Computational Sciences and Engineering 4(2) (2024) 259-268

University of Guilan

Computational Sciences and Engineering

journal homepage: https://cse.guilan.ac.ir/

Nose-to-ID: A Deep Learning Framework for Dog Identification Using Nose-Print Biometrics

Hediyeh Mosafer [a], Homa Taherpour Gelsefid [a], Rana Ghozat [a], Armin Azhdehnia [a], Javad Ghofrani [b], Ehsan Kozegar [a,*]

[a] Department of Computer Engineering, Faculty of Technology and Engineering- East of Guilan, University of Guilan, Guilan, Iran
[b] Department of Computer Science, University of Applied Sciences Bonn-Rhein-Sieg, Germany

ARTICLE INFO

Article history:
Received 6 June 2025
Received in revised form 22 June 2025
Accepted 25 June 2025
Available online 25 June 2025

Keywords:
Dog identification
Pet identification System
Nose-print biometrics
Dog nose print
Animal re-identification
Dog Nose Detection
Biometric retrieval

ABSTRACT

Accurate identification of individual dogs plays a crucial role in various applications including pet recovery, veterinary management, and animal welfare. This study proposes a fully automated dog identification framework based on unique nose-print biometric patterns, leveraging deep learning techniques to overcome limitations of traditional identification methods. The proposed approach processes user-submitted videos by selecting the most frontal frame via head pose estimation, detects the nose region using a fine-tuned YOLOv8 model, and extracts discriminative embeddings through a multi-resolution ResNeSt-based convolutional network enhanced with advanced augmentation strategies. The resulting embeddings are fused to produce robust identity descriptors capable of distinguishing between thousands of individual dogs. Experimental results demonstrate the system's efficacy under real-world conditions, emphasizing its potential for practical deployment in pet identification and management systems.

1. Introduction

In recent years, the accurate identification of individual animals, particularly domestic pets like dogs, has become an increasingly relevant task across multiple domains including pet recovery, veterinary services, and wildlife monitoring. Among these applications, dog identification based on visual features—such as facial structure or nose patterns—offers a promising biometric alternative to traditional identification methods like ID tags or implanted microchips, which are prone to being lost, removed, or damaged.

* Corresponding author.
E-mail addresses: kozegar@guilan.ac.ir

https://doi.org/10.22124/cse.2025.30957.1110

The need for reliable and scalable dog identification emerges in a variety of real-world scenarios. For instance, when a dog goes missing or is stolen, pet owners and animal control authorities often lack an efficient means to verify the animal's identity unless biometric information has already been registered. Additionally, animal shelters and veterinary clinics must manage and differentiate large populations of dogs, often with limited or no prior identification data, emphasizing the need for automated and accurate identification systems.

The motivation behind this work is to develop a deep learning-based visual recognition system capable of identifying individual dogs using unique and consistent visual cues—primarily the dog's nose or facial features. Such a system could be invaluable in multiple applications, including automated check-in procedures at veterinary clinics, digital pet management platforms, and, most critically, in the recovery of lost pets.

Recent studies have demonstrated the feasibility of using convolutional neural networks (CNNs) for pet identification tasks. For instance, a CNN-based approach utilizing soft biometrics—such as breed, color, and body shape—has been proposed to identify dogs in a non-invasive manner, achieving high accuracy through the combination of soft and hard biometric cues [1]. Similarly, another study introduced a deep learning system that uses facial features to retrieve lost dogs, highlighting the promise of CNNs over traditional methods for canine face recognition [2].

Although much work has focused on pet identification, techniques developed in livestock and wildlife contexts are also informative. A deep learning framework for cattle identification based on facial and body characteristics shows the generalizability of these methods across species [3]. In line with this, the fusion of facial and body features was shown to significantly improve pet recognition accuracy, supporting the need for multimodal biometric fusion [4].

More advanced approaches have also been explored. For instance, similarity learning methods such as Siamese and Triplet networks have been successfully applied for individual animal re-identification, even under occlusion or poor lighting conditions [5]. Likewise, contrastive learning-based transformers have recently been proposed for the retrieval of lost pets, leveraging visual embedding spaces to match missing animals with registered images [6].

Despite significant advances, many existing dog identification methods rely heavily on full-body features, breed classification, or require multiple views, which limits their effectiveness in real-world scenarios where only a single partial view—such as a close-up of the dog's nose—is available. To address this limitation, the proposed framework introduces a fine-grained, fully automated dog identification framework centered on the unique biometric patterns of the nose. Scientific evidence suggests that nose prints in dogs are as unique as human fingerprints, with features such as the shape of the nostrils, skin texture, wrinkles, and pigmentation forming a distinctive identity. Leveraging this, the proposed method applies a deep embedding network to generate compact, discriminative feature vectors that encode these visual patterns, enabling reliable identity matching. Unlike previous methods that depend on high-quality, pre-captured images, the pipeline operates directly on user-submitted videos. It incorporates head pose estimation to extract the most frontal frame, uses a dedicated nose detection model to localize the biometric region, and applies the embedding model to transform the nose image into a meaningful representation for comparison. This end-to-end approach is designed to function under real-world conditions, effectively bridging the gap between academic research and practical deployment.

2. Materials and Methods

2.1. Dataset Description

For training the nose detection model, we utilized a custom dataset consisting of facial images of various dogs, structured across five main directories named DogData1 to DogData5. Each of these directories contained multiple subfolders—340 in total—where each subfolder corresponded to a unique individual dog. Inside each subfolder, several images of that dog were stored, summing up to a total of 2,505 dog images across the entire dataset.

The images were originally collected from Instagram, featuring popular and well-known dogs whose accounts were publicly accessible. These dogs had multiple frontal and profile images posted by their owners, making them suitable candidates for structured visual datasets.

Since the original images were not annotated, we manually annotated a selected subset of the dataset using Label Studio, focusing specifically on labeling the nose regions in each image. The annotated portion was then used to fine-tune a YOLOv8-based object detection model for nose localization.

To evaluate the performance of the embedding-based identification model, we employed a modified version of the dataset originally introduced in the "CVPR2022 Biometrics Workshop – Pet Biometric Challenge." All images were preprocessed by resizing, augmentation, and cropping to extract only the nose region, with region sizes ranging from approximately 70 to 600 pixels, while maintaining the original resolution. The original dataset consisted of around 6,000 individual dog identities and 20,000 nose-print images, accompanied by a metadata CSV file. As the original challenge dataset is no longer publicly accessible, we utilized an enhanced open-access variant referenced in prior work for the embedding evaluation experiments [7].

To evaluate the performance of the embedding-based identification model, we constructed a dedicated test dataset derived from the same collection of dog facial images originally sourced from public Instagram profiles. The previously trained nose detection model was applied to all available images in this collection to automatically localize and crop the nose region from each dog image.

The resulting cropped nose images were then used to generate all possible pairwise combinations, forming a set of image pairs. These pairs were organized into a CSV file with two columns, A and B, each containing the path to one of the two nose images in the pair. Additionally, a target column was included to indicate whether the two images in each pair belonged to the same individual dog (matched) or to different dogs (unmatched).

This labeled pairwise dataset enables quantitative evaluation of the embedding model's ability to distinguish between different dogs and correctly identify instances of the same dog, serving as a reliable benchmark for retrieval-based biometric identification performance.

2.2. Methods

To provide a general overview of the system, the process begins when a user uploads a video through the client mobile application or dashboard. The uploaded video is forwarded to the backend services where a head pose estimation module processes the video frames and extracts the most frontal face image. This frontal image is then passed to the nose detection module which identifies and returns the coordinates of the nose. These coordinates are subsequently used by the nose embedding model to extract a feature representation. The extracted embedding can then be either used for enrollment or passed to the search pipeline for querying against previously stored embeddings. The overall architecture of this workflow is illustrated in ***Figure 1***.

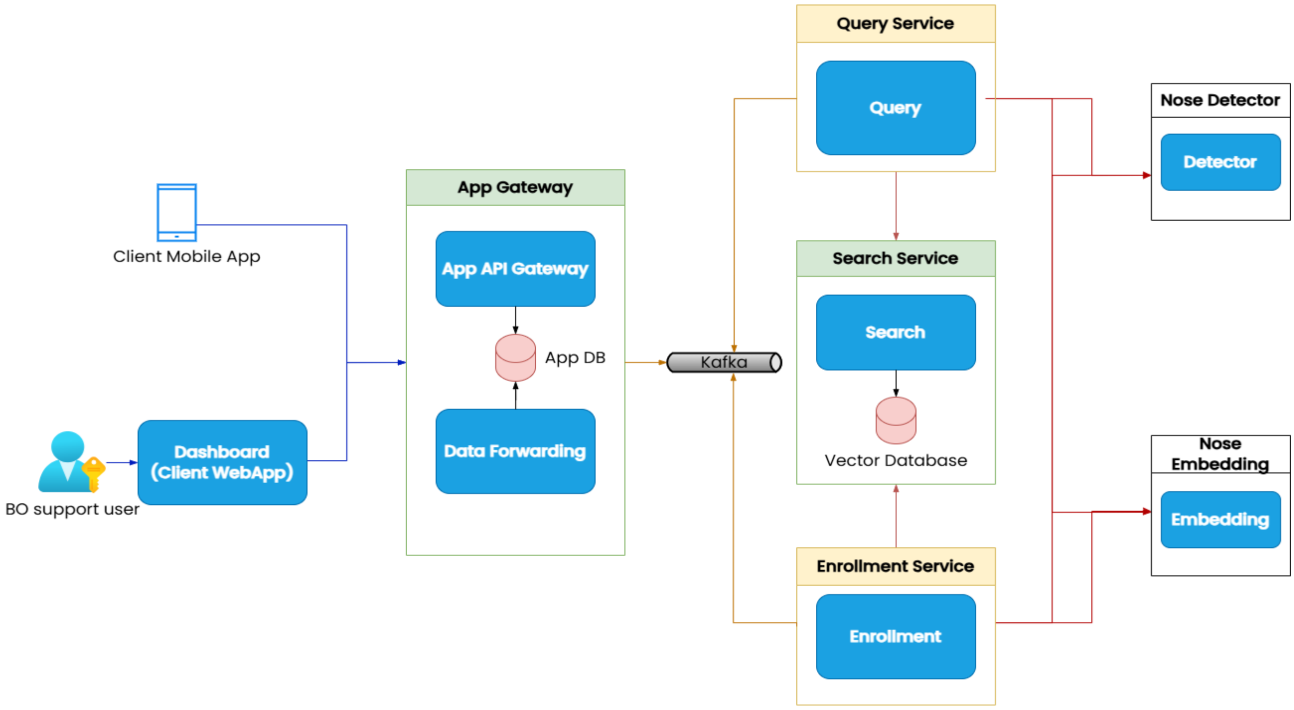

Figure 1. Overall system flowchart illustrating the main modules of the pipeline.

To improve the robustness and practicality of the dog identification system—especially in real-world scenarios where users may not have access to high-resolution frontal images of their pets—the proposed method begins by accepting short video submissions instead of single images. This decision was made to address challenges such as poor image quality and the difficulty of capturing well-aligned frontal shots in uncontrolled environments.

To extract a suitable frame from the video, we utilized head pose estimation to identify the most frontal view of the dog's face. Specifically, we employed dlib’s HOG-based face detector to locate faces in each video frame and used a pre-trained shape predictor to extract 68 facial landmarks. These landmarks, corresponding to key points such as the nose tip, eyes, and mouth corners, were used in combination with a 3D facial model and the cv2.solvePnP() algorithm to compute the yaw, pitch, and roll angles. Among up to 90 analyzed frames per video, the one with the smallest absolute yaw angle—indicating a head orientation closest to frontal—was selected for further processing.

The chosen frame was then passed to a YOLOv8-based nose detection model. We used the lightweight YOLOv8n variant, pre-trained on the COCO dataset [8], and fine-tuned it on our own dataset of annotated dog faces. Our dataset included images manually labeled with bounding boxes around the nose region using YOLO format annotations. The training was conducted using the official Ultralytics CLI interface with the following parameters: task=detect, mode=train, model=yolov8n.pt, epochs=100, and imgsz=640. These settings enabled efficient training for 100 epochs with resized input images of 640×640 pixels, balancing model accuracy with inference speed for real-time deployment.

The output of this module—the detected nose coordinates—is then forwarded to the embedding and identification modules.

In the next stage of the system, a robust embedding-based pipeline was employed for dog nose-print biometric identification. At its core, the pipeline leverages ResNeSt-based convolutional architectures to extract discriminative embeddings for retrieval-based identification (ReID). The design aims to ensure robustness to image resolution, texture variation, and illumination diversity.

To increase training diversity and enhance model generalization across geometric and photometric variations, several augmentation strategies were applied. All input images were resized and augmented using multiple techniques to improve model robustness. The augmentation pipeline included random cropping, horizontal flipping with a probability of 0.5, 10-pixel padding, and Gaussian blur with a (5, 9) kernel. In addition, more advanced augmentation methods such as AutoAugment, AugMix, affine transformations, and color jittering were also applied with a probability of 0.5. During different training stages, the input images were resized to 224×224, 256×256, or 288×288 pixels, depending on the configuration of the model. When the original image resolution was smaller than the target crop size, an internal resize operation was applied before cropping, resulting in a two-stage resizing process that preserved critical visual information.

The embedding model employed in this study was built upon the ResNeSt-101 backbone, a ResNet-based architecture that incorporates Split-Attention blocks to enhance feature representational power [9]. To further improve performance on fine-grained visual tasks, Instance Batch Normalization (IBN) and Non-Local blocks were added. These architectural components enabled the model to better capture subtle nose-print details. Pretrained weights on ImageNet were used to initialize the backbone, and to investigate the effect of network depth and input resolution, additional variants such as ResNeSt-200 were also employed during training.

The embedding pipeline followed a two-stage design. In the first stage, multiple backbone networks trained with different input resolutions were used to extract feature embeddings from the dog nose regions. In the second stage, a lightweight fusion model was introduced to aggregate the extracted embeddings into a single unified descriptor. This design allowed the system to benefit from the complementary information provided by different models.

Feature extraction was handled by four independently trained retrieval models with different architectural configurations: ResNet101 with 224×224 input, ResNet101 with 256×256 input, ResNet101 with 288×288 input, and ResNet200 with 224×224 input. These configurations were selected to systematically explore the influence of network depth and input scale on the quality and discriminability of the learned embeddings. Each model was fine-tuned individually on the annotated nose-print dataset, contributing diverse feature perspectives for fusion.

After the backbone stage, a Generalized Mean (GeM) pooling layer was used to aggregate spatial features. This pooling technique, which learns a pooling parameter, provides a flexible mechanism to generalize both average and max pooling. A Batch Normalization neck (BNNeck) was then added between the backbone and the classification head. Features extracted after the BNNeck were utilized for both identity classification during training and inference during evaluation.

The classification head used a Cosine Softmax layer with a margin of 0.35 and a scale factor of 64, allowing for more discriminative embeddings in the angular space. Although this classification objective was only used during training to guide the learning of the embedding space via cross-entropy loss, it was not used during inference. Instead, embeddings obtained after the BNNeck were directly compared using cosine similarity.

A lightweight fusion model combined the outputs from the four individual networks. This fusion network effectively merged multi-scale and multi-backbone representations into a single descriptor, improving overall retrieval accuracy. As the final stage of the pipeline, it facilitated more precise similarity computation and identity prediction.

The model was trained using a combination of Triplet Loss (with hard example mining), Circle Loss, and Cross-Entropy Loss with label smoothing ($\varepsilon = 0.1$). The models were optimized using Adam with an initial learning rate of 0.00035. A warmup strategy was employed, linearly increasing the learning rate over the first 400 iterations with a warmup factor of 0.1. The backbone parameters were frozen for the first 1000 iterations to allow the classifier head to stabilize. Training was conducted for 35 epochs, with a batch size of 80 for most models and 64 for deeper architectures. Checkpoints were saved every 5 epochs. Mixed precision training (AMP) was enabled for the model using a ResNeSt-200 backbone. The training sampler used the NaiveIdentitySampler with 4 instances per identity to facilitate effective batch hard mining in metric learning. Each data loading worker process was configured with 8 threads.

The overall workflow of the embedding model is illustrated in ***Figure 2***.

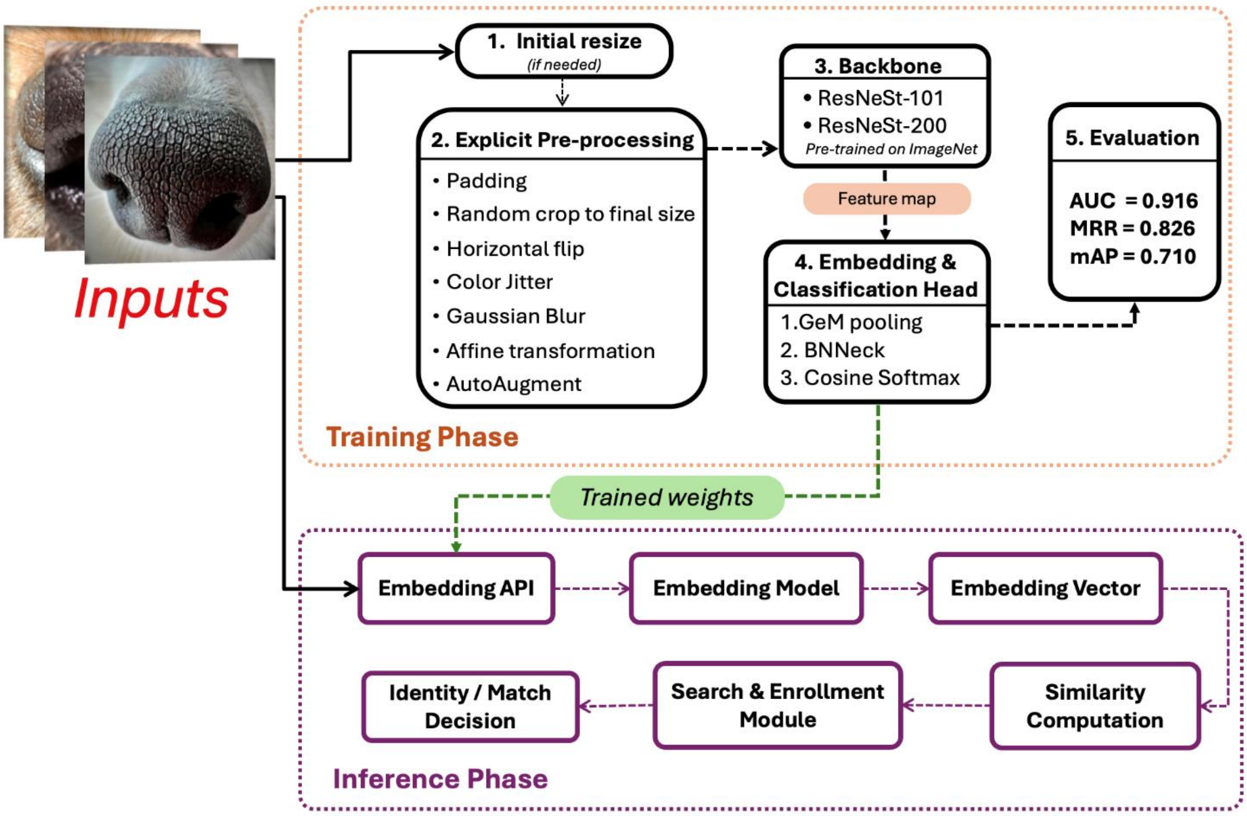

Figure 2. overall workflow of the embedding model

3. Results

To illustrate the effectiveness of the head pose estimation module, we present several representative frames extracted from input videos alongside the selected frontal frame determined by the model in

Figure 3. These samples demonstrate how the module identifies the frame with the most suitable head orientation for subsequent nose detection and embedding.

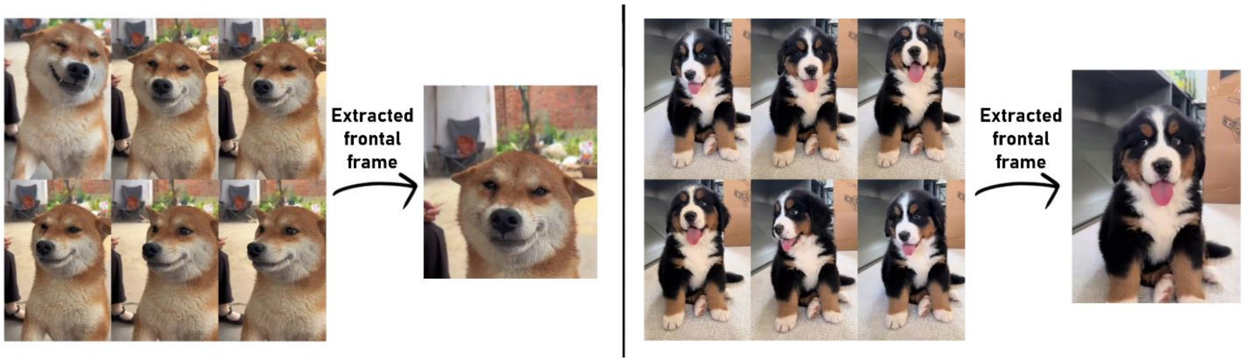

Figure 3. Representative video frames alongside the frontal frame selected by the head pose estimation module

We observed that the quality and nature of the input significantly impact the effectiveness of the system. When provided with a well-lit, high-resolution video of the dog, the head pose estimation algorithm is more likely to identify a truly frontal frame with clear and symmetric nose visibility. This results in a better-quality cropped nose image, leading to more accurate identification in the downstream embedding model.

In contrast, when the input is a single image rather than a video—especially if the image is not perfectly frontal, blurry, or taken in suboptimal lighting—the module often lacks sufficient spatial variation to find a better alternative. Consequently, the performance of the entire pipeline may degrade due to the lower quality of the input to the embedding model.

These observations highlight the importance of using short, high-quality video recordings rather than static images for optimal identification performance in real-world deployment.

To assess the performance of the nose detection model, we applied the YOLOv8-based detector on a set of 1,280 dog images. The model successfully detected the nose in 1,254 cases, among which 30 detections were incorrect (e.g., wrong regions or shifted bounding boxes). This yields a precise detection accuracy of approximately 95.63%.

Failure cases were typically associated with blurred images, poor lighting, or non-frontal head poses. These limitations motivated the integration of a head pose estimation module in the proposed pipeline to improve the selection of more frontal and high-quality frames, thus enhancing the reliability of the detection step.

To evaluate the identification performance of the embedding model, L2-normalized feature vectors were compared using cosine similarity. The resulting similarity scores were linearly rescaled to a 0–100 range for interpretability. Performance was assessed using three standard metrics: mean reciprocal rank (MRR), mean average precision (mAP), and area under the ROC curve (AUC). These metrics were computed per query and then averaged over all queries with at least one ground-truth match.

On the pet re-identification benchmark, the model achieved an MRR of 0.826, mAP of 0.711, and AUC of 0.9157, indicating strong performance in both ranking quality and discriminative power.

These results confirm the model's effectiveness in distinguishing between individual dogs based on nose-print embeddings, validating its applicability in real-world biometric identification scenarios.

To highlight the effectiveness of the proposed approach, we compare its performance with previous methods evaluated on the same nose-print re-identification task. Table X summarizes the AUC scores of several published solutions, including the winning entry of the CVPR2022 Pet Biometric Challenge and a standard VGG-16 baseline implemented for this study. The proposed method, based on a multi-resolution embedding ensemble and fusion strategy, achieves the highest AUC of 91.58%, outperforming all prior work.

The VGG-16 baseline was implemented using the pretrained ImageNet model, extracting feature embeddings from the features block followed by global average pooling. Cosine similarity was used for pairwise comparison, and performance was evaluated using the challenge-provided test CSV.

As shown in ***Table 1***, the proposed model surpasses prior works in terms of AUC, achieving the best performance overall.

Table 1. Performance Comparison with Prior Works

Methods	AUC (%)
Shen et al. [10]	86.70
Li et al. [11]	88.81
Li et al. (HUST) [12]	90.80
VGG-16 Baseline	78.00
The proposed method	**91.58**

4.Discussion

In this project, we presented a new and practical approach to identify individual dogs using the unique patterns on their noses, also known as nose prints. The proposed system is designed to work directly with videos recorded by users, rather than requiring high-quality still images or specialized equipment. It follows a complete step-by-step process: detecting the dog's head and choosing the best frame using head pose estimation, finding the nose using a detection model (YOLOv8), and then generating an embedding vector using a deep neural network.

What makes this method special is that it is fully automatic and easy to use in real-life situations. Many existing methods need carefully taken images or professional cameras, but the presented system can be used with simple mobile phone videos. This makes it useful in places like vet clinics, animal shelters, or pet adoption centers, where conditions may not be ideal. The use of an embedding model trained to capture fine nose patterns helps improve the system's accuracy, even when the input video is not perfect.

Another important strength is that this method focuses only on the dog’s nose, which has been proven to be unique for each dog — similar to fingerprints in humans. By using the texture, shape, and small ridges on the nose, the system can tell dogs apart, even when they look similar in other ways (like breed or color).

However, there are still some limitations to the proposed approach. One main issue is the quality of the videos. If the video is dark, blurry, or taken from a bad angle, the nose patterns may not be visible clearly, and that can affect the accuracy of the system. Sometimes, the system might miss the nose completely, or produce a false match if the image quality is too low. Also, since we only use the nose for identification, in some situations that might not be enough.

To make the system better in the future, we have several ideas:

- We can also extract features from the dog's face, such as eye position, muzzle shape, or fur texture. Combining both nose and face features into one embedding might give more reliable results.
- We plan to try using Siamese networks or other similarity-based models, which are very good at telling apart images that look very similar.
- Using higher-quality video, or applying image enhancement techniques like reducing noise or sharpening before detection, might improve nose visibility and help the model perform better.
- We may also explore data augmentation techniques to train the model to be more robust to different lighting conditions, angles, and image noise.

Lastly, while the proposed work focuses on identifying dogs, the same idea could be applied to other animals that have unique nose or muzzle patterns. For example, cows, cats, or even wild animals like zebras have been studied using similar biometric techniques. This makes the method flexible and suitable for future research or wider use in animal tracking, health monitoring, or identification systems.

References

[1] K. Lai, X. Tu and S. Yanushkevich, (2019) "Dog Identification using Soft Biometrics and Neural Networks," 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, pp. 1–8, doi: 10.1109/IJCNN.2019.8851971.

[2] T. Moreira, M. Perez, R. Werneck, and E. Valle, (2017) "Where is my puppy? Retrieving lost dogs by facial features," Multimedia Tools and Applications, vol. 76, doi: 10.1007/s11042-016-3824-1.

[3] Y. Qiao, D. Su, H. Kong, S. Sukkarieh, S. Lomax, and C. Clark, (2019) "Individual cattle identification using a deep learning based framework," IFAC-PapersOnLine, vol. 52, no. 30, pp. 318–323, doi: 10.1016/j.ifacol.2019.12.558.

[4] E. Azizi and L. Zaman, (2023) "Deep learning pet identification using face and body," Information, vol. 14, no. 5, p. 278, doi: 10.3390/info14050278.

[5] S. Schneider, G. W. Taylor, and S. C. Kremer, (2020) "Similarity learning networks for animal individual re-identification - beyond the capabilities of a human observer," in 2020 IEEE Winter Applications of Computer Vision Workshops (WACVW), Snowmass, CO, USA, 2020, pp. 44-52, doi: 10.1109/WACVW50321.2020.9096925.

[6] A. Voinea, R. Kock, and M. A. Dhali, (2025) "LostPaw: Finding lost pets using a contrastive learning-based transformer with visual input," in Proc. 14th Int. Conf. Pattern Recognit. Appl. Methods (ICPRAM), pp. 757–763, doi: 10.5220/0013261600003905.

[7] Alibaba Cloud Tianchi, "CVPR2022 Biometrics Workshop – Pet Biometric Challenge," [Online]. Available: https://tianchi.aliyun.com/competition/entrance/531952/information

[8] Ultralytics, "YOLOv8 models," Ultralytics Documentation, 2024. [Online]. Available: https://docs.ultralytics.com/models/.

[9] H. Zhang, et al., "ResNeSt: Split-Attention Networks," GitHub repository, [Online]. Available: https://github.com/zhanghang1989/ResNeSt/releases/download/weights_step1/resnest101-22405ba7.pth

[10] Fei Shen, Zhe Wang, Zijun Wang, Xiaode Fu, Jiayi Chen, Xiaoyu Du, and Jinhui Tang, “A Competitive Method for Dog Nose-print Re-identification,” Nanjing University of Science and Technology, DeepBlue Technology Co., Ltd, Guangdong University of Technology.
[11] B. Li, Z. Wang, N. Wu, S. Shi, and Q. Ma, "Dog nose print matching with dual global descriptor based on contrastive learning," Yunnan University, ShanghaiTech University, and Zhejiang A&F University.
[12] Z. Li, Z. Li, H. Zheng, B. Liu, R. Wang, Y. Shi, J. Chen, and H. Ling, "1st place solution to Pet Biometric Challenge 2022," Huazhong University of Science and Technology.