AREPS: A Robust Explainable Pawpularity Class Predicting System

Kausik Lakkaraju, Doctoral Student, University of South Carolina

December 2021

Abstract

Millions of stray animals suffer on the streets and some animals end up in pet shelters where they may or may not get adopted. PetFinder.my, a leading animal welfare platform, in Malaysia measures the attractiveness of a pet photo using a very basic cuteness meter. This tool is still at an experimental stage. There is a need for a system which could provide better results. Deep learning models are known to make good predictions when trained on a proper dataset. Many researchers have proposed robust architectures which make predictions with a very high accuracy. Despite their high accuracy, it is difficult to know the decision-making process of neural network model. Research into Explainable AI (XAI) has been increasing in the recent years to makes systems more transparent and trustworthy. Making the system explainable make the users trust the system more. This is very critical as AI systems are prone to bias and other trust issues. Kaggle recently created a competition called 'Pawpularity Contest' [5] where the given dataset has raw images of pets from pet shelters and the scores given to those photos which is called 'Pawpularity Score' based on how cute those images are. The task, as a part of competition, is to get a Root Mean Square Error (RMSE) value which is as less as possible on the test dataset that is provided. The additional module that I am adding to the project is the explainability. This will make us understand the reason behind the decision taken by the neural network model, which is otherwise a black-box.

1 Problem

Pets with attractive photos get adopted faster as they generate more interest. PetFinder.my is Malaysia's leading animal welfare platform, featuring over 180,000 animals with 54,000 animals already adopted. PetFinder.my uses a very basic cuteness meter to measure the attractiveness of a photo. Though the tool that they are using is being helpful, it is still at experimental stage. There is a need for a system which could provide better results.

Many AI researchers have been building better neural network models which could make a prediction with very high accuracy. Some of the best models which came out in the recent times are: InceptionV3 [9], ENet [4], VGG [8] and there are many other models which have received a commendable accuracy on datasets like imagenet. These neural network models can be used to predict the Pawpularity score for the pet photos. The model built by the winners of the competition will be used by PetFinder.my to make accurate predictions. This would ultimately improve the appeal of pet profiles. However, would model accuracies or other metrics be sufficient to say that the model is interpretable? [7]

An AI-powered medical diagnosis system which wrongly diagnoses a patient to not have a cancer when they actually have cancer or an autonomous vehicle that drives erratically and causes fatal accidents despite regular road conditions are some of the cases in which transparency of the system is very essential. If the reason for the AI system's mistake is known, AI system developers could learn why the model took a wrong decision and they could build better systems in the future.

In addition to building a model which would be able to assess how attractive a picture is, it is also important to make the system more transparent. That kind of transparency would create the trust between user and the AI system. Some systems might have high accuracy but still might be prone to bias and other trust issues. It is important to reach a mid point where the model is able to perform with good amount of accuracy and also provide a valid explanation which could make the system interpretable by the user.

This solution will have a significant impact as it not just predicts how attractive a photo is but also makes the black-box neural network more transparent and trustworthy. A robust neural network model and a good explanation method would make this work publishable and significant.

2 Related Work

AI research has become quite interdisciplinary. AI researchers are not just solving problems which are related to the field of computers but also exploring various other domains like biology, material design and chemistry. AI has been helping to solve many critical real-world problems. One such example is [2]. They created a software called Wildbook which could identify animals by their unique coat patterns or other hallmark features. This is now helping some scientists at North Kenya to preserve the endangered giraffe population. This shows how powerful computer vision can be in solving certain problems. Image Processing techniques when combined with deep learning would create such a powerful tool. Applications like this motivate me to use AI for animal welfare.

In a survey paper [3], the authors have explained various XAI algorithms, evaluation metrics along with an interesting case study which showed the growth of XAI from 2007 to 2020. This is one of the papers that motivated me to explore the concepts of XAI and use it in my project. Big companies like Google are also giving more importance to research in this area. [10] shows an increase in trend for XAI.

Though not much research has been done in XAI in this specific domain of animal protection, various researchers have used XAI in medical diagnosis and in detection of special conditions in humans. One such recent work is [1] where the authors have built a model to detect a condition called Autism Spectrum Disorder (ASD). They have built a model which generates autism subtypes and identify discriminatory factors among them. They have used logistic regression and built a SHAP explainer on top of that.

3 Methodology

3.1 Input

Input to the trained model would be an image of a pet.

3.2 Output

Output that we would receive from the trained model when given a test image is the prediction of the model on that image and the input image with ex-

	Id	Subject Focus	Eyes	Face	Near	Action	Accessory	Group	Collage	Human	Occlusion	Info	Blur	Pawpularity
0	0007de18844b0dbbb5e1f607da0606e0	0	1	1	1	0	0	1	0	0	0	0	0	63
1	0009c66b9439883ba2750fb825e1d7db	0	1	1	0	0	0	0	0	0	0	0	0	42
2	0013fd999caf9a3efe1352ca1b0d937e	0	1	1	1	0	0	0	0	1	1	0	0	28
3	0018df346ac9c1d8413cfcc888ca8246	0	1	1	1	0	0	0	0	0	0	0	0	15
4	001dc955e10590d3ca4673f034feeef2	0	0	0	1	0	0	1	0	0	0	0	0	72
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
9907	ffbfa0383c34dc513c95560d6e1fdb57	0	0	0	1	0	0	0	0	0	0	0	1	15
9908	ffcc8532d76436fc79e50eb2e5238e45	0	1	1	1	0	0	0	0	0	0	0	0	70
9909	ffdf2e8673a1da6fb80342fa3b119a20	0	1	1	1	0	0	0	0	1	1	0	0	20
9910	fff19e2ce11718548fa1c5d039a5192a	0	1	1	1	0	0	0	0	1	0	0	0	20
9911	fff8e47c766799c9e12f3cb3d66ad228	0	1	1	1	0	0	0	0	0	0	0	0	30

9912 rows × 14 columns

Figure 1: Metadata in the form of .csv

planation added to it emphasizing the prominent features which contributed in the model prediction.

3.3 Data

The provided dataset is a multimodal dataset. They have provided a .csv file with 14 columns giving details about each image. The ID column has the file name corresponding to that image. Other columns are focus, eyes, face, near, action, accessory, group, collage, human, occlusion, info, blur which describe various features of the image. In total, there are 9912 rows/images in the dataset. They have mentioned in their details about the data that the metadata which is provided in the form of .csv is not very critical and they were given to get a better understanding of the different features of the image. That is the reason why I have only extracted the Pawpularity score from the dataset and also the image ID to extract the corresponding image from the image directory they have provided. A snapshot of the data can be seen in Figure 1 and Figure 2.

3.4 System Overview

The overall system workflow representation can be seen in Figure 3. The images are preprocessed along with their pawpularity classes and is given as input to the neural network classifier. After the model is trained, Pawpularity

Figure 2: Sample image from the provided image data

class of a test image is predicted which will be given as an input along with the trained model to the explainer which will output the same image with marked prominent features.

3.5 Pre-processing

The pre-processing steps that have been followed before training the model are:

- Fetching the images from the given image folder by using the ID names given in the metadata and storing the corresponding Pawpularity score in a separate list.
- Converting images to arrays so that it can be processed by neural networks. The images cannot be converted to grayscale for this problem as the color of the spots also plays an important role to determine whether or not a picture is attractive.
- I have converted the problem to a classification problem rather than a regression problem as the training would be faster. For this I have created 5 classes based on the Pawpularity score which ranges from 0 - 100.
- Converting the image labels to binary arrays for easier training.
- Normalizing the pixel values by dividing it by 225 so that the values can be comparable.

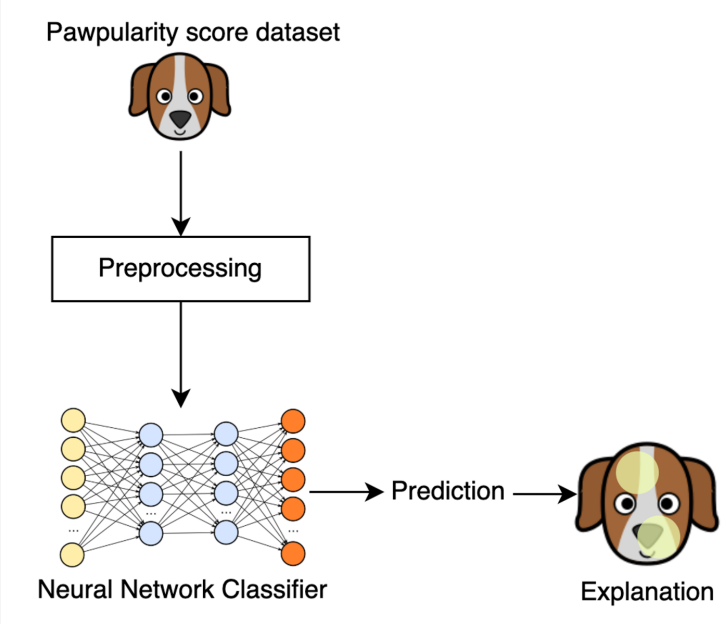


Figure 3: Overall workflow of the system

- Augmented the dataset by applying horizontal flip which would add more variation to the data while training.

3.6 Classification

For this project, I have used 2 kinds of classifiers: CNN and InceptionV3.

- **Convolutional Neural Network**: A Convolutional Neural Network is a class of artificial neural network, most commonly applied to analyze visual imagery. I have used a 5-layer CNN with 5-sets of 2D Convolutional layers, ReLU Activation layers, Batch Normalization layers, Max Pooling layers and Dropout layers. At the end, I have added some Dense and Batch Normalization layers. The whole architecture can be seen in the provided colab notebook.

- **InceptionV3**: Inception v3 [9] is a convolutional neural network for assisting in image analysis and object detection, and got its start as a module for Googlenet. It is the third edition of Google's Inception Convolutional Neural Network, originally introduced during the ImageNet

Recognition Challenge [11]. I initialized the model with pre-trained imagenet weights and applied transfer learning to use it for this project. I have applied 3 dense layers at the end. Two with ReLU activation functions and one with softmax activation function which gives out a multinomial probability distribution as output.

3.7 Explanation

To provide an explanation for the black-box models, I have used the LIME (Local Interpretable Model-agnostic Explanations) explainer [6]. It creates a local linear approximation for the model behavior. After training the model on the dataset, I have obtained the prediction of the models on images which are not present in either of the datasets (I have used images of the pets that we have adopted from a pet shelter). It highlights the portions of the image which have contributed to the decision taken by the model for that particular image. The results of these can be seen in section 4.

4 Results

I have trained two different models which were described in section 3.6 on the dataset. I have used the prescribed metric in the Kaggle competition description to evaluate my models. This metric is Root Mean Square Error (RMSE) which is given by the equation.

$$RMSE = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{1}$$

$\hat{yi}$ is the predicted value and yi is the actual value (ground truth) for each instance i.

These model results can be seen in Table 1. It can be clearly seen that CNN performed better when trained on the dataset for less number of epochs than InceptionV3. I have captured the loss curve while the CNN model was being trained. From the loss curve, it can be interpreted that this is an unrepresentative training dataset that means the training data does not provide sufficient information to learn the problem, relative to validation dataset used to evaluate it.

After getting the model predictions, I have used LIME explainer to get the explanation for model predictions. The results of this can be seen in

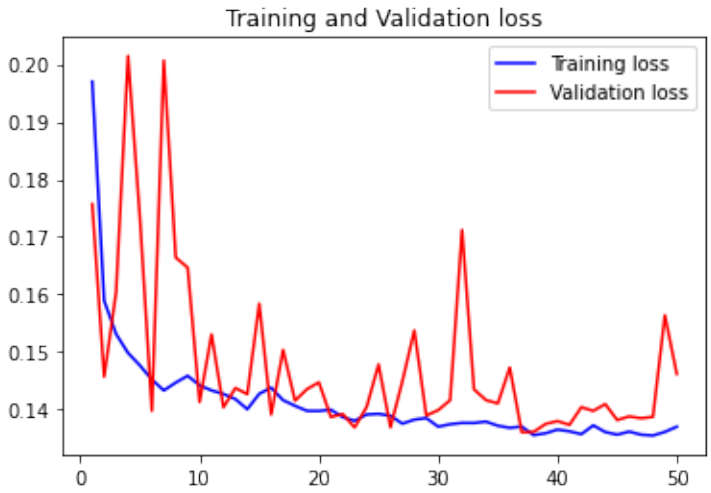
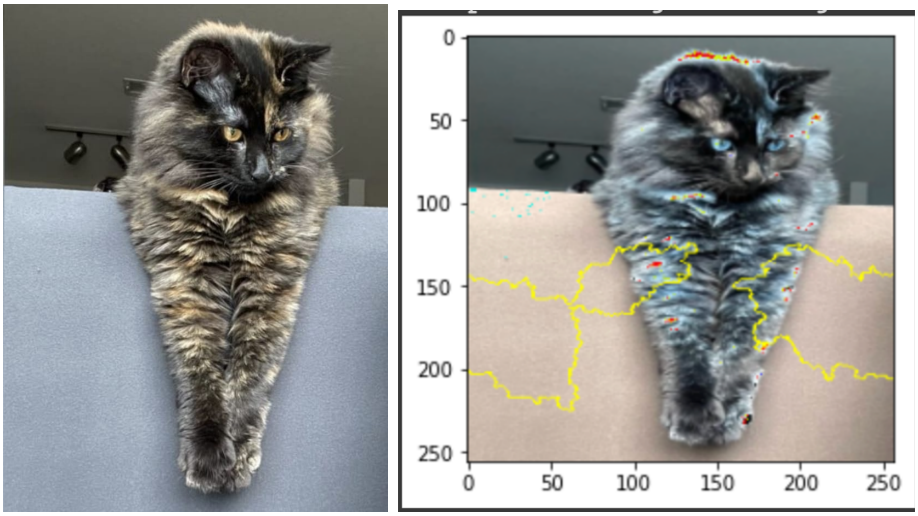


Figure 4: Loss curve while training the CNN model

Model	**RMSE**	**Loss**	**Epochs**
CNN	0.38	0.14 (MSE)	50
InceptionV3	0.45	0.97 (BCE)	60

Table 1: Root Mean Square Error (RMSE) and Loss are the metrics that I have included while training the models. I have used Binary Cross-entropy (BCE) loss for InceptionV3 and Mean Squared Error (MSE) loss for CNN. These are the results when the models are tested on the test datasets.



(a) Original Image (b) LIME Explanation

Figure 5: LIME Explanation for CNN prediction on the given test image

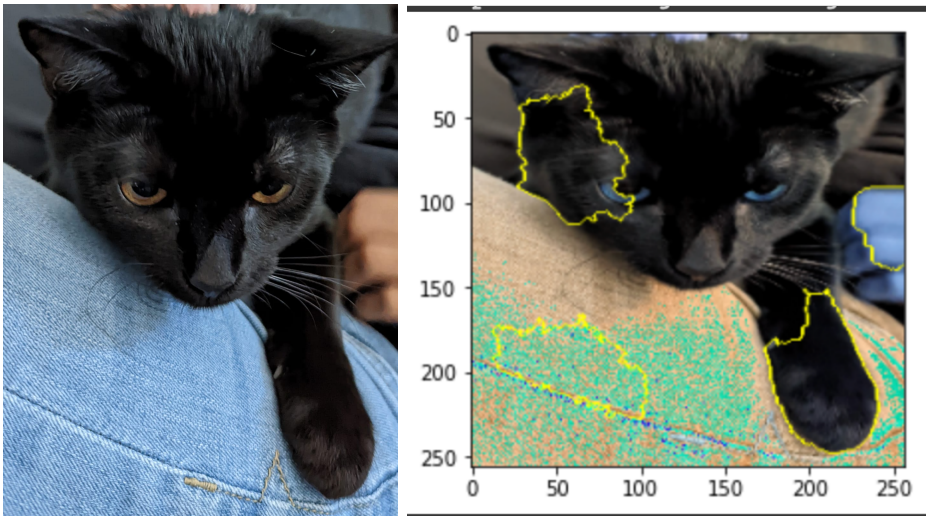
Figure 5 and Figure 6 for the models CNN and InceptionV3 respectively.

From the boundary marked by LIME in Figure 5, we can say that more than the cat posture, the background was responsible for providing the below given prediction. If the background is better, it would have given a better score. We can also find some red dots on the image, indicating a negative contribution to the prediction.

From the boundary marked by LIME in Figure 6, we can say that the paw, eye and some part of the background has contributed to the model prediction.

5 Conclusion

My project is an attempt to solve the problem of pet adoption. I have built two neural network models and trained them on the Pawpularity score dataset. I have also provided explanation for why the model made that particular prediction using LIME explainer. There is still a lot of scope for further improvement to model training. This project has been a good learning experience and motivated me to participate in more Kaggle competitions. AI researchers should realize that, in addition to building robust neural net-



(a) Original Image (b) LIME Explanation

Figure 6: LIME Explanation for InceptionV3 prediction on the given test image

work architectures which could achieve high accuracies, it is also important to make the systems more transparent and trustworthy.

6 Future Work

As this dataset was released as a part of the Kaggle competition, I would like to build a better model which would achieve high accuracy on the test set. In addition to making the architecture robust, I would also like to use better explanation methods to explain the model prediction to the users. What I did is just the tip of the iceberg. There are still a lot of exciting experiments that I could do with the data.

References

[1] Milon Biswas, M. Shamim Kaiser, Mufti Mahmud, Shamim Al Mamun, Md. Shahadat Hossain, and Muhammad Arifur Rahman. An xai based autism detection: The context behind the detection. In Mufti Mahmud, M. Shamim Kaiser, Stefano Vassanelli, Qionghai Dai, and Ning Zhong, editors, *Brain Informatics*, pages 448–459, Cham, 2021. Springer International Publishing.

[2] Anne Casselman. https://www.nationalgeographic.com/animals/article/artificial-intelligence-counts-wild-animals.

[3] Arun Das and Paul Rad. Opportunities and challenges in explainable artificial intelligence (XAI): A survey. *CoRR*, abs/2006.11371, 2020.

[4] Adam Paszke, Abhishek Chaurasia, Sangpil Kim, and Eugenio Culurciello. Enet: A deep neural network architecture for real-time semantic segmentation. *CoRR*, abs/1606.02147, 2016.

[5] PetFinder.my. Petfinder.my - pawpularity contest.

[6] Marco Túlio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should I trust you?": Explaining the predictions of any classifier. *CoRR*, abs/1602.04938, 2016.

[7] Cynthia Rudin and Joanna Radin. Why are we using black box models in ai when we don't need to? a lesson from an explainable ai competition. *Harvard Data Science Review*, 1(2), 11 2019. https://hdsr.mitpress.mit.edu/pub/f9kuryi8.

[8] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.

[9] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. *CoRR*, abs/1512.00567, 2015.

[10] Google Trends. https://trends.google.com/trends/explore?date=2016-01-01%202019-07-07q=explainable%20ai.

[11] Wikipedia. https://en.wikipedia.org/wiki/inceptionv3.