



Improving Hard Negative Handling for Individual Dog Re-identification with Facial Biometrics

Punyanuch Borwarnginn¹, Worapan Kusakunniran², Qiang Wu³ and Thanongchai Siriapisith⁴

ABSTRACT

Visual biometrics, such as facial recognition, play an important role in authentication and identification. This concept extends to animal biometrics, particularly for identifying individual dogs, which is useful for veterinary care, ownership verification, population control, and disease monitoring. However, distinguishing visually similar dogs within the same breed remains challenging due to hard negative samples. In this work, we propose a deep learning model for individual dog re-identification using contrastive learning with a binary cross-entropy loss function. Instead of treating each dog as a separate class, the proposed framework learns a similarity function from positive and negative image pairs, including a high proportion of hard negative pairs during training, to better capture fine-grained facial differences. Experimental results show that the proposed model achieves 96.83% accuracy on hard negative samples and 81.48% accuracy on the public Flickr dog dataset in a zero-shot setting, demonstrating improved performance in same-breed dog re-identification over traditional multiclass classification approaches.

Article information:

Keywords: Contrastive Learning, Convolutional Neural Network, Deep Learning, Hard Negative, Visual Biometrics

Article history:

Received: January 14, 2026

Revised: March 26, 2026

Accepted: May 21, 2026

Published: June 13, 2026

(Online)

DOI: 10.37936/ecti-cit.2026203.265723

1. INTRODUCTION

1.1 The Critical Role of Anomaly Detection in WSNs

Dogs are the most common domestic animals. According to a recent estimate, the population of dogs in Thailand is 12.8 million, with a significant number being non-owned dogs [1]. It presents several challenges to manage these large numbers, including population control, disease outbreaks, vaccination tracking, and legal ownership verification. To address these issues effectively, tools for individual dog management and identification are required.

While conventional methods like body marker systems (such as dye, paint, and tattoo) have been used for dog identification, they have some limitations. Body markers can easily fade or be altered. Although microchipping is reliable, it is expensive and requires expert implantation.

In recent years, many applications have successfully used deep learning algorithms for human re-identification [2] using visual biometrics such as fingerprints, iris scans, and facial recognition. As a result, researchers have extended these techniques to animal recognition. Similar to humans, animal biometrics can classify species, identify individuals, and monitor behaviors [3]. However, different species have unique modalities for identifying individuals, such as zebra coat patterns [4], elephant's ears [5], muzzle print cattle images [6], and pet animals' facial images [7].

Previous studies on visual dog biometrics have been limited and have primarily focused on dog breed classification using pattern-recognition techniques. For instance, the Stanford dogs dataset [8] was used in the Kaggle dog breed identification challenge. Chanvichitkul *et al.* [9] proposed dog breed classification using face contours as a coarse classification with a Principal Component Analysis (PCA)

^{1,2}The authors are with the Faculty of Information and Communication Technology, Mahidol University, Nakhon Pathom, Thailand, Email: punyanuch.bor@mahidol.ac.th and worapan.kun@mahidol.ac.th

³The author is with the Faculty of Engineering and Information Technology, University of Technology Sydney, NSW, Australia, Email: qiang.wu@uts.edu.au

⁴The author is with the Department of Radiology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand, Email: thanongchai@gmail.com

²Corresponding author: worapan.kun@mahidol.ac.th

classifier as a fine classification. An extended version [10] of their work included local features such as ears and faces to enhance classification within the same group. This method utilized Normalized Cross-Correlation (NCC) for each local part, followed by PCA for breed classification. The results showed a four times improvement in runtime and an accuracy of 88% across 35 dog breeds.

Similarly, various feature extraction techniques have been used to extract features for classifying breeds of domestic animals. For example, one approach involved using a combination of shape and appearance features including Histogram of Oriented Gradient (HOG) [11] and Scale-Invariant Feature Transform (SIFT) [12]. The method achieved an accuracy of 69% in identifying 37 breeds of cats and dogs. Another study by Liu *et al.* [13] proposed a method using a combination of grayscale SIFT descriptors, color histograms, and appearance models of facial parts such as eyes, noses, and ears to train an SVM classifier for detecting dog breeds. The model achieved a 67% recognition rate across 133 dog breeds.

Recently, there has been a growing interest in using deep learning techniques to identify individual dogs in images. Moreira *et al.* [14] compared four human facial recognition approaches (EigenFaces, FisherFaces, LBPH, and Sparse methods) against their own two CNN networks, BARK and WOOF. They applied these methods to their public dataset (Flickr-dog) which contained 42 dogs from 2 breeds: huskies and pugs. The WOOF model achieved the highest accuracy of 66.90%, with the most misclassified dogs from the same breeds.

Similarly, Tu *et al.* [7] proposed dog identification using deep learning techniques. The study employed three datasets: Stanford dogs, Columbia dogs for breed classification, and Flickr-dog for testing dog identification. They proposed a transfer learning approach using GoogleNet architecture to train a model called BreedNet, which identifies breeds from dog faces. BreedNet is then fine-tuned with the Flickr-dog dataset to produce DogNet, which identifies individual dogs. The study achieved 83.94% accuracy. Furthermore, Lai *et al.* [15] developed a revised version of the model which incorporated soft biometrics information such as gender and breeds. The revised model achieved an accuracy of 84.94% when identifying a mix of Pug and Husky breeds. The study also found that when each breed was trained in a separate model, the accuracy was 88.67% for Huskies and 81.40% for Pugs.

Batic [16] presented an initial work on using social media images to identify individual dogs. Their method applied transfer learning on the Inception V3 architecture which was initially trained on the Flickr-dog data (Pet2NetID v1) and later fine-tuned on their Pet2Net Dataset to identify 16 individual dogs from

10 different breeds. The model achieved 94.59% accuracy.

Most studies in dog identification rely on a classification model trained on relatively small datasets. This method represents individual dogs in a class, which requires model retraining when new dogs are added. Thus, reported performance is typically tested on a few dogs, which may not accurately represent real-world scenarios. Although CNN models can identify dogs from different breeds, in reality, distinguishing dogs of the same breed remains a significant challenge due to their similar physical characteristics, resulting in higher rates of false positives.

To address this issue, an image verification and similarity model can be used, like in human face identification. These models leverage feature embeddings, similar to FaceNet [17], and have demonstrated promising results without frequent model retraining. Recent research in contrastive learning offers a promising result in training image representation [18, 19].

Contrastive learning learns feature representations by contrasting pairs of samples using a contrastive loss function. This method has been applied to various image domains such as unpaired image-to-image translation [20] and reducing the impact of hard negatives [21].

Furthermore, Sundaram *et al.* [22] conducted a comprehensive animal detection and classification review, outlining the datasets that are currently available in animal studies, with a focus on species classification such as identifying different breeds of animals. Thus, there is a lack of available datasets that can be used to identify individual animals. Therefore, there is a need for more comprehensive datasets for this research area.

This paper presents a method for dog re-identification by extending the use of contrastive learning. The focus is on improving re-identification within the same breed, which can often produce false positives or “hard negative samples”. The contributions of this study are as follows:

1. We propose an extension of contrastive learning that uses a binary cross-entropy loss function to enhance dog re-identification, particularly for small datasets.
2. The results demonstrate that our model achieves higher accuracy compared with traditional multiclass classification approaches, particularly in recognizing individual dogs from some difficult breeds. These results demonstrate the potential of a visual biometric system in supporting tasks such as collecting dog populations, registering dogs, and identifying dogs using images.
3. To support this research, we introduced the MVITdog dataset, consisting of 300 individual dogs, each with a minimum of 10 images per dog. The images have been manually cropped to fo-

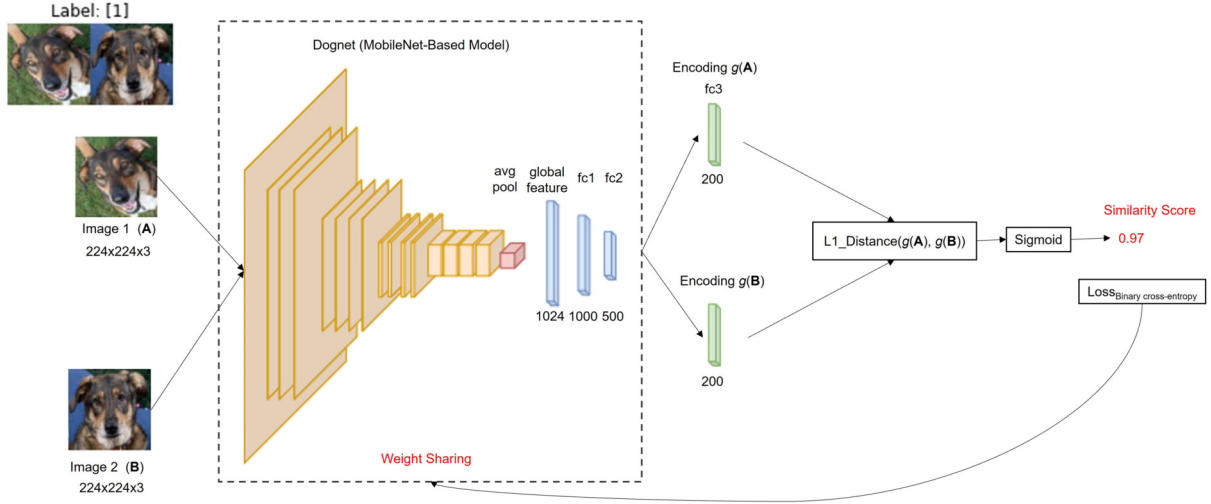


Fig.1: Overview of Dog ID using Contrastive Learning technique with binary cross-entropy loss function.

cus on dog faces, making it a resource for animal visual biometric identification.

2. PROPOSED METHOD

2.1 Dog Identification Model

Traditional approaches to dog identification view each individual dog as a separate class in a multiclass classification framework. While effective for a fixed set of known dogs, such methods require full model retraining whenever new dogs are introduced, making them impractical for large-scale or open-set scenarios. Furthermore, training becomes increasingly resource-intensive and time-consuming as the number of dogs increases. In contrast, our method applies contrastive learning to develop a similarity model that learns a function for directly comparing pairs of dog images. This approach enables the model to identify unseen dogs at inference time without retraining. The model structure is shown in Fig. 1.

The model accepts a pair of images as input and produces a similarity score as output. It comprises two main components: the base model and the contrastive learning module.

2.1.1 Base Model: DogNet

Due to the number of datasets, this approach requires a pre-trained Convolutional Neural Network (CNN) model with similar initial learning weights. For this purpose, we adopted a base model from the architecture in [7] called DogNet. DogNet is a MobileNet-based CNN model trained using a 2-step training process as illustrated in Fig. 2. This 2-step training process aims to focus the feature selection focus on a similar domain: transitioning from dog breed classification (BreedNet) to dog identification (DogNet). The initial phase involves fine-tuning the pre-trained CNN model on a larger dataset, such as ImageNet, for dog breed classification. To initialize

BreedNet, we used a MobileNet model pre-trained on the ImageNet dataset. BreedNet was then fine-tuned on the Columbia Dog Breed Dataset. Its classification head consists of a fully connected layer followed by a softmax activation function, with 133 output units corresponding to the 133 breed classes in the Columbia Dog Breed Dataset. We used the learned weights from BreedNet to initialize DogNet, which was then trained end-to-end on the MVITDog dataset with all network parameters updated. Thus, the second stage involves further fine-tuning of the initialized model. The result of this two-step training process is DogNet, which serves as the base model for our contrastive training approach.

2.1.2 Contrastive Learning Module

In contrastive learning, the model is trained using pairs of inputs, denoted as A and B, associated with a binary label: 0 (dissimilar dogs) or 1 (the same dog). To extract features from these inputs, each input is connected to DogNet model. Then, we apply dropout and a series of fully connected layers (fc1, fc2, and fc3) after DogNet.

The encoding of each input is obtained from the final fully connected layer (fc3), which is $g(A)$ and $g(B)$. This output feature is then used to calculate the L1 distance between the two encoders. The L1 distance is the sum of absolute differences defined as follows:

$$L1 = \sum_{i=1}^n |a_i - b_i| \quad (1)$$

After calculating the L1 distance, the output is connected to the final output layer. At this layer, we apply the sigmoid function to generate a similarity score that falls within the range of 0 to 1. During the training process, the weights of DogNet model are shared for each pair of inputs. These weights are adjusted using a loss function, which is the binary

cross-entropy loss function. Given that our training labels are binary, this loss function is suited for the purpose of weight adjustment. Binary cross-entropy is calculated as follows:

$$Loss = -\frac{1}{m} \sum_{i=1}^m (y_i \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i)) \quad (2)$$

where m is the batch size, y_i is the ground-truth label (either 0 for dissimilar or 1 for similar), and $\hat{y}_i$ is the predicted similarity score for the i -th pair.

2.2 Verification

Once the contrastive learning model is trained, it transforms images into feature representations or embeddings. These embeddings capture the visual characteristics of the images and can be used to evaluate the verification task, similar to FaceNet [17]. One of the main advantages is that we can generate embeddings for new dogs without having to retrain the model with new images.

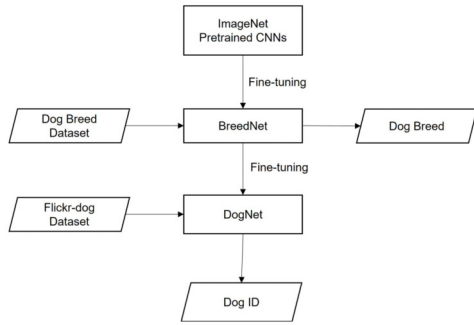


Fig.2: 2-Step training framework using dog breed as the initial model [7, 15].

To verify the similarity of dogs, we use these embeddings to compute feature maps and measure their similarity using the cosine similarity. Cosine similarity is calculated as the cosine of the angle between two vectors. If the angle between these vectors is small, the cosine similarity will be close to 1, indicating high similarity. However, in our task, the closest distance will likely be the same dog. The cosine similarity is defined as follows:

$$similarity(A, B) = \cos(\theta) \quad (3)$$

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^n (A_i B_i)}{\sqrt{\sum_{i=1}^n (A_i^2)} \sqrt{\sum_{i=1}^n (B_i^2)}} \quad (4)$$

2.3 Implementation

The implementation of the dog identification model consists of three main steps: data preparation,

model training, and verification as shown in Fig. 3. Data preparation includes cleansing, data augmentations, and preparing data in set of pairs. We use the data to train a dog identification model as illustrated in Fig. 1. The model training process uses contrastive learning to train the feature representation. The trained model is then used as the embedding for verification by comparing similarities and calculating the results.

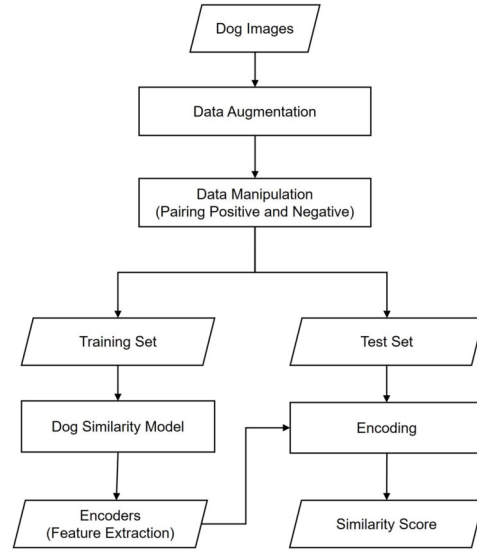


Fig.3: The overview of training and testing process.

In the data preparation step, we extract and crop the dog's face using their face regions as input images. However, training the dog similarity model requires input data in the form of positive and negative pairs. Since the number of dog images is limited, we applied data augmentations. These included image rotation, random noise, and horizontal image flipping to increase the dataset size.

We generate positive pairs by selecting two images from the same dog. In contrast, negative pairs are generated from different dogs, especially the hard negative pairs involving dogs of the same breed are more challenging. We ensure that our pairs consist of hard negative pairs to train the model to distinguish between dogs with visual similarities. Based on the literature, it was reported that a large proportion of negative pairs could help in the model learning process. Therefore, in our dataset, negative pairs are four times more than positive pairs due to the memory resources. Fig. 5 shows sample pairs from the dataset.

2.3.1 Contrastive Learning for Dog Face Image Representation

Our training set consists of 100 individual dogs. Each dog has 10 original images. We applied data augmentation techniques to enhance the training set, doubling the number of images for each dog to 20 and

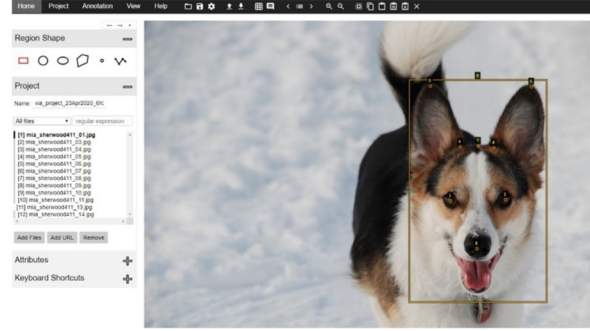


Fig.4: Sample manual labeled using VIA software [24].



Fig.5: Sample negative (0) and positive pairs (1).

2 images per dog for the test set. As in contrastive learning, the training process requires input in the form of pairs. For each dog, 190 positive pairs were generated, resulting in 19,000 positive pairs for the entire training set. In contrast, negative pairs are four times positive pairs. These negative pairs are from two types: hard negative pairs and different breed pairs. Hard negative pairs are a quarter of the total negative pairs, 19,000 pairs.

These pairs are dogs from the same breed, representing a challenging task for the model to learn and capture the dissimilarity. Meanwhile, the remaining 57,000 negative pairs are from dogs belonging to different breeds which could help the model generalize. We selected 2 images per dog for the test set to form a positive pair. This yields 100 positive pairs combined with 400 negative pairs as in the training set.

Our proposed model was implemented and empirically trained using Tensorflow 2.0. During training, we set the batch size to 100, the number of training epochs to 200, and the image size to 224×224 pixels. The model is trained using Nvidia A100 GPU and the training process was over 3 days. The model was compiled with Adam optimizer with the learning

rate starting from (1×10^{-2}) to (1×10^{-6}) . Each training epoch is approximately 15 minutes. However, we encountered a memory resource challenge due to the size of our dataset. Therefore, we optimize the data loading process by changing from TensorFlow's image dataset from directory to TensorFlow's `tf.data.Dataset` with image files.

This change effectively reduces the time required for each training epoch by half. The proposed model achieved 93% validation accuracy at the selected checkpoint, which was used as the embedding model for the dog re-identification task. Fig. 8 shows the model loss and accuracy while training the contrastive learning. The validation loss decreases, aligning with the training loss.

3. EXPERIMENTS

Our experiments consist of two main tasks, i.e., the image classification approach and the image verification approach. The dog re-identification model was trained using a 2-step training framework for image classification and evaluated using predicted results from the softmax layer. For image verification, the dog re-identification model was trained using our proposed contrastive model then the model was used as feature embeddings to calculate distance using cosine similarity on 1) the mixed dog dataset regardless of the dog breeds and 2) the hard negative dataset. The dataset is divided into two parts, the training and validating set for contrastive learning and the test set for dog re-identification. In the image verification experiments, query images used for identification are referred to as probe images, whereas reference images used for matching are referred to as gallery images. The trained contrastive learning model used to generate feature embeddings is referred to as the dog similarity model. In contrast, the full identification framework is referred to as ContrastiveDogID.

3.1 Experiment 1: Dog re-identification using the image classification approach

Previous studies have viewed dog re-identification using an image classification approach. This method treats each dog as an individual class, which means

that the model needs to be retrained every time new dogs are added to the dataset. To address this issue, we conducted an experiment to investigate this limitation.

We adopted a model architecture from the 2-step learning framework shown in Fig. 2. Specifically, BreedNet was trained using the Columbia Dog Dataset, and then using it to fine-tune DogNet on our MVITDog Dataset.

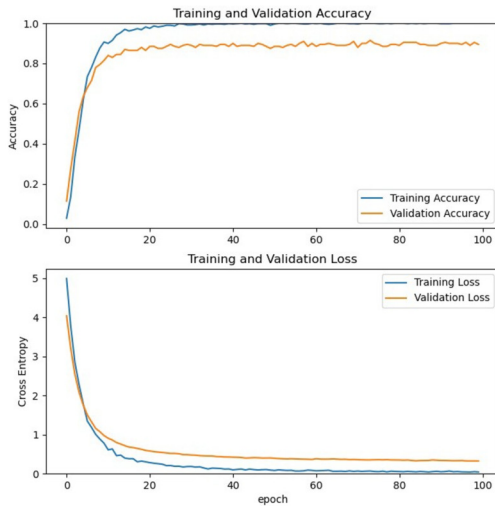


Fig.6: Training and validation accuracy and loss from classification approach on 100 dogs.

To conduct this experiment, we randomly selected 100 dogs representing 20 different breeds, with a minimum of 10 images per dog from MVITDog. Our experiments compared the performance of the model across 4 pre-trained CNN models. Table 1 presents a summary of the results showing that as the number of dogs (classes) in the dataset increased, the accuracy of the model decreased. In addition, the Confusion Matrix in Fig. 7 shows that prediction results were more likely to be confused as the number of classes increased. MobileNet achieved the highest accuracy among the models, with 94.50% accuracy for 100 dogs. Fig. 6 shows the training graph for MobileNet using the pre-trained model on our dataset.

3.2 Experiment 2: Dog re-identification using image verification approach (Contrastive Learning)

In this study, we investigate the effectiveness of our proposed contrastive model as feature encoders for dog re-identification using an image verification approach. We evaluated the model's performance under three different scenarios: identifying unseen dogs regardless of breed, breed-oriented dogs for hard negative samples, and a public dataset.

For the first scenario, we selected 200 unknown dogs from the MVITdog dataset, disregarding their breeds. The test set was split into 3 groups accord-

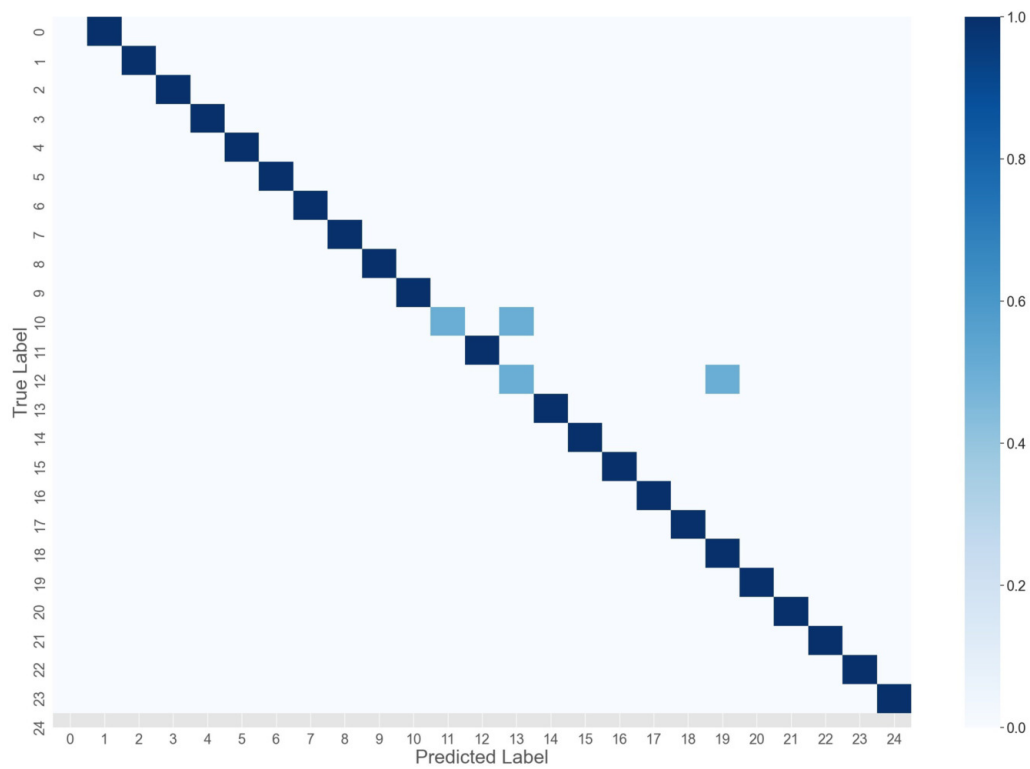
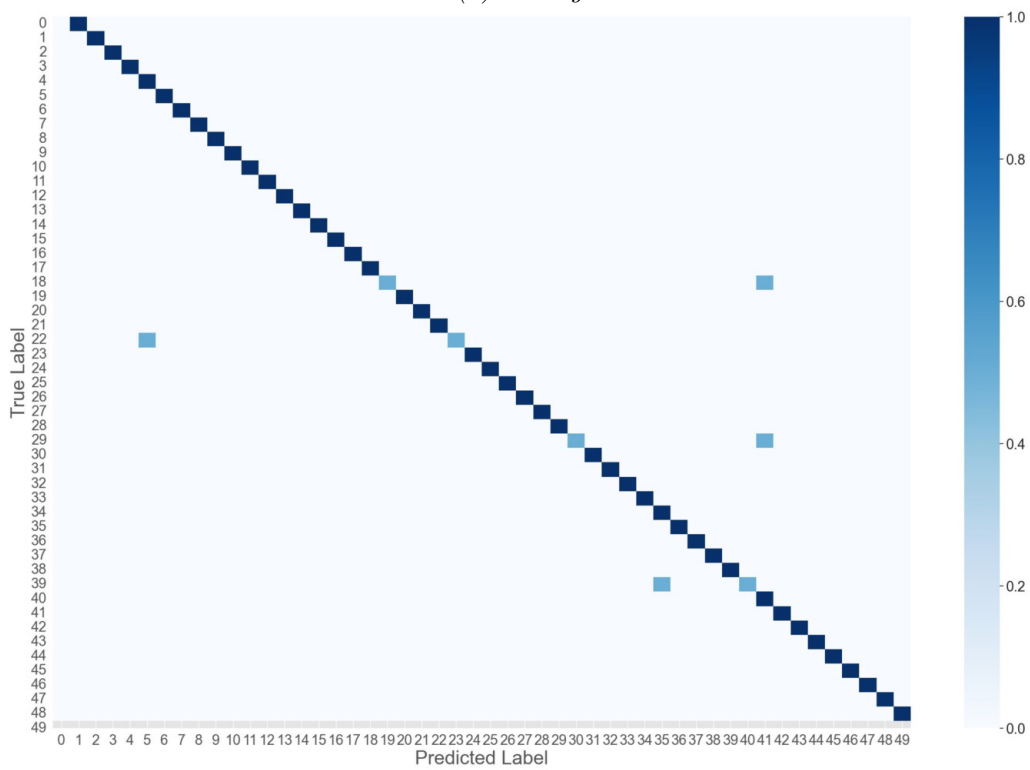
ing to the number of gallery and probe images for each set. Table 2 shows the accuracy achieved by the model in each set.

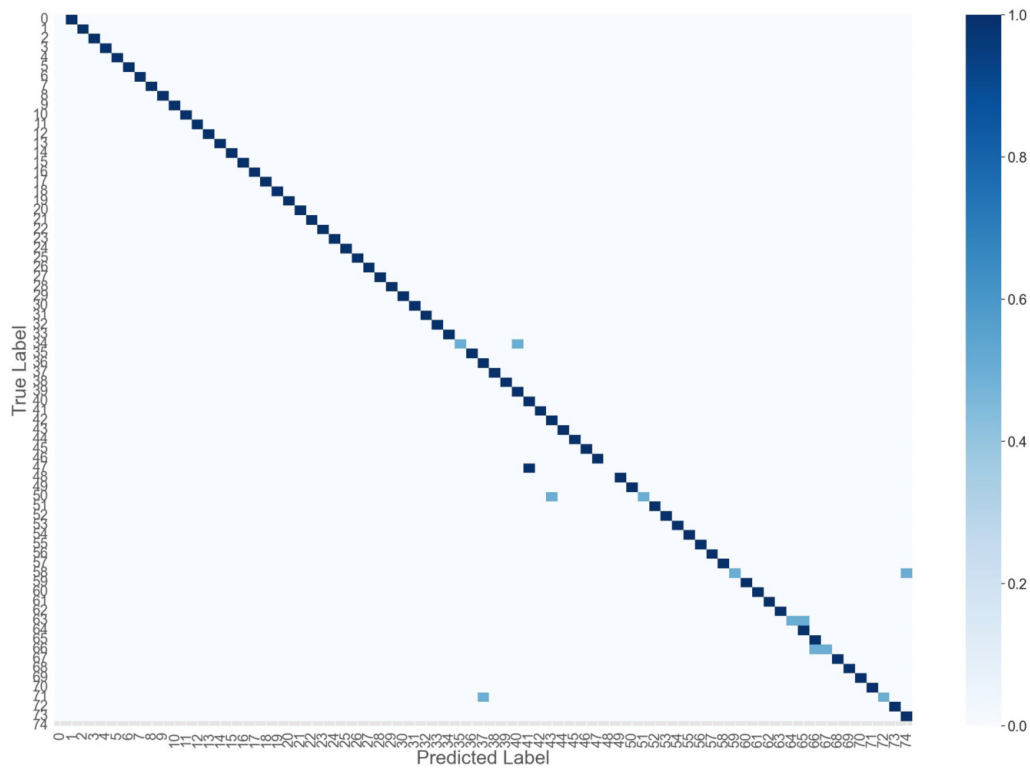
For the second scenario, the test sets were selected from MVITdog under specific conditions to simulate the verification of hard negative samples. We selected breeds containing more than one dog, focusing on pure breeds and excluding mixed breeds. This resulted in a total of 79 dogs from 25 breeds. Within this set, the test set was divided into 2 probe images and 5 gallery images. Our proposed model generated embeddings for each image and then calculated a similarity matrix using cosine similarity for all probe and gallery images. To demonstrate the model's ability, we focus solely on reidentifying dogs within their own breeds. The top-1 accuracy, F1-score, precision, and recall are reported in Table 3. The confusion matrix of each model is shown in Fig. 9a and Fig. 9b.

In the mixed-breed scenario (Table 2), the proposed model achieved slightly lower accuracy than DogNet (86.75% vs. 89.00% with 6 gallery images). In contrast, remaining comparable for dog re-identification in this setting. However, unlike DogNet, our model does not require retraining when new dogs are introduced, which makes the approach more scalable in practice. Recognizing dogs within the same breed remains a more challenging problem. In the hard negative scenario (Table 3), our model achieved higher accuracy and normalized accuracy than DogNet. The results show that our model can detect and differentiate dogs from the same breed, including challenging breeds such as Samoyed and American PitBull Terrier. For example, Samoyed is known to have the same color and furry coat, making it difficult to correctly reidentify individuals, but our model achieved 87.50% accuracy in dog identification within Samoyed. Meanwhile, our model outperformed DogNet's 87.50% accuracy in reidentifying American Pitbull Terrier, achieving a 100% accuracy rate. Table 4 reports the accuracy of the misidentification breeds. In contrast, Fig. 10 shows sample images of breeds where our model achieved higher accuracy.

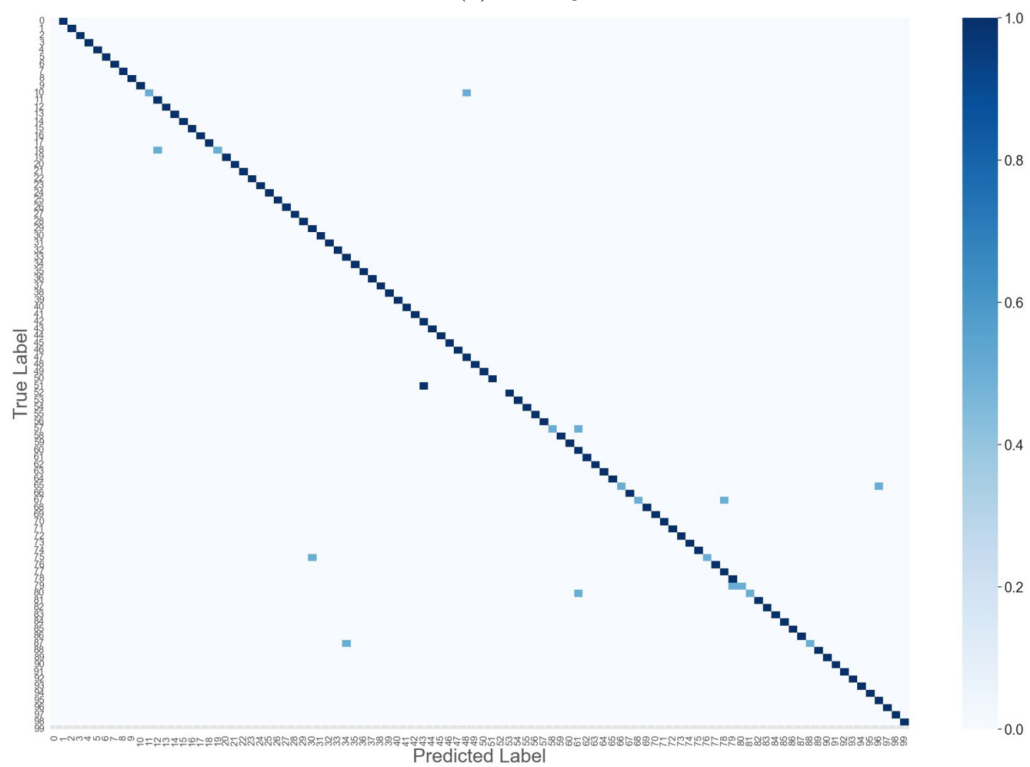
The improved results of ContrastiveDogID over DogNet in the hard negative scenario can likely be attributed to the contrastive learning framework, which explicitly trains the model to distinguish between visually similar dog pairs. By incorporating a high proportion of hard negative pairs, dogs from the same breed, during training, the model is encouraged to learn fine-grained facial features that capture individual differences within a breed, rather than relying primarily on coarse inter-breed distinctions.

Despite this improvement, certain breeds remain challenging. Samoyed and Australian Shepherd showed lower identification accuracy, which is likely related to three factors: (1) high intra-breed visual homogeneity, Samoyed dogs share a distinctively

*(a) 25 Dogs**(b) 50 Dogs*



(c) 75 Dogs

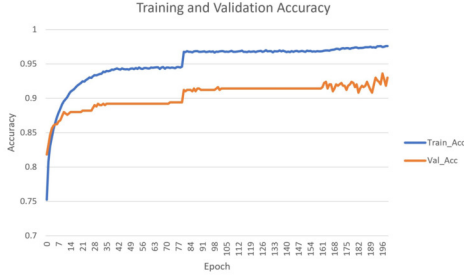


(d) 100 Dogs

Fig.7: Confusion matrices of the classification approach using MobileNet on datasets with different numbers of individual dogs: (a) 25 dogs, (b) 50 dogs, (c) 75 dogs, and (d) 100 dogs. Rows denote true labels and columns denote predicted labels.

Table 1: Accuracy from 4 pretrained CNN models on different numbers of individual dogs.

Pre-trained Models	Number of dogs			
	25	50	75	100
InceptionV3	96.00%	94.00%	91.33%	91.00%
Xception	98.00%	95.00%	94.00%	90.50%
VGG-16	96.00%	94.00%	92.67%	89.00%
MobileNet	96.00%	96.00%	94.67%	94.50%



(a) Training and validation accuracy



(b) Training and validation loss

Fig.8: Training and validation accuracy and Loss from the contrastive learning.

white, fluffy coat with limited facial variation. In contrast, Australian Shepherds often exhibit similar coat patterns across individuals; (2) the limited number of training images per dog (minimum 10 images), which restricts the model’s ability to capture sufficient within-dog variation; and (3) pose and illumination variation across the collected images.

Table 2: The dog re-identification accuracy on the mixed breed dataset.

Methods	(#Gallery, #Probe)		
	(2,2)	(5,2)	(6,2)
DogNet	78.50	87.00	89.00
ContrastiveDogID (Proposed Model)	74.00	84.00	86.75

These failure cases are illustrated in Fig. 9 and Fig. 10. Representative misclassification patterns can also be observed in the confusion matrices shown in Fig. 9, which further highlight the difficulty of distinguishing visually similar dogs within the same breed.

For the last scenario, we evaluate our model on a public dataset, the Flickr dog dataset [14]. The dataset contains 42 individual dogs from 2 breeds: Pug and Siberian Husky. Our proposed model demonstrated that it could generate their embedding without retraining. The model achieved a top-1 accuracy of 81.48%, with an F1-score of 81.48%, a precision rate of 74.07%, and a recall rate of 81.48%. It is worth noting that our model was trained solely on the MVITDog dataset and evaluated on the Flickr dog dataset without any retraining, i.e., in a zero-shot generalization setting. Although a direct comparison with DogNet [7] (83.94%) is not straightforward, since DogNet was explicitly trained on the Flickr dog

dataset, our model still achieved 81.48% without using any Flickr images during training. Furthermore, our model outperformed the method of Moreira *et al.* [14], which reported 66.90% on the same Flickr dog dataset.

4. CONCLUSION

This paper presents an extended contrastive learning method for improving dog re-identification tasks. Specifically, our focus is on enhancing the identification of dogs within the same breeds, which are called hard negative samples. Our approach involves using contrastive learning with a binary cross-entropy loss function to train feature representations on a limited dataset. The trained model is then used to generate feature embeddings for dog re-identification. Based on our experiments, the proposed model achieved 96.83% accuracy on hard negative samples, outperforming the traditional multiclass DogNet (94.93%) in same-breed dog re-identification. Furthermore, the model generalized to an unseen public dataset (the Flickr dog dataset) in a zero-shot setting, achieving 81.48% accuracy without retraining. This result is comparable to DogNet (83.94%), which was trained on that dataset, and higher than the result reported by Moreira *et al.* (66.90%). These results could potentially be applied to a broader range of applications. In summary, our experiments demonstrate the potential of using contrastive learning to handle diverse datasets, distinguish dogs from the same breed, and adapt to new data without retraining the model. These advantages could be beneficial for further applications in the field.

Table 3: The dog re-identification accuracy on hard negative samples.

Methods	Precision	Recall	F1-Score	Accuracy	Normalized Accuracy
DogNet	96.20	94.94	94.93	94.93	96.95
ContrastiveDogID (Proposed Model)	96.62	96.84	96.83	96.83	98.13

Table 4: The top-1 accuracy for difficult breeds.

Breeds	(Accuracy (%))	
	DogNet	ContrastiveDogID (Proposed Model)
Samoyed	75.00	87.50
American Pitbull Terrier	87.50	100.00
Australian Shepherd	86.36	90.90

ACKNOWLEDGEMENT

This research was supported by the Royal Golden Jubilee (RGJ) Ph.D. Programme under the Thailand Research Fund (No. PHD/0053/2561).

AUTHOR CONTRIBUTIONS

Conceptualization, P.B., W.K. and Q.W.; methodology, P.B., W.K. and Q.W.; software, P.B.; validation, P.B., W.K., Q.W. and T.S.; formal analysis, P.B.; investigation, P.B.; data curation, P.B.; writing, original draft preparation, P.B.; writing, review and editing, P.B., W.K., Q.W. and T.S.; visualization, P.B.; supervision, W.K., Q.W. and T.S. All authors have read and agreed to the published version of the manuscript.

References

- [1] W. Thanapongtharm *et al.*, "Spatial distribution and population estimation of dogs in Thailand: Implications for rabies prevention and control," *Front. Vet. Sci.*, vol. 8, p. 790701, Dec. 2021.
- [2] I. Yaqoob, M. U. Hassan, D. Niu, X. Zhao, I. A. Hameed and S.-U. Hassan, "A novel person re-identification network to address low-resolution problem in smart city context," *ICT Express*, vol. 9, no. 5, pp. 809–814, Oct. 2023.
- [3] H. S. Kühl and T. Burghardt, "Animal biometrics: Quantifying and detecting phenotypic appearance," *Trends in Ecology & Evolution*, vol. 28, no. 7, pp. 432–441, Jul. 2013.
- [4] M. Lahiri, C. Tantipathananandh, R. Warungu, D. I. Rubenstein and T. Y. Berger-Wolf, "Biometric animal databases from field photographs: Identification of individual zebra in the wild," in *Proceedings of the 1st ACM International Conference on Multimedia Retrieval (ICMR '11)*, no.6, pp. 1–8, 2011.
- [5] A. Ardovini, L. Cinque and E. Sangineto, "Identifying elephant photos by multi-curve matching," *Pattern Recognition*, vol. 41, no. 6, pp. 1867–1877, Jun. 2008.
- [6] S. Kumar and S. K. Singh, "Automatic identification of cattle using muzzle point pattern: A hybrid feature extraction and classification paradigm," *Multimedia Tools and Applications*, vol. 76, pp. 26551–26580, Dec. 2016.
- [7] X. Tu, K. Lai and S. Yanushkevich, "Transfer Learning on Convolutional Neural Networks for Dog Identification," *2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)*, Beijing, China, pp. 357–360, 2018.
- [8] A. Khosla, N. Jayadevaprakash, B. Yao and F.-F. Li, "Novel dataset for fine-grained image categorization: Stanford Dogs," in *Proceedings CVPR Workshop Fine-Grained Visual Categorization (FGVC)*, 2011.
- [9] M. Chanvichitkul, P. Kumhom and K. Chamnongthai, "Face recognition based dog breed classification using coarse-to-fine concept and PCA," *2007 Asia-Pacific Conference on Communications*, Bangkok, Thailand, pp. 25–29, 2007.
- [10] P. Prasong and K. Chamnongthai, "Face-recognition-based dog-breed classification using size and position of each local part, and PCA," *2012 9th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, Phetchaburi, Thailand, pp. 1–5, 2012.
- [11] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, CA, USA, vol. 1, pp. 886–893, 2005.
- [12] O. M. Parkhi, A. Vedaldi, A. Zisserman and C. V. Jawahar, "Cats and dogs," *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, USA, pp. 3498–3505, 2012.
- [13] J. Liu, A. Kanazawa, D. Jacobs and P. Belhumeur, "Dog breed classification using part localization," in *Computer Vision – ECCV 2012*, Florence, Italy, pp. 172–185, Oct. 2012.
- [14] T. P. Moreira, M. L. Perez, R. d. O. Werneck and E. Valle, "Where is my puppy? Retrieving lost dogs by facial features," *Multimedia Tools and Applications*, vol. 76, pp. 15325–15340, 2017.
- [15] K. Lai, X. Tu and S. Yanushkevich, "Dog Identification"

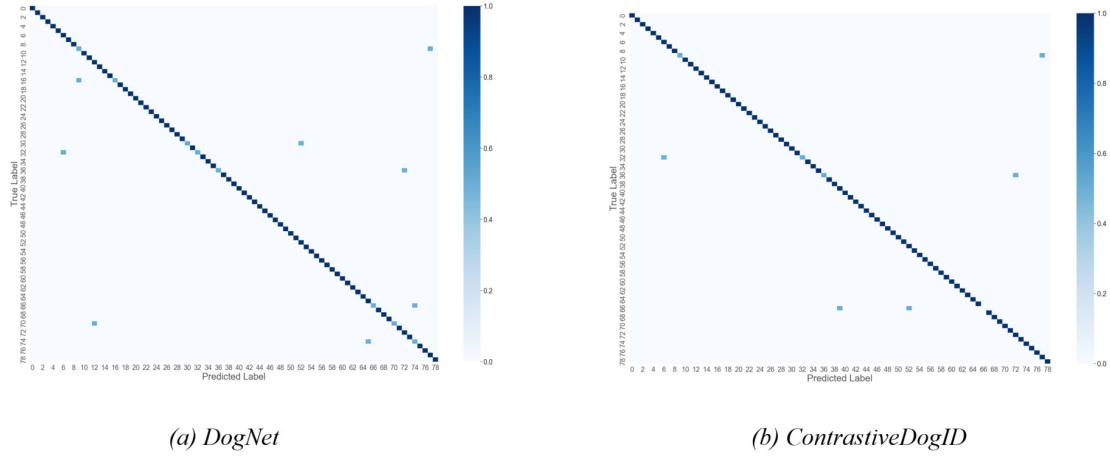


Fig.9: Confusion matrices on the hard negative dataset for (a) DogNet and (b) ContrastiveDogID. Rows denote true labels and columns denote predicted labels.

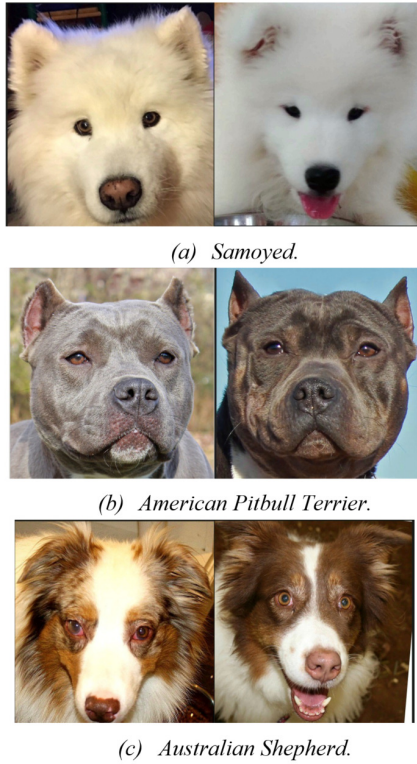


Fig.10: Sample images of difficult breeds for which the model achieved high accuracy.

tification using Soft Biometrics and Neural Networks,” *2019 International Joint Conference on Neural Networks (IJCNN)*, Budapest, Hungary, pp. 1-8, 2019.

- [16] D. Batic and D. Culibrk, “Identifying individual dogs in social media images,” arXiv:2003.06705, 2020. [Online]. Available: <https://arxiv.org/abs/2003.06705>.
- [17] F. Schroff, D. Kalenichenko and J. Philbin, “FaceNet: A unified embedding for face recog-

nition and clustering,” *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, pp. 815-823, 2015

- [18] P. Khosla *et al.*, “Supervised contrastive learning,” *Proceedings of the 34th International Conference on Neural Information Processing Systems*, no.1567, pp. 18661–18673, 2020.
- [19] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “A simple framework for contrastive learning of visual representations,” in *Proceedings of the 37th International Conference on Machine Learning*, no. 149, pp. 1597–1607, 2020.
- [20] W. Wang, W. Zhou, J. Bao, D. Chen and H. Li, “Instance-wise Hard Negative Example Generation for Contrastive Learning in Unpaired Image-to-Image Translation,” *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada, pp. 14000-14009, 2021.
- [21] W. Cui, L. Bai, X. Yang and J. Liang, “A new contrastive learning framework for reducing the effect of hard negatives,” *Knowledge-Based Systems*, vol. 260, no. 110121, 2023.
- [22] N. Sundaram and S. D. Meena, “Integrated animal monitoring system with animal detection and classification capabilities: A review on image modality, techniques, applications, and challenges,” *Artificial Intelligence Review*, vol. 56, pp. 1–51, 2023.
- [23] A. G. Howard *et al.*, “MobileNets: Efficient convolutional neural networks for mobile vision applications,” arXiv:1704.04861, 2017. [Online]. Available: <http://arxiv.org/abs/1704.04861>.
- [24] A. Dutta and A. Zisserman, “The VIA annotation software for images, audio and video,” in *Proceedings of the 27th ACM International Conference on Multimedia*, pp. 2276–2279, 2019.



Punyanuch Borwarnginn received a B.Sc. degree in information and communication technology from Mahidol University, Nakhon Pathom, Thailand, in 2009, and an M.Sc. degree in Informatics from the University of Edinburgh, Edinburgh, United Kingdom, in 2011, and a Ph.D. degree in Computer Science from Mahidol University in 2026.



Qiang Wu received the B.Eng. and M.Eng. degrees from the Harbin Institute of Technology, Harbin, China, in 1996 and 1998, respectively, and the Ph.D. degree from the University of Technology Sydney, Australia, in 2004. He is currently a Professor and a Core Member of the Global Big Data Technologies Centre, University of Technology Sydney. His research interests include computer vision, image processing, pattern recognition, machine learning, and multimedia processing. His research outcomes are applied span over fields such as video security surveillance, biometrics, video data analysis, and human computer interaction. His research outcomes have been published in many premier international conferences, including ECCV, CVPR, ICIP, and ICPR, and the major international journals, such as IEEE TIP, IEEE TSMC-B, IEEE TCSVT, IEEE TIFS, PR, PRL, and Signal Processing.



Worapan Kusakunniran received the B.Eng. degree in computer engineering from the University of New South Wales (UNSW), Australia in 2008, and the Ph.D. degree in computer science and engineering from UNSW, in cooperation with the Neville Roach Laboratory, National ICT Australia, Australia in 2013. He is currently an Associate Professor with the Faculty of Information and Communication Technol-

ogy, Mahidol University, Thailand. He is the author of several papers in top international conferences and journals. He served as a program committee member for many international conferences and workshops. Also, he has served as a reviewer for several international conferences and journals, such as ICPR, ICIP, PR, TIP, and TIFS. He was a recipient of the ICPR Best Biometric Student Paper Award in 2010, and also a winner of several national and international innovation contests. His research interests include biometrics, pattern recognition, medical image processing, computer vision, multimedia, and machine learning.



Thanongchai Siriapisith is a professor of radiology at the Faculty of Medicine Siriraj Hospital Mahidol University. He received his M.D. and Ph.D. degrees from Mahidol University in 1997 and 2018, respectively. He is the author of several papers in international journals. His current research interests include cardiovascular imaging, medical image processing, and machine learning.