance on the validation dataset.

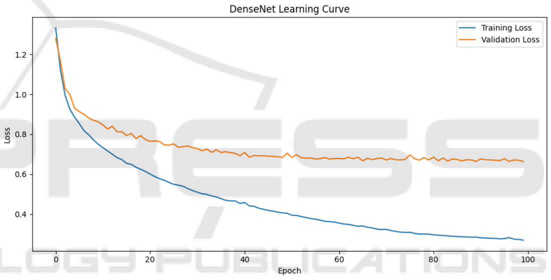

Figure 3: The learning curves of the model (Picture credit: Original).

Figure 4 shows the accuracy curves of the model. At the outset of Epoch 1, the model displays an accuracy of 36.67% on the training data, while the validation data yields a slightly lower accuracy of 31.67%. These initial accuracy values are in line with expectations, considering that training begins with randomly initialized model weights. However, as training progressed, both the training and validation accuracy consistently improved. This is a promising trend, signifying the model's proficiency in extracting pertinent features from the pet facial expression images and enhance its ability to accurately classify emotions. Fast forward to the end of Epoch 100, and the model showcased remarkable growth. It achieved an impressive training accuracy of 96.48%, signifying its proficiency in correctly classifying emotions for most of the training samples.

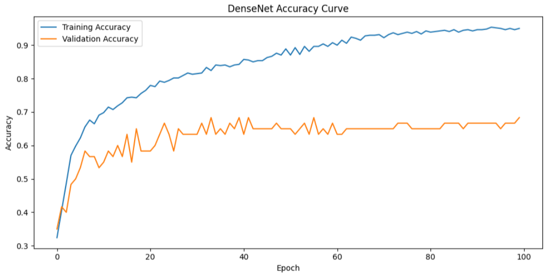

Figure 4: The Accuracy curves of the model (Picture credit: Original).

Nevertheless, it's worth noting that the validation accuracy, at 55%, didn't quite reach the same levels as the training metrics. Discrepancies between training and validation performance are common, suggesting that the model's generalization to unseen data may still be improving. Further training epochs or fine-tuning could potentially help bridge this performance gap and further enhance the model's effectiveness on the validation dataset. In summary, the DenseNet model demonstrated substantial learning and achieved high accuracy in classifying emotions in the training data. However, it's important to acknowledge that the validation accuracy fell short compared to the training metrics. This discrepancy is a frequent occurrence in machine learning, indicating that the model might require additional fine-tuning to better generalize to unseen data. By conducting further analysis and enhancing the model through additional training epochs or fine-tuning, it is possible to narrow the performance gap between the training and validation data. This iterative process would lead to an improved ability to recognize pet facial expressions in real-world scenarios.

While the DenseNet model exhibited promising results and a strong learning capability, there is still room for refinement. Through continued optimization efforts, it has the potential to excel in accurately classifying emotions and capturing pet facial expressions in diverse and unobserved contexts. Patience and iterative improvements will play a crucial role in maximizing the effectiveness of this model.

4 CONCLUSION

This study effectively closes the loop on the quest to understand and interpret pet emotions through their facial expressions. Building upon the introduction's premise of the importance of pet facial expression recognition, this paper's main methodology and contribution, the proposed model, have been introduced and thoroughly examined. The results of extensive experiments conducted on the proposed model showcase its potential in recognizing pet emotions. Notably, the model exhibits substantial progress in extracting meaningful features from diverse pet facial expressions. While the training accuracy reaches impressive levels, a key limitation emerges in the form of a performance gap between training and validation data, suggesting room for further fine-tuning and exploration of data augmentation strategies. In the future, research plans include expanding the study to a wider range of animal species than traditional pets, which is promising. Practical applications extend to veterinary care, animal well-being assessments, and even the burgeoning field of animal-human communication. Furthermore, the research aims to delve deeper into the analysis of pet behaviors and their correlation with facial expressions, thereby enriching understanding of pet emotions. This study underscores the importance of understanding pet emotions, not only for the welfare of animal companions but also for strengthening the bonds between humans and animals.

REFERENCES

V. Sreenivas, V Namdeo, E. Vijay Kumar, "Modified deep belief network based human emotion recognition with multiscale features from video sequences," Software: Practice and Experience, vol. 51, 2021, pp. 1259-1279.

H. Liao, D. Wang, P. Fan, et al. "Deep learning enhanced attributes conditional random forest for robust facial expression recognition," Multimedia Tools and Applications, vol. 80, 2021, pp. 28627-28645.

K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," Biological cybernetics, vol. 36, 1980, pp. 193-202.

Q. Chen, Y. Huang, R. Sun, et al. "An efficient accelerator for multiple convolutions from the sparsity perspective," IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 28, 2020, pp 1540-1544.

L. Lin, Y. Zhang, W. Zhang, et al. "A real-time smile elegance detection system: a feature-level fusion and SVM based approach," Electronic Imaging, 2017, pp. 80-85.

H. Yao, F. Dai, S. Zhang, et al. "Dr2-net: Deep residual reconstruction network for image compressive sensing," Neurocomputing, vol. 359, 2019, pp. 483-493.

A.T. Lopes, E. De Aguiar, A.F. De Souza, et al. "Facial expression recognition with convolutional neural networks: coping with few data and the training sample order," Pattern recognition, vol. 61, 2017, pp. 610-628.

J. Li, K. Jin, D. Zhou, et al. “Attention mechanism-based CNN for facial expression recognition,” Neurocomputing, vol. 411, 2020, pp. 340-350.

S. Qin, Z. Zhu, Y. Zou, et al. “Facial expression recognition based on Gabor wavelet transform and 2-channel CNN,” International Journal of Wavelets, Multiresolution and Information Processing, vol. 18, 2020, p. 2050003.

X. Li, H. Niu, “Feature extraction based on deep ‐ convolutional neural network for face recognition,” Concurrency and Computation: Practice and Experience, vol. 32, 2020, p. 1-1.

Dataset https://www.kaggle.com/datasets/anshtanwar/pets-facial-expression-dataset