

Combining Deep Learning Computer Vision and Geometric Morphometric Facial Landmarks Cat Facial Expression Recognition

| | |
|-------------------------|------------|
| Wanlipa Chamchum | 6520422003 |
| Budsadee Sareerasart | 6520422009 |
| Therarat Srisaswatakul | 6520422012 |
| Thanyakorn Hovongratana | 6520422018 |

INTRODUCTION

Cats are generally quiet and subtle in their facial expressions. Their behaviors are also difficult to interpret, so this makes it challenging for humans to accurately understand their moods and interact with them. Unlike other humans, we cannot communicate with them using verbal means, so there is no way to ascertain cat emotions using words. Humans who want to deeply and accurately understand how cats feel, such as caring cat owners or veterinarians, would benefit from an easier or more accurate method to understand cat emotions. That is, to understand how cats feel is important, particularly in environments where many cats are gathered together with humans. One good example is cat cafe, where so-called cat lovers may be less aware of cats' behaviors. Another, more serious, environment is in Veterinary hospitals where detecting cats' mood or pain is vital. Additionally, adoption centers or shelters, where cats along with other animals are rescued and taken care of, can also benefit from facial expression recognition to better manage cats' stress and possibly improve the chance of the cats being adopted [1].

Understanding cats' facial expressions can possibly change our ways of interaction with cats. With accurate interpretation of the cats' subtle facial expressions, we can enhance our ability to handle tasks such as pain management in Veterinary hospitals to reduce acute possible cases due to cats' unstable conditions or behavioral control in both cat cafe and adoption centers to avoid possible negative cats' reaction.

Humans may believe that they can understand cats well due to the time they spend with them, whether due to years of living with their pet cats for cat owners, having daily interactions with numerous cats for shelter workers, or professional training of veterinarians to handle cats' health issues, but according to a study [2], when cats exhibit more subtle emotions, humans cannot reliably detect such emotions. This calls for a more systematic and objective approach to identifying cat emotions.

RELATED WORK

The first work that needs to be mentioned in cat emotion recognition is CatFACS [1, 3], an objective cat facial action description system inspired by the well-known Facial Action Coding System (FACS) [4] used in humans. As the basis of CatFACS is facial action, in terms of media requirements, the user would need videos that show cat movement to use CatFACS if they are not observing a live cat. Another method researchers have used for recognition of cat emotions, such as being angry, happy or neutral, is through cat sound [5]. Although these approaches provide good results, facial expression recognition through static images is still of interest as finding and evaluating cat images are much simpler than processing of videos or sounds.

Surprisingly, cat facial emotion recognition based on static images is not as well-studied. One notable approach we will explore in our study is based on geometric morphometric of cat facial landmarks [6, 7] specifying 4 regions of interest (eyes, ears, nostrils and mouth) to analyze cats' facial structure and expressions in pain recognition. In direct relation to our work, there have also been other studies that use deep learning image classification for cat facial emotion recognition with promising results, but with limited depth [8, 7, 9]. Feighelstein et al. used pre-trained Resnet50 network to distinguish cat in pain and not in pain [7], while Jain et al. [9] and Vishwakarma et al. [8] used custom convolutional neural networks to identify 2 cat emotions (happy and sad) based on images. Since there are very few studies on cat facial expressions recognition, we look further for related work in other animals. One study uses CNN-based deep learning models with improved whale optimization algorithm (IWOA) achieving 89.91% accuracy on 5 dog emotions [10]. Another study uses SVM classifiers based on encoded facial data of non-human primates achieving 82% accuracy on 4 classes of emotions [11].

Drawing comparison from human facial expression recognition is a natural next step as there is much more research on humans compared to animals with regards to facial expression recognition. Common experiments on facial expression recognition in humans involve obtaining humans' straight facial images [12] [13] as an established dataset and are less likely to use images of human faces that are tilted. According to Li and Deng [12], the state-of-the-art static-based human facial expression recognition is still largely based on convolutional neural network (CNN), albeit with more advanced techniques such as auxiliary blocks (layers to enhance the expression-related representation capability of learned features), network ensemble and cascaded networks (sequentially combining modules for different tasks). Some of the well-known benchmark datasets for human facial expression recognition include

CK+ [14]: a laboratory-controlled images of a change in expression from neutral to peak expression

FER2013 [15]: an image database of human facial expressions in the wild collected from Google image search

AffectNet [16]: large-scale database of over 1 million images queried from the internet using emotions-related tags

RAF-DB [17]: a crowd-sourced collection of 29,672 annotated human facial emotion images where each image is annotated about 40 times

On the contrary, as noted earlier, there is no established state-of-the-art for cat facial emotion recognition machine learning algorithms. There is also no benchmark dataset for cat facial expression recognition. From these facts, it becomes clear that cat facial expression recognition based on machine learning algorithms remains very much underexplored.

In addition to the scarcity of research in cat facial expression recognition, there are other challenges when trying to identify cat facial expressions. Most importantly, animals cannot tell us what they feel, causing ground truth verification to be more complicated than in humans. Solving this issue by using experts to rate animals' affective states could also introduce bias due to difference in perception from humans to animals [7]. Animals may also display different facial expressions to human under similar emotional state [18]. Capturing facial expressions using action units is also subject to errors caused by complexity of facial musculature and may not cover facial emotions in specific species, not to mention the action units' inherent unsuitability to use on static images [6]. Difference in individuals is another factor that complicates facial emotion recognition. In human facial expression recognition, difficulties include not only visible muscle on human faces, but also flexibility in appearance or shape of the facial expression and probably inter-person differences in facial expression [13]. Likewise, in animals such as cats, their variations of facial expression are not in control due to greater variation in their facial morphology and certain absent facial features like pain feelings in some breeds. This variability can make population-level evaluations challenging for cats. In addition, hair covering may also obstruct certain facial features such as wrinkles [6].

To address these difficulties, we propose combining deep learning computer vision techniques with geometric morphometric from facial landmarks to predict cat emotion from cat facial images. This approach leverages both the strength of deep learning to learn subtle features and the explainability of geometric morphometric, which consist of vectors obtained from different landmarks of cat faces. There are studies showing that the use of geometric morphometric can help in identifying facial expressions in cats [6] and in humans [19].

Another challenge we face with cat facial expression recognition (FER) is the lack of directly comparable labeled datasets for cat expressions, unlike those available for humans. To this end, in addition to running supervised classification model, we will also take one of the more academically established cat facial datasets, the Cat Facial Landmarks in the Wild dataset or CatFLW [20] and try to label the images. We follow the state-of-the-art in human facial expression recognition semi-supervised learning according to Roy and Etemad [21], using FixMatch and Mean Teacher as the 2 best-performing semi-supervised methods out of 8 methods they examined. The study also shows that performances of the aforementioned

methods are not far below that of the fully supervised methods on *FER13*, *RAF-DB* and *AffectNet* [21]. Finally, we will run the classification again on the combined dataset.

Summary of Related Work

Facial Expression Recognition in Humans

- Human FACS as qualitative approach to capture human emotions
- State-of-the-art machine learning approach established for human static-based facial expressions recognition
- Countless benchmark datasets including CK+, FER2013, AffectNet and RAF-DB

Facial Expression Recognition in Cats

- CatFACS for capturing cat facial emotions qualitatively, inspired by human FACS
- Cat sound also used in cat emotion classification
- Cat facial landmarks – geometric morphometric used to recognize cat pain
- No state-of-the-art machine learning approach for cats
- No benchmark datasets

Our contributions

- Introduce a multimodal deep learning classification approach combining image input and facial landmark vector input
- Use the resulting model to label the CatFLW and re-train the classification model on the labeled CatFLW dataset

METHODOLOGY

1) Initial Dataset Preparation

- a) We obtain an open image dataset having 3 specific classes of cat facial emotions including Anger, Happiness and Sadness.
- b) Due to the low volume of images from the above dataset, we search for more cat facial pictures using 3 keywords (Anger, Happiness and Sadness) and apply such keywords as our cat emotion classes.
- c) All cat face images acquired are of straight-faced pose
- d) We then combine images from the 2 approaches above into a single image dataset for further usage.

2) Facial Landmark Extraction

- a) We use an animal facial landmark detector belonging to an experiment which devoted to Cat facial landmark automated detection using Ensemble Landmark Detector (ELD) [20]. By using this detector, it returns 48 landmarks on cat face.

3) Feature Engineering

- a) Identify 4 regions on cat face using 48 landmarks: Ears, Left eye, Right eye and Nose-Mouth-Whiskers.
- b) Determine the mean point (center point) of each region.
- c) Generate vectors by using the center point of each region as the origin and creating a vector from the center point

4) Model Implementation

- a) Step 1: Supervised Learning
 - i) Utilize vectors obtained from Facial Landmark Extraction state for self-supervised learning using classical ML algorithms (non-deep learning) to see the most suitable model for further classification.
 - ii) Create deep learning algorithms combining facial image inputs and vectors obtained from facial landmarks

- b) Step 2: Semi-Supervised Learning
 - i) Use the resulting model to label the CatFLW dataset with the 2 semi-supervised learning algorithms
 - (1) FixMatch
 - (2) Mean Teacher
- c) Performance Evaluation
 - i) TBD

REFERENCES

- [1] C. Caeiro, A. Burrows and B. Waller, "Development and application of CatFACS: Are human cat adopters influenced by cat facial expressions?," *Applied Animal Behaviour Science*, vol. 189, pp. 66-78, 2017.
- [2] L. Dawson, J. Cheal, L. Niel and G. Mason, "Humans can identify cats' affective states from subtle facial expressions," *Animal Welfare*, vol. 28, no. 4, p. 519–531, 2019.
- [3] C. C. Caeiro, B. M. Waller and A. M. Burrows, *The Cat Facial Action Coding System Manual*, 2013.
- [4] P. Ekman and W. V. Friesen, "Facial Action Coding System (FACS)," *APA PsycTests*, 1978.
- [5] Pandeya, Y. R. a. Kim, D. a. Lee and Joonwhoan, "Domestic Cat Sound Classification Using Learned Features from Deep Neural Nets," *Applied Sciences*, vol. 8, 2018.
- [6] L. R. Finka, S. P. Luna, J. T. Brondani, Y. Tzimiropoulos, J. McDonagh, M. J. Farnworth, M. Ruta and D. S. Mills, "Geometric morphometrics for the study of facial expressions in non-human animals, using the domestic cat as an exemplar," *Scientific Reports*, no. 9, p. 9883 (2019), 2019.
- [7] I. S. L. R. F. S. P. L. L. D. S. M. a. A. Z. Marcelo Feighelstein, "Automated recognition of pain in cats," *Scientific Reports*, 2022.
- [8] M. VISHWAKARMA, K. JANGID, K. PAL and L. RODRIGUES, "Cat Breed & Emotion Detection Using Yolo, CNN & Canny Edge Detection," *IRE Journals*, vol. 7, no. 8, 2024.
- [9] P. JAIN, A. K. PANDEY, V. M. K. PANDEY and B. YADAV, "Facial Emotion Recognition of Cat Breeds by Using Convolution Neural Network," *IRE Journals*, vol. 7, no. 6, pp. 58-63, 2023.
- [10] Y. Mao and Y. Liu, "Pet dog facial expression recognition based on convolutional neural network and improved whale optimization algorithm," *Scientific Reports*, vol. 13, p. 3314, 2023.

- [11] G. Blumrosen and D. Hawellek, "Towards Automated Recognition of Facial Expressions in Animal Models," in IEEE International Conference on Computer Vision (ICCV), 2017.
- [12] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," IEEE Transactions on Affective Computing, vol. 13, no. 3, pp. 1195-1215, 2022.
- [13] S. Bazrafkan, T. Nedelcu, P. Filipczuk and P. Corcoran, "Deep Learning for Facial Expression Recognition: A step closer to a SmartPhone that Knows your Moods," 2017 IEEE International Conference on Consumer Electronics (ICCE), pp. 217-220, 2017.
- [14] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, San Francisco, CA, USA, 2010.
- [15] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, Y. Zhou, C. Ramaiah, F. Feng, R. Li, X. Wang, D. Athanasakis and J. Sha, "Challenges in Representation Learning: A Report on Three Machine Learning Contests," in Neural Information Processing , Berlin, 2013.
- [16] A. Mollahosseini, B. Hasani and M. H. Mahoor, "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild," IEEE Transactions on Affective Computing, vol. 10, no. 1, pp. 18-31, 2019.
- [17] S. Li, W. Deng and J. Du, "Reliable Crowdsourcing and Deep Locality-Preserving Learning for Expression Recognition in the Wild," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017.
- [18] C. Caeiro, K. Guo and D. Mills, "Dogs and humans respond to emotionally competent stimuli by producing different facial actions," Scientific Reports, vol. 7, p. 15525, 2017.
- [19] Y. Qiu and Y. Wan, "Facial Expression Recognition based on Landmarks," in 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chengdu, China, 2019.
- [20] G. Martvel, I. Shimshoni and A. Zamansky, "Automated Detection of Cat Facial Landmarks," International Journal of Computer Vision, vol. 132, p. 3103-3118, 2024.
- [21] S. Roy and A. Etemad, "Analysis of Semi-Supervised Methods for Facial Expression Recognition," in 2022 10th International Conference on Affective Computing and Intelligent Interaction (ACII), Nara, Japan, 2022.

