

Article

Deep Learning Pet Identification Using Face and Body

Elham Azizi *  and Loutfouz Zaman 

Faculty of Business and Information Technology, Ontario Tech University, Oshawa, ON L1G 0C5, Canada; loutfouz.zaman@ontariotechu.ca

* Correspondence: elham.azizi@ontariotechu.ca

Abstract: According to the American Humane Association, millions of cats and dogs are lost yearly. Only a few thousand of them are found and returned home. In this work, we use deep learning to help expedite the procedure of finding lost cats and dogs, for which a new dataset is collected. We applied transfer learning methods on different convolutional neural networks for species classification and animal identification. The framework consists of seven sequential layers: data preprocessing, species classification, face and body detection with landmark detection techniques, face alignment, identification, animal soft biometrics, and recommendation. We achieved an accuracy of 98.18% on species classification. In the face identification layer, 80% accuracy was achieved. Body identification resulted in 81% accuracy. When using body identification in addition to face identification, the accuracy increased to 86.5%, with a 100% chance that the animal would be in our top 10 recommendations of matching. By incorporating animals' soft biometric information, the system can identify animals with 92% confidence.

Keywords: convolutional neural networks; transfer learning; animal identification



Citation: Azizi, E.; Zaman, L. Deep Learning Pet Identification Using Face and Body. *Information* **2023**, *14*, 278. <https://doi.org/10.3390/info14050278>

Academic Editors: Mohd Nor Akmal Khalid, Hanhan Maulana, Masnizah Mohd, Nursuriati Jamil and Hiroyuki Iida

Received: 11 March 2023

Revised: 28 April 2023

Accepted: 4 May 2023

Published: 8 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

According to recent data from the Animal Protection Index (API) (<https://api.worldanimalprotection.org/country/canada>, accessed on 3 May 2023), when it comes to animal protection laws, Canada has not been performing adequately. The low marks received were for not having adequate legislation protecting wildlife in captivity, working animals, farm animals, animals used in research, and companion animals. At the same time, public support for improving animal welfare is high (<https://spca.bc.ca/news/volunteer-week>, accessed on 3 May 2023). The Ontario Society for the Prevention of Cruelty to Animals (SPCA) and Humane Society have reported that more than 10 million pets go missing every year in North America, and only 25% of them are claimed and returned home (<https://ontariospca.ca/blog/lost-pet-recovery-101>, accessed on 3 May 2023). In another study, Weiss et al. [1] found that between 11–16% of dogs and 12–18% of cats are lost at least once in five years in the US. On average, 75% of cats and 93% of dogs are found, but only 1.5–4.5% of cats and 10–30% of dogs are returned to their owners. This means that most stray dogs and cats are never reunited with their owners once they end up in an animal shelter.

Different jurisdictions have their own strategies when it comes to unclaimed lost pets. Some animals end up being adopted, some are euthanized, yet some suffer from serious behavioral problems (<https://www.diamondpet.com/blog/adoption/strays/what-happens-after-a-dog-arrives-at-a-shelter>, accessed on 3 May 2023). There are statistical reports in many jurisdictions on lost cats and dogs where a common pattern emerges: lost rates are high, and the Return-to-Owner (RTO) rate is low. Conventional tools and methods help to prevent pet loss or to expedite their search. The simplest and most economical method is using collar tags (<https://archive.nytimes.com/well.blogs.nytimes.com/2011/09/21/the-importance-of-pet-tags>, accessed on 3 May 2023). These tags usually include the owner's phone number and/or other contact information. However,

many pet owners delay or resist using these tags, resulting in the increased likelihood that a pet ends up lost (<https://archive.nytimes.com/well.blogs.nytimes.com/2011/09/21/the-importance-of-pet-tags>, accessed on 3 May 2023). On the other hand, collar tags are vulnerable and likely to be lost, worn out, or damaged. The contact information otherwise could also be tattooed on the animal's body, but it is not popular due to the higher costs. Tattooing also can cause pain and risk developing infections in animals (<https://pethelpful.com/pet-ownership/tattooing-a-cat>, accessed on 3 May 2023). Additionally, tattooing pets may not be suitable for certain species and breeds due to the specifics of the coat, which makes noticing the tattooed information difficult (<https://pethelpful.com/pet-ownership/tattooing-a-cat>, accessed on 3 May 2023). A more high-tech solution is to use GPS tags and microchips. GPS tags are tied to the pet allowing the owner to locate their lost pet directly. Similar to the tags, these devices are vulnerable to impact damage and weather interference [2]. Some pets, reportedly, have been suffering from the side effects of allergic reactions to these devices [3]. Microchipping is also another novel method used for pet identification. Some reports indicate that microchipping increases the risk of cancer in pets [4]. Meantime, microchipping is not a prevalent means because of its expensive price, and not all pet owners expect to lose a pet until it happens [5]. It is important to note that all the methods mentioned above assume that the lost pets have been located. Some lost pets manage to find their way back home, while others are located. When located, they are usually taken to shelters or returned to the owners directly. Some pets may wander away to other jurisdictions, and animal control services in smaller towns may not pick up stray cats or dogs, so they end up taken to rescues or shelters farther from where they were found. There are online websites and platforms to help bring pets back home. PetFBI (<https://petfbi.org>, accessed on 3 May 2023) and PetFinder (<https://www.petfinder.com>, accessed on 3 May 2023) expedite the search of a lost pet by requiring information such as the location of the lost animal. These platforms can help expedite the procedure of finding lost pets, but, at the same time, they can be time consuming and not efficient, since the owners should go through all the posts one by one. More intelligent and effective suggestions should be provided to help owners or shelters.

Photos of lost animals are usually taken in shelters. Owners also post some photos of their pets for common notice. People can also take photos or videos while encountering a lost animal to report later. This data can help identify pets. To speed up the identification of lost cats and dogs, we introduce a new framework based on deep learning. We apply classification for species detection and use transfer learning to improve accuracy. The results are connected to the next module dealing with individual animal identification. Capturing lost animals' full-face images, which are fully aligned with the camera direction, is not always possible and can make the animal feel uncomfortable. To be able to work around this, we identify animals using two methods: face and body identification. For face identification, we first detect the face and use landmark detection, which refers to the task of marking key points in the faces that appear in typical animal faces, such as lips or eyes [6]. Face alignment helps to achieve a more accurate match by establishing a common basis for the comparison of animal photos. The animal does not have to look directly at the camera. Both face and body identification run in parallel. We apply feature extraction in both methods, and the results of one method complement the other in situations where the data is not available for one of the methods. The contribution of this work is a framework to expedite finding lost cats and dogs, which uses deep learning and the following features:

- A full-body animal identification system, which matches animals based on face and body where the results of one method complement the other;
- Improved species classification using transfer learning;
- The use of facial landmark detection for animal face detection and comparison of different face alignment functions;
- The design of a recommendation system for animal owners who lost their pets;

- The use of additional information, which are breed, gender, age, and species of lost animals to reduce the search area and to expedite finding as well as improve the accuracy.

2. Related Work

Deep learning (DL)—a subcategory of machine learning which deals with artificial neural networks—has recently gained popularity due to the ability to maximize the utilization of unstructured data, eliminate the need for feature engineering, unnecessary costs, and data labeling, and the ability to deliver high-quality results. To help animal welfare and wellbeing, various deep learning applications are used for different purposes.

One of the applications of DL is the identification of animal species, where convolutional neural networks (CNN) are used for classifying species [7]. CNNs were previously used to classify marine animals [8]. Komarasamy et al. [9] introduced a new method for animal face classification, which was built based on the score-level fusion of CNNs. Komarasamy et al. observed the positive effect of extracted features, achieving 95.31% accuracy. A new system of characterization for animal faces was proposed. It was reported that their designed model improved the accuracy of animal classification. Norouzzadeh et al. [10] utilized a CNN to identify multiple species by counting and classifying their behavior. A dataset of 3.2 million images was created for 48 species, and a 93.8% accuracy was achieved by Norouzzadeh et al. with a plan that was shared to improve the accuracy.

In farming, the identification of individual cattle is needed for farm management. CNNs were used, e.g., by Yongliang et al. [11], to recognize beef cattle via both CNN and LSTM (Long Short-Term Memory) network methods. The authors used the combination of CNN and LSTM for better results. Similarly, here, we also use the combination of image- and video-based classification to get satisfactory results.

In a study by Barbedo [12], the cattle were detected using a UAV (Unmanned Aerial Vehicle), which captured aerial view images via a CNN. In this study, different breeds of cattle were detected, and 15 different CNN architectures were tested. Breed classification for cats and dogs was investigated by Omkar et al. [13], which achieved 59% accuracy on 37 different breeds. In their work, Omkar et al. introduced a pet breed classification model. Two main architectures were compared: hierarchical and flat. The hierarchical approach first assigns the animal to one of the categories of cats or dogs using a classification method and then identifies their breed. Using the flat approach, the breed is obtained directly. The breed classification is based on both the shape and appearance of the animals in the images. It is important to mention that mixed breeds were not considered in this work.

Kenneth et al. [14] used deep neural networks to determine dog identity. They used both the photos and extra information about dogs, gender, and breed, to help with identification. Dog breed classification is performed prior to the identification of each individual dog. Kenneth et al. reported that using breed classification before identification not only reduced the search space but also helped to improve the accuracy by 6.2%. The accuracy based on photos only was 78.09%, which was improved using “soft” biometric information to 84.94%. The approach of Kenneth et al. only considered dogs and only two breeds. However, in practice, multiple breeds and mixed breeds of dogs exist, making this approach not practical. Moreover, lost pets usually hide or run away from people who they are not familiar with, making it hard to capture close photos of the full front face. However, the dataset that was used contained animal images featuring full front faces looking at the camera. Additionally, it is possible that part of their face or body might not be visible.

Using Flickr-dog and Snoopybook datasets, Thierry et al. [15] contrasted four human recognizers, named EigenFaces [16,17], FisherFaces [18], Sparse method [19], and LBPH [1], to two convolutional neural networks, WOOF and BARK, to address the problem of dog identification. WOOF has a deeper architecture than BARK. The results showed that human facial recognizers performed poorly for dogs. However, the two CNNs can potentially be a proper solution for cat identification based on their faces.

3. Data

The data that we have collected has been divided into two groups: classification and identification data. Publicly available datasets such as Flickr-dog [15] could not help us with addressing the full body identification problem since the images are usually from full front faces limited to only a few breeds. Additionally, finding data for individual cats was even more challenging than for dogs due to limited past studies on them. Accordingly, we introduce our new dataset. The data was collected from different sources. Each source was used for a specific task.

COCO Dataset. Most of our data was obtained from this dataset. Microsoft Common Objects in Context (COCO) (<https://cocodataset.org>, accessed on 3 May 2023) is a large-scale dataset of 328,000 images that contain objects. We only used images under the dog and cat categories. The number of images in this dataset in the cat category is 4298, and in the dog category, it is 4562. We have used only 1669 images from the cat category and 1952 images from the dog category. The reason for not using all the images is due to having enough for training, as we used other datasets as well, which are mentioned below. All of the data were used for training purposes. Animals in this dataset are not perfectly visible, be it their bodies or faces. This is the case for most images in this dataset.

Flickr-dog dataset. The Flickr-dog dataset [15] is specifically for dogs. It consisted of 374 images. These photos reside in 42 classes, each having at least 5 images of a single animal.

Pet owners. We asked pet owners to upload photos of their pets via a form. This helped to make the results closer to the real-life use case. We were able to collect more than 1100 images, creating 70 pet profiles.

BC SPCA. Most of the lost pets are brought to animal shelters. The British Columbia Society for the Prevention of Cruelty to Animals (BC SPCA) (<https://spca.bc.ca>, accessed on 3 May 2023) is one of the organizations that operate animal shelters. We were able to use some of the available data that they provided for us. A total of 57 classes of individual animals for identification and more than 1800 pictures of cats and dogs were obtained from BC SPCA.

Crawled: In addition, 521 images of 76 animals, for identification purposes, were crawled or downloaded from the Internet.

Kaggle: Some data for testing and evaluating was obtained from the Kaggle website (<https://www.kaggle.com>, accessed on 3 May 2023). We used mostly the Asirra dataset (<https://kaggle.com/competitions/dogs-vs-cats>, accessed on 3 May 2023), which contains 25,000 images of dogs and cats. Only 5411 images of the Asirra dataset were used for testing the classification model.

Columbia Dogs Dataset. Each image in Columbia Dogs Dataset [20] is annotated by the coordinates of eight key points. The location of these eight points is dependent on the fur, pose, or breed of the dog. The right eye, left eye, nose, right ear tip, right ear base, top of the head, left ear base, and left ear tip are the locations of the key points. The size of this dataset is 8351. In this dataset, 3575 images were used for testing and 4776 for training.

Stanford Dogs Dataset: This dataset is built using annotation files that highlight the body of the dog in each image. It includes 20,580 images split into 12,000 images for training and 8580 for testing. We used all 12,000 images for training in animal detection and classification.

Data Augmentation. Aside from the collected data mentioned above, data augmentation techniques were applied to the identification dataset. This technique adds slightly modified copies of images we have and helps prevent overfitting. The augmentation technique used is the combination of rotation, flipping, and cropping techniques, as well as change in brightness, contrast, and saturation of the images.

Sample images of cats and dogs from all datasets are shown in Figure 1. Since the data is collected from different sources, integrating it in a consistent format is essential. Some of the data was in the video format as opposed to being stills. For this data, we have extracted

the frames from different time slots in the videos. The following steps were taken in the data preparation stage, which are as follows:

1. **Filtering.** After collecting the dataset, irrelevant photos or videos were removed, and only the ones where the whole or partial body of the animals was present were selected for both classification and identification;
2. **Categorizing.** In this stage, cats and dogs were separated. In addition, photos of the same animals were put together in the same directory. Additionally, the dataset is split into test and train categories. For species classification, 17,421 images were used for training, and 5411 for testing;
3. **Labeling.** Each class of animals was labeled. The labels at this stage were just cats and dogs. Each animal was assigned an ID that was unique to the animal itself. We assigned a directory for each animal and the name of the directory is the ID assigned to that animal. Additionally, the name of the animal along with other extra information, including breed, age, and gender, were added to the animal's profile. We created two different CSV files for cats and dogs. Each CSV file contains information about the animal, including their ID, breed, age, and gender. If no extra information is provided for the animal, for example, their age is not applicable, the respective entry would be a NaN value, which is a numeric data type with an undefined value;
4. **Structuring.** Defining a logic for the algorithm to access the dataset was performed at this stage. Moreover, in the identification data, we considered the worst-case scenario in which only one image of an animal is registered in our database. Although multiple images of an animal can be registered in its profile to increase the chances of finding a match, we only considered one such image. The image taken from the found animal is called a query image. Multiple images can be query images. However, in the worst-case scenario, the query image would be compared only with one registered image. This is so that we are aware of the performance of the model in the worst-case scenario. We used the remaining data from our model for testing;
5. **Preprocessing.** In the data preprocessing stage, normalizing and resizing techniques were applied.

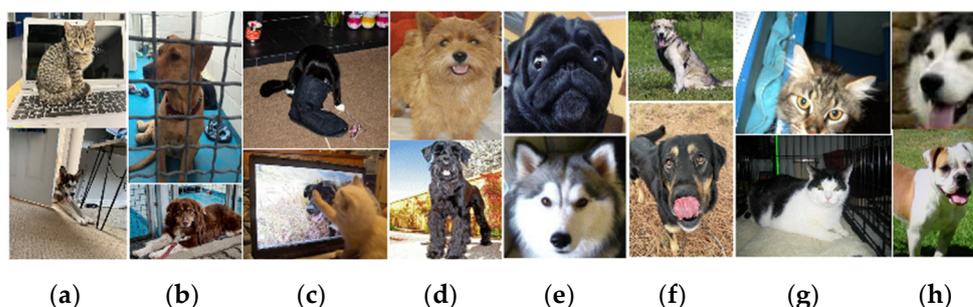


Figure 1. Sample images from different sources of the collected dataset. Each column represents one source: (a) Pet Owners, (b) BC SPCA, (c) COCO dataset, (d) Stanford Dogs, (e) Flickr-dog, (f) Crawled, (g) Kaggle, and (h) Columbia Dog dataset. (Figure courtesy of Devyn Kelly, Elham Azizi, BC SPCA.)

4. Methodology and Implementation

In this work, we have followed the foundational methodology for data science by John Rollins (<https://www.linkedin.com/in/john-b-rollins-ph-d-115460/overlay/50681155/single-media-viewer>, accessed on 3 May 2023), which is an iterative process beginning from Business Understanding to deployment and feedback. The framework consists of different layers. Each layer includes different models and algorithms. The overall architecture of the framework is demonstrated in Figure 2. Each of the layers uses different algorithms and trained models. After the data preprocessing layer as the first layer, the following layers are assigned for a specific task:

1. **Animal Detection and Species Classification.** In this layer, the animal is detected, and its class is recognized as either cat or dog (Figure 2a). For training, COCO, BCSPCA, and Stanford dog dataset were used. We used Convolutional Neural Networks (CNN) as the classification algorithm [21]. We trained the data with YOLO (You Only Look Once) V4 [22,23] with transfer learning to detect animals and perform classification. YOLO V4 is a pre-trained one-stage object detection network, consisting of 53 convolution layers. We have applied transfer learning with pre-trained weights on the COCO dataset and then trained the model on our new data, starting with YOLO's initial weights. The images of the detected animal with the use of segmentation techniques were cropped and connected to the next layers for identification purposes. By doing so, the noise in the image is reduced, and we only consider the animal part in the photos for identification. Animal detection is a crucial component that enhances identification efficiency, as it reduces the search space and minimizes the time required for identification. Furthermore, the classification step can be bypassed if the animal finder inputs the species into an API that connects user inputs to the framework. Our research is designed to be integrated into an app, which means that if the user specifies the species, the system will skip the classification step and proceed to step 2;
2. **Face Detection and Land Marking.** We used a pretrained MobileNetV2 model for face detection for both cats and dogs. The detected faces were cropped and transferred to the face alignment layer (Figure 2b). The rest of the animal's body was then considered for body identification. Additionally, if no face was detected, the image would be directed to the body section, which only applies feature extraction. Both the face and body identification run concurrently afterward. The number of key points used for landmarking for cats is five (two eyes, two ears, and the nose), and for dogs is six (two eyes, two ears, the nose, and the top of the head). Figure 3 illustrates these points. There is no best number of points for landmarking for cats and dogs, but since cats mostly have their ears upwards, having one key point at the top of the head does not make a significant difference. Meanwhile, due to the variation in breeds of dogs, one extra key point (key point number 1) is helpful in cases where ears are downwards;
3. **Face alignment.** Based on the labeled dataset, we elaborated on the above; we consider the eyes of the animals to be the baseline for face alignment (Figure 2c). We then calculate the angle (β) between the x -axis and the line connecting the animal's eyes. Based on this angle, we rotate the image to align the eyes horizontally. Figure 4 illustrates an example of this procedure. We used homography transformation in the OpenCV (<https://opencv.org>, accessed on 3 May 2023) library. The homography function maps points in one image, in our case, eyes and other landmarks, to the corresponding points in the other image, which are the aligned images in our scenario. The homography function takes into account detected landmarks and applies the appropriate transformation. As a result, the homography function is performed based on the specific features of each animal. We compared homography with *getAffineTransform* (https://docs.opencv.org/3.4/d4/d61/tutorial_warp_affine.html, accessed on 3 May 2023), *getRotationMatrix2D*, and *getPerspectiveTransformation* (https://docs.opencv.org/3.4/da/d54/group__imgproc__transform.html#gafbbc470ce83812914a70abfb604f4326, accessed on 3 May 2023), which are transformation functions in OpenCV. We observed that the homography function maps and aligns images more accurately than the others;
4. **Feature Extraction.** After obtaining animals' faces and bodies, we transform these images into numerical features (Figure 1d);
5. **Comparison.** Face alignment techniques provide us with a baseline for comparing different animal faces together. However, due to numerous possible poses of an animal's body captured from various angles, it is cumbersome to define landmarks for each pose for each animal. In turn, this will negatively affect the time and efficiency of the algorithm. Accordingly, the alternative solution is to encourage animal owners

to upload more images of their animals to their profile to increase the possibility of having a similar pose to the query image of the lost pet. Thus, the comparison of the faces is performed after alignment, and the comparison of the body is performed without alignment (Figure 1e). We used cosine similarity. This measurement is defined in Equation (1) as follows:

$$S_c(x, y) := \cos \varnothing = \frac{x \cdot y}{\|x\| \|y\|} \tag{1}$$

In this equation, x and y are vectors. The dot product of the vectors is divided by the cross product of the length of the vectors. The smaller the distance, the closer the two animals are, which means that the possibility of matching the query image with the suggested one by the model is higher;

6. Recommendations. The top N results of the comparison are presented to this layer (Figure 1f). The top match means that the similarity value of comparison from the previous step is the highest. We searched the database to find the top animals' names with the IDs found and recommend them to the user who queried the image. In our case, having false positives is more useful than obtaining more accurate results since we are recommending the top possibilities of the match to an external entity to confirm.

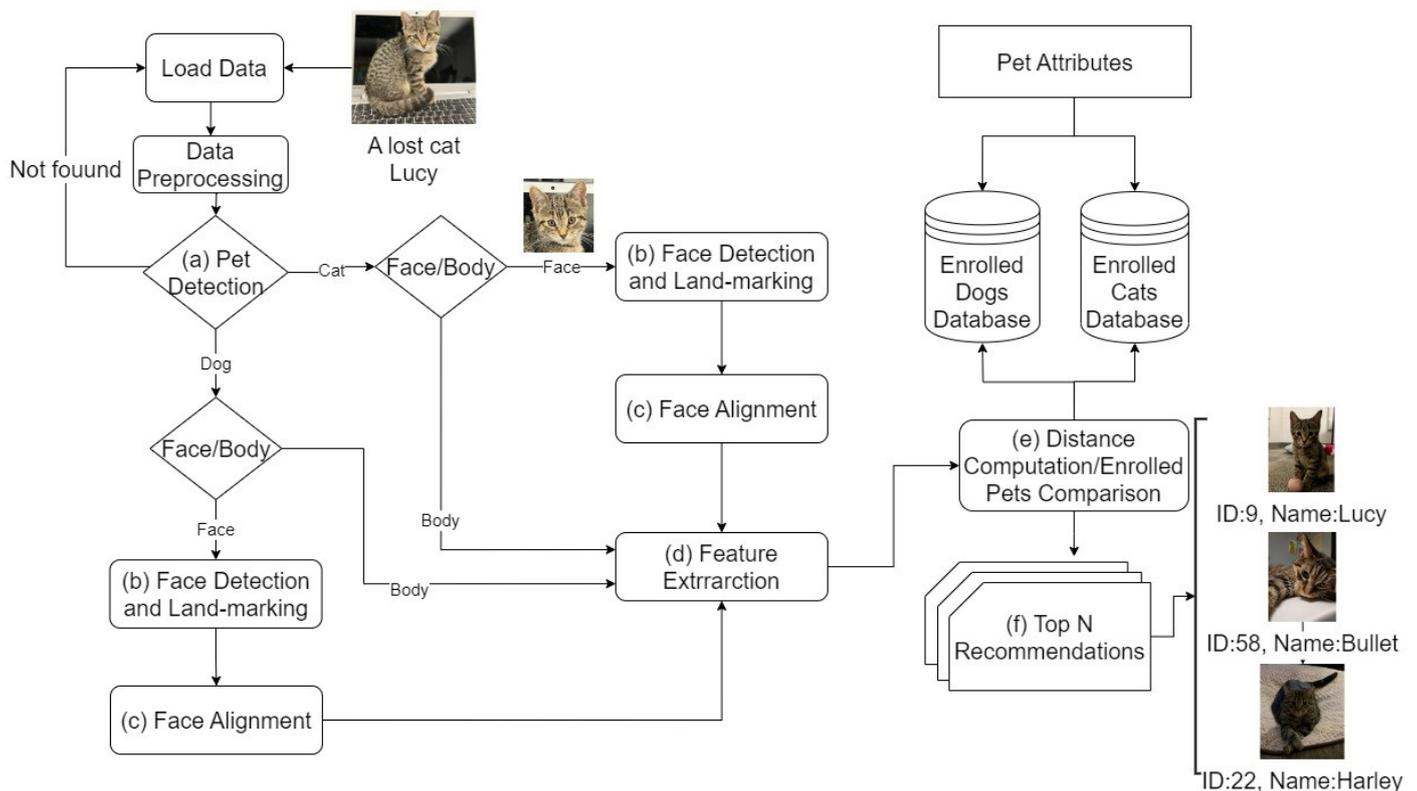


Figure 2. Framework Architecture (Figure courtesy of Elham Azizi, Alison Jones, and Jennifer Neundorf).

To evaluate classification and comparison techniques in machine learning, the following metrics are normally used:

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$recall = \frac{TP}{TP + FN} \tag{3}$$

$$f1_score = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (4)$$

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

where True Positive (TP) is the number of correctly predicted dogs, False Negative (FN) is the total number of dogs incorrectly classified as cats, True Negative (TN) is the number of correctly predicted cats, and False Positive (FP) is the total number of cats predicted as dogs. Precision points out how accurate a model is when it comes to images that are predicted as dogs, while recall calculates how many of the actual dogs are labeled as dogs. Accuracy is a measure of correctly predicted labels in comparison to all the predictions. Our framework was evaluated based on *precision*, *recall*, *f1_score*, and *accuracy* metrics.



Figure 3. An illustration of landmarking key points for cats and dogs. (Figure courtesy of Devyn Kelly).

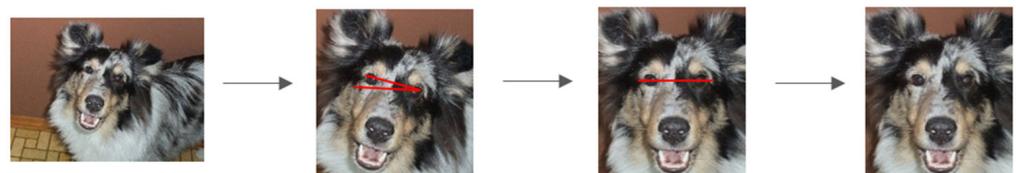


Figure 4. An example illustration of the face detection, cropping, and alignment of the model. (Figure courtesy of J. Marie Brown).

5. Experimental Results

5.1. Species Classification

In the animal detection and species classification layers, the COCO dataset's cat and dog category, BC SPCA, Stanford Dog Dataset, and augmentation techniques were used for training. Images collected from Kaggle were used to test the model.

Table 1 shows the confusion matrix of the classification on the COCO dataset. We achieved 85.92% accuracy in training and 98.18% in testing. It can be inferred that for both classes, we maintain high precision and recall throughout most of the range of thresholds. The classification reports are also shown in Table 2 in detail. The difference in the accuracy of training and testing, shown in Table 2, is due to the difference in the source of the datasets. As stated above, the COCO dataset was a strict data for training, since there are

many images in which the animal is not clearly visible or just a small part of the animal is detectable. In contrast, the images from the *Kaggle* dataset show more of the animal and less noise.

Table 1. Confusion Matrix of the animal species classification layer.

		Predicted		
		Dog	Cat	Other
Actual	Dog	2679	58	
	Cat	40	2634	

Table 2. Performance measurements for training and testing in the species classification layer.

Metric	Result
Training accuracy on COCO dataset	85.92%
Testing accuracy on Kaggle dataset	98.18%
Testing precision on Kaggle dataset	97.88%
Testing recall on Kaggle dataset	98.52%
F1 score on Kaggle dataset	98.19%

Table 2 shows performance metrics for training and testing in the species classification layer which are *precision*, *recall*, *f1-score*, *training accuracy*, and *testing accuracy* (for both COCO and Kaggle datasets).

5.2. Identification and Recommendations

The Flickr-dog dataset, pet owners' collected dataset, BC SPCA individual animals' data, and crawled open-source images were used for recognizing and identifying the animals. Furthermore, data augmentation techniques were applied. Before any comparison, the animal must be registered to the database, and an ID must be assigned to it. This ID is unique and differentiates animals from each other. The registration is completed only using a single image of the animal, which is the image that has the highest probability of matching with the queried image. The total number of individual pet profile IDs in the database is 245.

At first, the algorithm is tested without any animals' soft biometric information (gender, breed, age, and species). However, we have compared the results of both using animal images alone and together with their soft biometric information. Additionally, the results of using face and body separately and together have been compared to reveal the effectiveness of using body features for identification.

In the first phase of the experiment, the algorithm detects the animal face and applies preprocessing such as normalization methods before performing face alignment and landmark detection. If any face is detected, the identification model compares the cropped image of the face with the whole database via feature extraction. The model compares the animal with the registered images and assigns a probability of finding a match. The model assigns a probability number to each animal comparison and, in the end, sorts the probabilities in ascending order. The model then returns the image from the most probable animal to the least probable. Only the top 10 best matches are returned via the algorithm as a recommendation for the user. The results of the predictions based on face recognition, shown in Table 3, demonstrate that the accuracy of the system is 80% if the first recommendation is the same animal. If the animal is second probable, the model's performance indicates that in 14% of the predictions, the animal is the second returned ID. In other words, the chance that the animal is found by the top two recommendations is 94%. Similarly, the third prediction is 3%, and the remaining 3% indicates that the animal is

the fourth to the tenth identity in our recommendations. As also seen, there are no cases in our predictions that the animal is not among the top 10 recommended identities.

Table 3. Accuracy report on face and body identification for top 10 recommendations.

Nth Recommendation	Face Identification Accuracy	Body Identification Accuracy	Face and Body Identification Accuracy
1	80%	81%	86.5%
2	14%	15%	10.5%
3	3%	2%	2%
4 to 10	3%	2%	1%

It is possible that the algorithm is unable to find the animal’s face because it is not present or only a portion of it is visible. In this case, we enable full-body identification with face matching. Both the face and body identification run simultaneously. Additionally, if only the face of the animal is presented in an image, the algorithm still compares the face with the database but this time without any alignments. In this case, even if only a partial part of the body is visible together with the face, that part of the body will be effective for prediction. The reason is to use as many features as we can obtain from an image. The results of only full body identification do not indicate a significant improvement in the predictions. From Table 3, it can be inferred that the probability that the lost animal is our first returned recommendation is 81%. However, the combination of both face and body identification improves the first recommendation for image prediction up to 86.5%. We calculated the combination of the results in body and face using Equation (5) as follows:

$$P(ID | face , body) = \frac{Score(ID; face) \cdot Score(ID; body)}{N(face, body)} \tag{6}$$

N in Equation (6) is our normalization factor. We further improved the identification process with the use of animal soft biometric information (gender, species, age, and breed).

The results show an improvement from 86.5% to 92% accuracy for the first recommendation. This means that 5.5% of the predictions, which were not the first recommendation, are now the best match due to the use of soft biometric data. It is worth mentioning that only 40% of our dataset contains extra information about the animal. So, 60% of the individuals did not provide any other information besides the images. It can be suggested that if these individuals had provided their soft biometric properties, the predictions would have been more accurate and precise. The likelihood of the identity prediction is computed as follows for the whole framework:

$$P(ID|face, body, g, b, s, a) = \frac{Score(ID; face) \cdot Score(ID; body) \cdot f(id, g) \cdot f(id, b) \cdot f(id, s) \cdot f(id, a)}{N(face, body)} \tag{7}$$

In this equation, f indicates a function, which is defined as follows:

$$f(x, y) = \begin{cases} 1, & \text{if } x \text{ matches } y \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

where x is the ID of the recommended animal and y is the type of attribute. In Equation (8), g represents gender, b is the breed, and a is the age of the animal.

Our study employs body identification in addition to face and soft biometrics and utilizes different methodologies in framework architecture. We compared our findings with the most recent research on pet identification conducted by Kenneth et al. [14], and the summary of this comparison can be found in Table 4. According to the table, our framework achieved an accuracy of 80% for the top similarity match for both cat and

dog faces, whereas [17] only achieved 78.09% accuracy for dog faces. Furthermore, our framework achieved a soft biometrics accuracy of 92%, compared to 84.94% in [14]. Apart from accuracy, our framework is more practical due to its ability to identify animals in different poses, making it more adaptable in real-world settings. Figure 1 provides examples of various animal poses. In addition to the face detection layer, the body identification layer performs feature extraction on the animal's body and compares the resulting values with those stored in the database. Extracting the features is not dependent on the poses since there would be no need for body alignment due to the wide range of possible poses an animal can adopt.

Table 4. Performance comparison.

Study	Metric	Identification Accuracy Using Face	Overall Accuracy (Soft Biometrics)
Method by Kenneth et al. [14]		78.09%	80%
Our method		84.94%	92%

6. Conclusions

In this work, a new pet identification system was proposed. The framework consists of seven layers: data preprocessing, species classification, landmark detection and face/body detection, alignment, identification, soft biometrics, and recommendation. We further improved the performance by applying pet information such as gender and breed. Using DNNs, we were able to classify cats and dogs with 85.92% accuracy in training and 98.18% accuracy in testing. We were able to match animals using their face with 80% confidence along with 100% confidence that the animal would be found via the system in the top 10 recommendations. For the face, landmarking and face alignment techniques were applied to increase the likelihood of the match. Using only animals' bodies, we achieved 81% accuracy and a 100% likelihood for the animal to appear in our top 10 recommendations. The combination of face and body identification improved the performance of the system by up to 86.5%. Using animals' soft biometric information, 92% accuracy was achieved for the overall framework. The system's performance can be attributed to the integration of various layers for face, body, and soft biometrics identification. By utilizing face alignment techniques, the face identification layer was able to establish a baseline of comparison, while the body identification feature extracted more general features of the animal's body, thereby contributing to the system's enhanced performance. We believe the framework's performance holds promise for pet identification and can enhance the Return-To-Owner rate of lost pets. The system can automate the search and find procedure, which is typically conducted manually by pet owners or in shelters. To assist in reuniting lost pets with their owners, the system will be integrated into an app with a user-friendly interface.

7. Limitations and Future Work

The current system is tested on the collected dataset from different sources such as shelters. However, in the future, the system should be tested on larger datasets of lost animals from shelters to verify the performance in real use-case scenarios. Additionally, having a high similarity probability in the body section is dependent on the presence of an image of the animal with a similar pose in the database. The feature extraction can identify an animal from one side properly, even if the pose from that side of the animal has changed. However, if the camera angle has completely changed, e.g., from the left side of the body to the back of the animal, the system is unable to accurately match the animal. At the production level, when the algorithm is used in practice by users, it is highly encouraged to upload as many photos as possible. This is also a strategy to follow for non-shedding dogs that change significantly when lost. On the contrary, a high volume of images means an increase in prediction time. One way to tackle this is to use soft biometrics before the query and filter the animals with the same information. The other solution that can be

pursued in the future is using batch predictions. In order to improve the performance of the framework, nose print identification [24] and voice recognition [25] can be included in the future. Nose prints are unique to each cat or dog and can positively affect true positive or true negative predictions. In the future, capsule networks [26] can be used in the system to add more flexibility in the body poses, as well as nose prints. Voice recognition can be useful in scenarios where the images are not available, e.g., in the darkness, or if animals hide when lost or are scared. Additionally, in the future, more specifications can be explored in addition to dogs and cats. Finally, at present, there is limited research directly targeting pet identification to the extent of our proposed framework. Therefore, we were unable to provide more performance comparisons to fully showcase the effectiveness of our body identification method. In the future, different models can be compared within the same proposed method. This could potentially lead to improved performance.

Author Contributions: Conceptualization, E.A. and L.Z.; methodology, E.A. and L.Z.; software, E.A.; validation, E.A.; formal analysis, E.A.; investigation, E.A. and L.Z.; resources, E.A.; data curation, E.A.; writing—original draft preparation, E.A. and L.Z.; writing—review and editing, E.A. and L.Z.; visualization, E.A.; supervision, L.Z.; project administration, L.Z.; funding acquisition, L.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by a Mitacs Accelerate grant to which Petnerup Inc. contributed. Funding request reference FR75730.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available as it is owned by Petnerup Inc.

Acknowledgments: We would like to extend our sincere gratitude to Kwene H. Low (khlowe-mail@gmail.com) and Petnerup Inc. for initiating this research and providing support and assistance with conceptual directions. This work also acknowledges BC SPCA's support in dataset collection.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Weiss, E.; Slater, M.; Lord, L. Frequency of Lost Dogs and Cats in the United States and the Methods Used to Locate Them. *Animals* **2012**, *2*, 301–315. [CrossRef] [PubMed]
2. Zafar, S. The Pros and Cons to Dog GPS Trackers-Blogging Hub. 12 December 2021. Available online: <https://www.cleantechloops.com/pros-and-cons-dog-gps-trackers/> (accessed on 19 February 2023).
3. Hayasaki, M.; Hattori, C. Experimental induction of contact allergy in dogs by means of certain sensitization procedures. *Veter. Dermatol.* **2000**, *11*, 227–233. [CrossRef]
4. Albrecht, K. Microchip-induced tumors in laboratory rodents and dogs: A review of the literature 1990–2006. In Proceedings of the 2010 IEEE International Symposium on Technology and Society, Wollongong, Australia, 7–9 June 2010; pp. 337–349. [CrossRef]
5. Kumar, S.; Singh, S.K. Biometric Recognition for Pet Animal. *J. Softw. Eng. Appl.* **2014**, *7*, 470–482. [CrossRef]
6. Dryden, I.L.; Mardia, K.V. *Statistical Shape Analysis: With Applications in R*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2016. [CrossRef]
7. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269. [CrossRef]
8. Cao, Z.; Principe, J.C.; Ouyang, B.; Dalgleish, F.; Vuorenkoski, A. Marine animal classification using combined CNN and hand-designed image features. In Proceedings of the OCEANS 2015-MTS/IEEE Washington, Washington, DC, USA, 19–22 October 2015; pp. 1–6. [CrossRef]
9. Komarasamy, G.; Manish, M.; Dheemanth, V.; Dhar, D.; Bhattacharjee, M. Automation of Animal Classification Using Deep Learning. In *Smart Innovation, Systems and Technologies, Proceedings of the International Conference on Intelligent Emerging Methods of Artificial Intelligence & Cloud Computing*; García Márquez, F.P., Ed.; Springer International Publishing: Cham, Switzerland, 2022; pp. 419–427. [CrossRef]
10. Norouzzadeh, M.S.; Nguyen, A.; Kosmala, M.; Swanson, A.; Palmer, M.S.; Packer, C.; Clune, J. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E5716–E5725. [CrossRef] [PubMed]
11. Qiao, Y.; Su, D.; Kong, H.; Sukkarieh, S.; Lomax, S.; Clark, C. Individual Cattle Identification Using a Deep Learning Based Framework. *IFAC-PapersOnLine* **2019**, *52*, 318–323. [CrossRef]

12. Barbedo, J.G.A.; Koenigkan, L.V.; Santos, T.T.; Santos, P.M. A Study on the Detection of Cattle in UAV Images Using Deep Learning. *Sensors* **2019**, *19*, 5436. [[CrossRef](#)] [[PubMed](#)]
13. Parkhi, O.M.; Vedaldi, A.; Zisserman, A.; Jawahar, C.V. Cats and dogs. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 3498–3505. [[CrossRef](#)]
14. Lai, K.; Tu, X.; Yanushkevich, S. Dog Identification using Soft Biometrics and Neural Networks. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–8. [[CrossRef](#)]
15. Moreira, T.P.; Perez, M.; Werneck, R.; Valle, E. Where is my puppy? Retrieving lost dogs by facial features. *Multimed. Tools Appl.* **2017**, *76*, 15325–15340. [[CrossRef](#)]
16. Turk, M.; Pentland, A. Eigenfaces for Recognition. *J. Cogn. Neurosci.* **1991**, *3*, 71–86. [[CrossRef](#)] [[PubMed](#)]
17. Sirovich, L.; Kirby, M. Low-dimensional procedure for the characterization of human faces. *J. Opt. Soc. Am. A JOSAA* **1987**, *4*, 519–524. [[CrossRef](#)] [[PubMed](#)]
18. Belhumeur, P.; Hespanha, J.; Kriegman, D. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**, *19*, 711–720. [[CrossRef](#)]
19. Xu, Y.; Zhang, D.; Yang, J.; Yang, J.-Y. A Two-Phase Test Sample Sparse Representation Method for Use with Face Recognition. *IEEE Trans. Circuits Syst. Video Technol.* **2011**, *21*, 1255–1262. [[CrossRef](#)]
20. Liu, J.; Kanazawa, A.; Jacobs, D.; Belhumeur, P. Dog Breed Classification Using Part Localization. In *Lecture Notes in Computer Science, Proceedings of the Computer Vision–ECCV 2012, Florence, Italy, 7–13 October 2012*; Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 172–185. [[CrossRef](#)]
21. ImageNet Classification with Deep Convolutional Neural Networks | Communications of the ACM. Available online: <https://dl.acm.org/doi/abs/10.1145/3065386> (accessed on 12 April 2023).
22. Bochkovskiy, A.; Wang, C.; Liao, H. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934. [[CrossRef](#)]
23. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [[CrossRef](#)]
24. Bae, H.B.; Pak, D.; Lee, S. Dog Nose-Print Identification Using Deep Neural Networks. *IEEE Access* **2021**, *9*, 49141–49153. [[CrossRef](#)]
25. Yeo, C.Y.; Al-Haddad, S.; Ng, C.K. Animal voice recognition for identification (ID) detection system. In Proceedings of the 2011 IEEE 7th International Colloquium on Signal Processing and Its Applications, Penang, Malaysia, 4–6 March 2011; pp. 198–201. [[CrossRef](#)]
26. Hinton, G.E.; Krizhevsky, A.; Wang, S.D. Transforming Auto-Encoders. In *Lecture Notes in Computer Science, Proceedings of the Artificial Neural Networks and Machine Learning–ICANN 2011, Espoo, Finland, 14–17 June 2011*; Honkela, T., Duch, W., Girolami, M., Kaski, S., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 44–51. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.