Pose-Guided Complementary Features Learning for Amur Tiger Re-Identification

Ning Liu[1], Qijun Zhao[1, 2, 3], Nan Zhang[2], Xinhua Cheng[2] and Jianing Zhu[2]

[1]National Key Laboratory of Fundamental Science on Synthetic Vision, Sichuan University
[2]College of Computer Science, Sichuan University
[3]School of Information Science & Technology, Tibet University

Abstract

Re-identifying different animal individuals is of significant importance to animal behavior and ecology research and protecting endangered species. This paper focuses on Amur tiger re-identification (re-ID) using computer vision (CV) technology. State-of-the-art CV-based Amur tiger re-ID methods extract local features from different body parts of tigers based on stand-alone pose estimation methods. Consequently, they are limited by the pose estimation accuracy and suffer from self-occluded body parts. Instead of estimating elaborated body poses, this paper simplifies tiger poses as right-headed or left-headed and utilizes this information as an auxiliary pose classification task to supervise the feature learning. To further enhance the feature discriminativeness, this paper learns multiple complementary features by steering different feature extraction network branches towards different regions of the tiger body via erasing activated regions from input tiger images. By fusing the pose-guided complementary features, this paper effectively improves the Amur tiger re-ID accuracy as demonstrated in the evaluation experiments on two test datasets. The code and data of this paper are publicly available at https://github.com/liuning-scu-cn/AmurTigerReID.

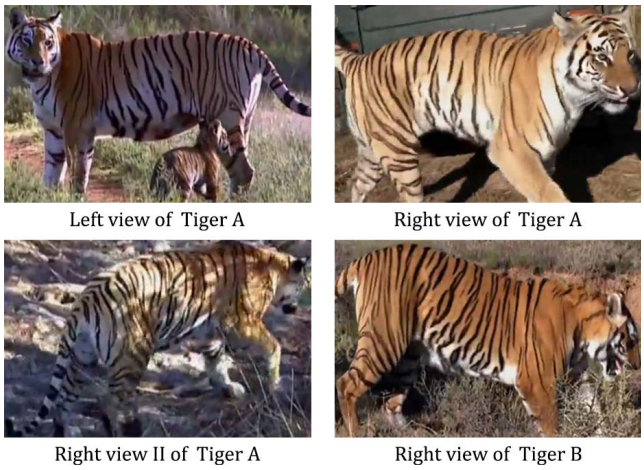


Figure 1. Example tiger images. Right and left sides of one tiger could have quite different stripe patterns (e.g., left view and right view of Tiger A), and are defined as different tiger entities [15]. Two different tigers may have similar appearance on different sides of their bodies (e.g., left view of Tiger A and right view of Tiger B). The appearance of the same side of one tiger could look different in different images due to varying pose and illumination (e.g., the two images of the right view of Tiger A).

1. Introduction

Distinguishing different animal individuals is an essential prerequisite for animal behavior and ecology research and for protecting endangered species. Without the knowledge of which individual an animal is in its population, it is impossible to understand the animal and its population's habitual behaviors or status, and thus difficult to effectively protect them. For a long time, people have to spend lots of efforts to determine the identity of individual animals and to acquire precise up-to-date information of the animal population by visual inspection, collecting excreta and biology samples of the animals, or attaching emitters or microchips to the animals. Therefore, it is highly demanded to develop automatic methods for identifying individual animals [18]. In this paper, we aim to propose an effective method for automatically recognizing individual Amur tigers in the wild.

In recent years, thanks to the widespread surveillance cameras installed in the habitats of wildlife, some researchers have made attempts to recognize individual animals by using computer vision technology for a number of species, such as African penguins [1], northeast tigers [23], lemurs [3], great white sharks [12], primates [4], ringed

seals [2], giant pandas [14] and red pandas [7]. They extract discriminative features from certain body parts of the animals, and differentiate different individuals based on the extracted features. Similar techniques have also been applied to cattle [21], diary cows [16], and pigs [5] in agriculture applications. Despite the impressive accuracy achieved in these studies, they mostly assume that the animals are imaged under well-controlled conditions. Such assumption is however more than often unrealistic for recognizing animals in the wild because of the large variations in the body pose of the animals and the illumination on the images.

As for the task of recognizing individual tigers, Peng [23] and Hiby et al. [10] demonstrated the effectiveness of the stripe patterns on the body of tigers for differentiating individual tigers. However, they did not consider distracting factors in practical tiger images, like varying poses and illuminations, occlusions and cluttered background (see Fig. 1). Li et al. [15] for the first time studied the full pipeline of automatically recognizing individual tigers in the wild. They approached the in-the-wild individual tiger recognition problem as a re-identification (re-ID) problem, and constructed the largest public dataset (called ATRW) of annotated in-the-wild Amur tiger images with specific evaluation protocol for various tasks. Motivated by the re-ID methods for pedestrians and vehicles, they provided several baseline re-ID methods for Amur tigers by using different losses or different features. Particularly, they proposed a pose part based model (PPbM) to deal with the large pose variation of tiger body caused by non-rigid motion. PPbM divides tiger body into different parts according to a standalone pose estimation method, and combines the features extracted from these parts to determine the identity of individual tigers. Although PPbM achieves state-of-the-art accuracy, it could be easily affected by the pose estimation errors and self-occluded body parts.

In this paper, we simplify tiger poses into right-headed or left-headed (see Fig. 1, Left view of Tiger A and Right view of Tiger A), and integrate the tiger pose classification task into the re-ID feature learning procedure as an auxiliary task. Besides, inspired by recent advances in object localization and person re-ID [24, 22] which learn diverse features by forcing the feature extraction network to pay attention to various image regions, we propose a multi-branch feature extraction network in which one branch learns a feature representation from the images that have been partially erased according to the activation maps of previous branches. This way, different branches are enabled to learn complementary features. Our experiments on two test datasets show that improved Amur tiger re-ID accuracy can be obtained based on the combination of the learned complementary features. To summarize, this paper makes the following contributions.

- A novel end-to-end method is proposed to learn diverse features for re-identifying in-the-wild Amur tigers based on multi-task learning and complementary feature learning. By using an auxiliary task of simplified pose classification to guide the feature learning, the proposed method is not dependent on third-party pose estimation modules and can more effectively cope with large pose variations.
- A set of in-the-wild images of 27 tiger entities are collected as a supplementary to ATRW that can be used as a validation set.
- Evaluation results show that the proposed method puts forward the state-of-the-art Amur tiger re-ID accuracy by large margins.

2. Problem Statement

Following the formulation in [15], we treat the left-view and right-view of the same tiger as two different classes (or entities), though they belong to the same individual tiger. As can be seen from Fig. 1, this is reasonable, because one tiger could have substantially different stripe patterns on the two sides of its body. The Amur tiger re-ID problem can be thus stated as follows. Given a probe image of an Amur tiger (either left-headed or right-headed, but unknown in advance), the goal of re-ID is to retrieve from a gallery of Amur tiger images the images that belong to the same view of the same tiger, according to which the identity of the tiger in the probe image can be determined. Note that the pose of the tiger images in the gallery is unknown too.

Because the pose of the tiger images is unknown in advance, it is possible to match the left-view of one tiger to the right-view of another tiger (or vice versa), especially considering that one view of one tiger could have very similar appearance to the other view of another tiger (see Fig. 1, Left view of Tiger A and Right view of Tiger B). To avoid such mistake, it is necessary to take into consideration the pose of tiger images when learning discriminative re-ID features. In this paper, instead of using elaborated pose keypoints as in [15], we simplify the tiger body pose into two categories according to the heading direction of the tigers, i.e., left-headed (or left view) and right-headed (or right view). This reduces the cost of annotating training data and enables our proposed method to more flexibly deal with large pose variations and self-occlusions.

3. Proposed Method

3.1. Overview

Fig. 2 shows the flowchart of the proposed pose-guided complementary features learning (PGCFL) method for Amur tiger re-identification. PGCFL mainly consists of four modules, base module, image-erasure module, complementary module, and fusion module. The base module

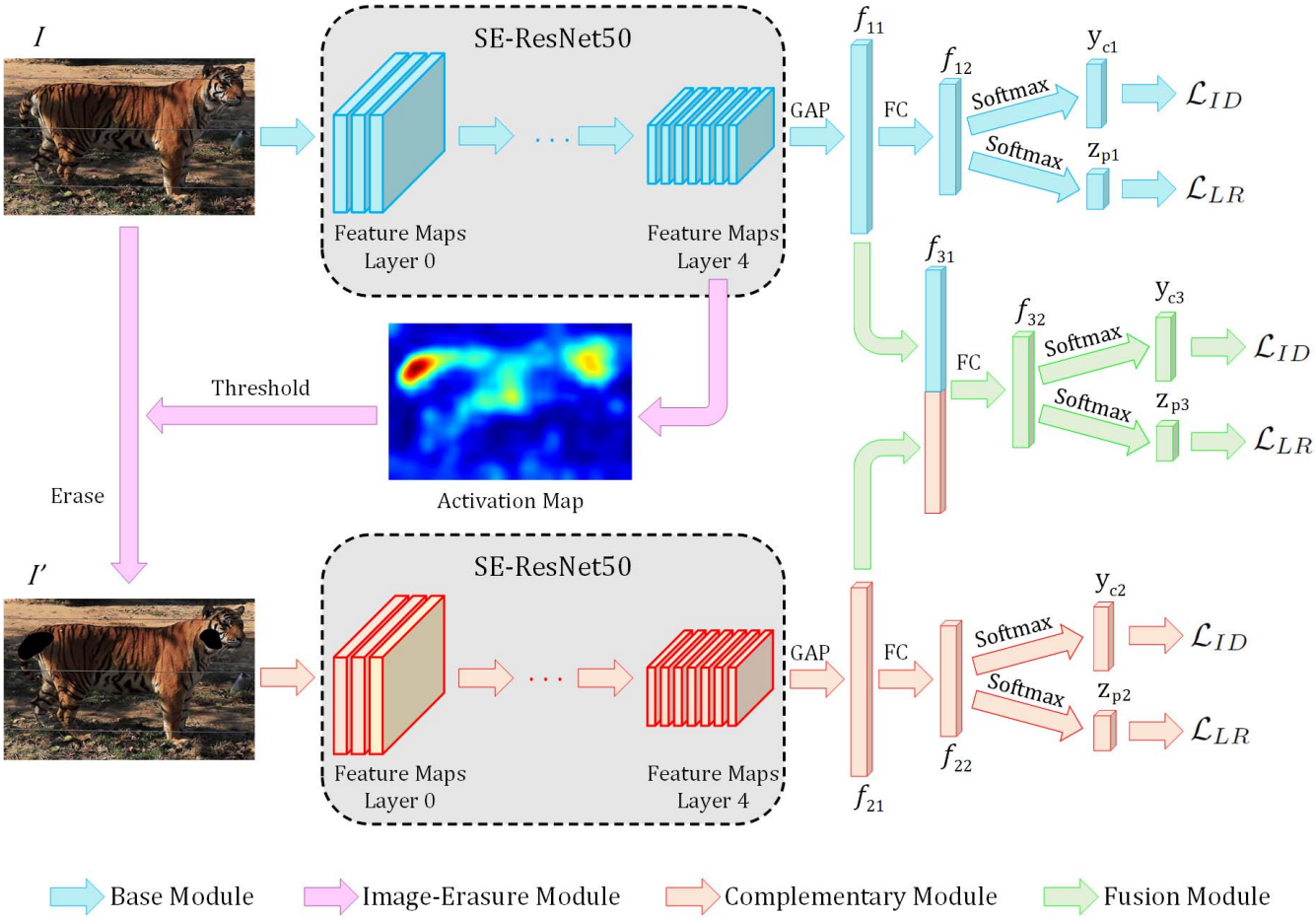


Figure 2. Overview of the proposed pose-guided complementary features learning (PGCFL) method for Amur tiger re-identification.

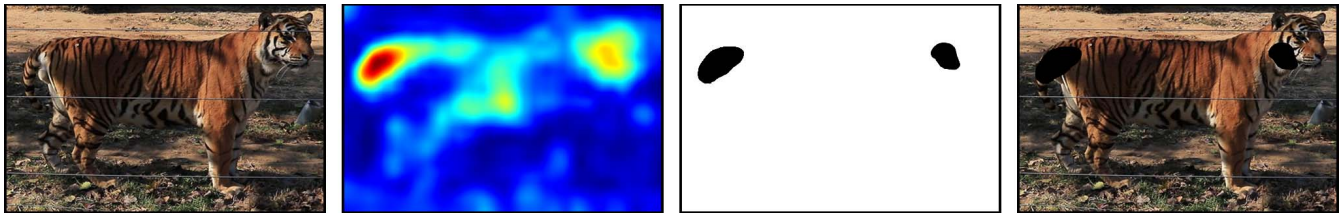
takes the original tiger image as input and learns base re-ID features (f_{11}). Based on the feature maps of the backbone network of the base module, the image-erasure module constructs the activation map that reveals which parts of the tiger image the base module focuses on, and then erases these activated parts from the tiger image to generate a masked tiger image, which is used as the input to the complementary module during training. The complementary module thus learns new discriminative features (f_{21}) from the remaining parts of the tiger image. These features are finally unified into a single feature representation by the fusion module. Note that multiple image-erasure and complementary modules can be stacked when implementing the PGCFL method, and the image-erasure module is used only during training (in other words, both base and complementary modules take the original tiger images as input when using the trained PGCFL model to extract features from tiger images for re-ID). In the rest of this section, we introduce in detail the modules as well as the training of PGCFL and the application of PGCFL for Amur tiger identification.

3.2. Base Module

The base module employs the SE-ResNet50 [6, 11] as the backbone with the residual block adapted according to [8]. More specifically, the convolution layers from Layer0 to Layer4 of SE-ResNet50 are used in PGCFL. The feature maps of Layer4 are fed into a Global Average Pooling (GAP) layer [17] to generate the base feature vector $f_{11} \in \mathbb{R}^{2048\times1}$, which is then transformed to another feature vector $f_{12} \in \mathbb{R}^{512\times1}$ via a fully connected (FC) layer followed by a batch normalization (BN) layer [13] and a dropout layer [20]. Based on f_{12}, the identity (y_c) and pose (z_p) of the input tiger image (I) are respectively predicted via softmax layers.

3.3. Image-Erasure and Complementary Modules

Given an input tiger image I, the backbone network of the base module generates feature maps through Layer0 to Layer4. Let $F \in \mathbb{R}^{K\times H\times W}$ denote the feature maps on Layer4, where $H \times W$ is the spatial size, K is the number of channels (or feature maps). We accumulate these feature



Input Image | Activation Map | Binary Mask | Erased Image

Figure 3. Illustration of image erasure. From left to right: input tiger image, activation map of Layer4 in the base module, binary mask (threshold= 0.6), input erased tiger image for the complementary module during training.

maps via pixel-wise summation crossing channels,

$$M = \sum_{k=0}^{K-1} F^k, \tag{1}$$

where F^k is the k^{th} feature map. As shown in Fig. 3, the accumulated feature map M reveals which parts of the input image the feature extraction layer focuses on. Therefore, M is also called activation map. In this paper, the image-erasure module up-samples the activation map of Layer4 to the same spatial size as the input tiger image, and forms a mask by binarizing it with a threshold α to locate the image regions that have been utilized by the base module. The image-erasure module then sets these regions on the input image to black, resulting in a new tiger image I' in which the already-used regions are excluded from the subsequent feature learning. See Fig. 3.

The complementary module takes the masked tiger image I' as input, and is enforced to discover additional discriminative features from the remaining regions in the original tiger image. It employs the same network structure as the base module, and learns another feature $f_{21} \in \mathbb{R}^{2048\times 1}$. Fig. 4 visualizes the activation maps of Layer4 in the base and complementary modules for some input tiger images. Obviously, the base and complementary modules do focus on different regions of tiger images and thus extract complementary features for re-identifying the tigers.

3.4. Fusion Module

The objective of fusion module is to combine the base and complementary features to form a unified diverse feature representation. This is done by first concatenating the base feature f_{11} and the complementary features f_{21} to get a single feature vector f_{31}, and then transforming it to the unified feature $f_{32} \in \mathbb{R}^{512\times 1}$ via a FC layer followed by a BN layer and a dropout layer. Based on this unified feature, another prediction of the identity and pose of the input tiger image is obtained via softmax layers.

3.5. Training

The proposed PGCFL model is trained in three phases. In Phase I, the backbone network (i.e., SE-ResNet50) of all branches in PGCFL is pre-trained on ILSVRC [19]. In Phase II, the entire PGCFL model is trained in an end-to-end manner under the supervision of two tasks, identification ($\mathcal{L}_{ID}$) and pose classification ($\mathcal{L}_{LR}$). Both tasks are evaluated with cross-entropy loss. In Phase III, we fine-tune the PGCFL model with TriHard loss [9]. Specifically, TriHard loss is applied over the unified features f_{32} (note that the features are normalized such that they have unit L_2 norm).

3.6. Application to Tiger Re-ID

Once the PGCFL model is trained, we can apply it to tiger images to extract the unified diverse feature (i.e., f_{32}) for each image. Given a tiger image, we first feed it into the trained PGCFL model and obtain a feature vector f_{32}^1. Then, we horizontally flip the tiger image, and get another feature vector f_{32}^2 by feeding the flipped image into the trained PGCFL model. We finally concatenate the two feature vectors to get the final unified diverse feature of the tiger image. It is worth mentioning that when using the trained PGCFL model to extract feature from an image, both the base and the complementary modules take the original image as input. This is different from the training phase during which the complementary module uses the partially-erased image rather than the original image as input. By using original images as input for both base and complementary modules during feature extraction can not only reduce the computational complexity, but also avoid potential negative impact of the erasure operation.

After the features are extracted for all the tiger images in the gallery and for the probe tiger image, we compute the cosine similarity between the features of probe image and each gallery image, and generate as the re-ID result a rank list of all the tiger entities in the gallery according to their similarities with the probe image.

4. Experiments

4.1. Experiment Setup

Datasets. To evaluate the effectiveness of the proposed method, we use the training data in ATRW [15] to train the

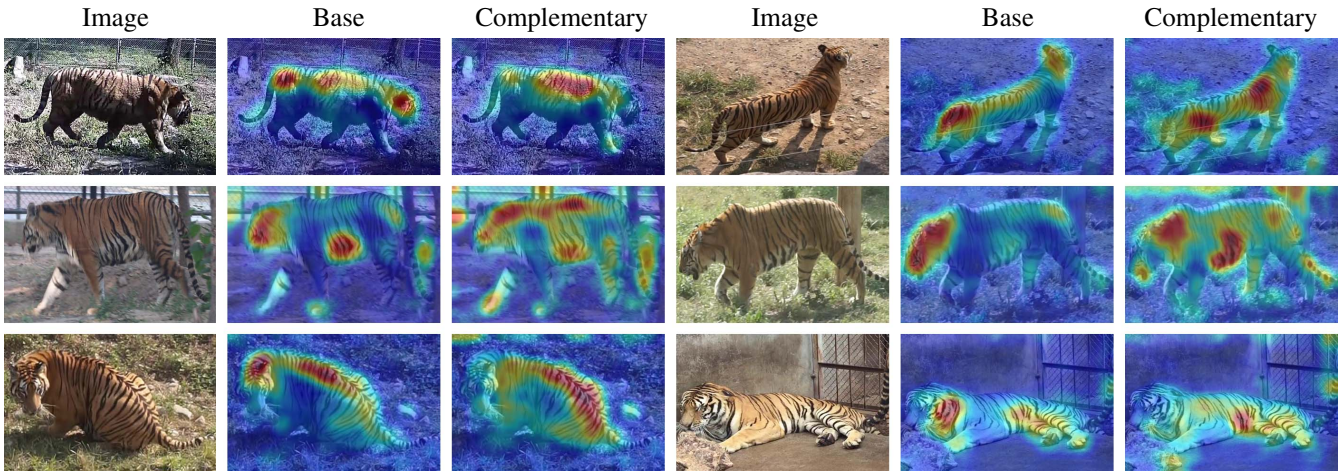


Figure 4. Visualization of the activation maps in the base and complementary modules for some input tiger images. It can be clearly observed that the two modules learn complementary features in different regions on the tiger images.

PGCFL model, and the test data in ATRW as well as a set of tiger images collected by ourselves to assess the performance of the trained PGCFL model. The ATRW training set contains $1,887$ images of 107 Amur tiger entities from 75 different tigers. After removing the low quality tiger images due to blurring or heavy occlusion, we use the remaining $1,824$ images to train the proposed PGCFL model.

The ATRW test set provides $1,762$ images of 75 entities from another 58 individual tigers. These test images are divided into single-camera and cross-camera cases according to whether the query and gallery images are captured by the same camera or by different cameras. The single-camera case has 701 images of 47 tiger entities, and the cross-camera case has $1,061$ images of 28 tiger entities. For more detail please refer to [15].

Besides, we collected an additional test set of Amur tiger images. These tiger images were cropped from in-the-wild videos of Amur tigers that are publicly available in the Internet. When cropping the images, we tried to remove the high correlation between different tiger images by using as large time interval as possible to sample video frames. Following the protocol of ATRW, we finally acquired 276 images of 27 tiger entities as another test set, which we call SCU-Tiger.

Training and Testing Settings. To train the PGCFL model, we resize all the images to 324×504, and augment the training data by randomly cropping images of 288×448 (which is the input spatial size required by the PGCFL model) from the original images, randomly erasing parts of the images, and randomly horizontally flipping the images. Note that for a horizontally flipped image, its pose label should be reversed, but its identity label remains unchanged.

We implement PGCFL in Pytorch, and train it on a PC with two NVIDIA GTX-1080Ti GPUs. In the Phase II training, for the first three epochs, we fix the backbone network and warm-up the rest layers with learning rate from 0.0001 to 0.01. We use SGD with momentum of 0.9, basic learning rate $lr = 0.01$, weight decay $\lambda = 0.0001$, label smoothing of 0.1 and batch size of 10. The learning rate of backbone is set to $0.1 * lr$, and the training stops after 30 epochs with the learning rate decreased by a factor of 10 at epoch 20. In the Phase III training, when fine-tuning the PGCFL model with TriHard loss, we do not carry out warm-up training, and use the same parameter setting as in Phase II except setting the basic learning rate as 0.001 and the margin of TriHard loss as 0.4.

When using the trained PGCFL model to extract features from tiger images during testing, we resize the tiger images to 324×504, from the center of which we crop images of 288×448 as input for PGCFL.

Evaluation Protocol. For evaluation on the ATRW test data, we follow the official protocol in which each image in the test set is taken as probe image, while all the remained images are used as gallery. The re-ID result for each probe image is a rank-list of the gallery images. For evaluation on the SCU-Tiger test data, we randomly choose 30% of each tiger entity as gallery and use the rest images as probe. To further increase the re-ID difficulty, we enlarge the gallery with all the training tiger images from ATRW. We report the top-1 and top-5 identification rates in the Cumulative Match Characteristic (CMC) curves as well as the mean average precision (mAP) results.

4.2. Comparison with State of the Art

In Table 1, we compare our method with the baseline (CE, Triplet Loss, and Aligned-reID) methods and the state-of-the-art PPbM methods in [15] in terms of mAP, top-1 and top-5 identification rates on the ATRW test data. As can

be seen, our method achieves the best results in all cases. Specifically, our method substantially improves the state-of-the-art mAP and top-1 identification rate by 15.7% and 7.2% for the single-camera case, and by 12.6% and 13.7% for the cross-camera case.

Table 1. Comparison with state-of-the-art re-id methods on the ATRW test data.

Method	Single-camera (%)			Cross-camera (%)		
	mAP	top-1	top-5	mAP	top-1	top-5
CE	59.1	78.6	92.7	38.1	69.7	87.8
Triplet Loss	71.3	86.6	96.0	47.2	77.6	90.6
Aligned-reID	64.8	81.2	92.4	44.2	73.8	90.5
PPbM-a	74.1	88.2	96.4	51.7	76.8	91.0
PPbM-b	72.8	89.4	95.6	47.8	77.1	90.7
PGCFL (ours)	**89.8**	**96.6**	**97.7**	**64.3**	**91.3**	**95.8**

We believe that the improvement achieved by our method owes to the following factors. (i) Our method learns complementary features from adaptively-determined different regions in tiger images, whereas the counterpart methods in [15] manually select local parts on tiger body based on the pose keypoints located by a standalone method. (ii) Our method utilizes pose classification as an auxiliary task to supervise the feature learning procedure, whereas the counterpart methods use the pose information only as input for feature extraction. (iii) A simplified pose definition (i.e., right-headed or left-headed) is employed in this paper, which makes our method easier to train and more robust to large pose variation and self-occlusion.

4.3. Ablation Study

Contribution of Pose Classification and Complementary Learning. We evaluate the contribution of the auxiliary pose classification task and the complementary feature learning module to the superior performance of the proposed PGCFL model by using the SCU-Tiger dataset. We adapt the PGCFL model by enabling or disabling pose classification task and complementary module. When the complementary module is disabled, the image-erasure and fusion modules are disabled also; in other words, only the base module is active. When the pose classification task is disabled, only the identification loss is used to train the model. The results are summarized in Table 2. Obviously, the best mAP and top-1 identification rate are obtained when both pose classification task and complementary learning module are enabled. These results demonstrate the effectiveness of our proposed method in learning discriminative features for in-the-wild Amur tiger re-identification.

Importance of Diverse Features. In this experiment, we evaluate the necessity of learning diverse features for Amur tiger re-ID. According to Section 3, the proposed PGCFL model can generate three types of features, base features (f_{11}), complementary features (f_{21}) and unified features (f_{32}), among which the unified features are fusion of the other two types of features and thus capture more diverse information of the tigers. Given the PGCFL model trained on ATRW, we use it to extract these features for the tiger images in the SCU-Tiger dataset, and then compute the re-ID accuracy based on them. The results are shown in Table 3. Not surprisingly, the unified diverse features achieve the best results for all the performance metrics, which proves the importance of learning diverse features in Amur tiger re-ID.

Table 2. Performance of the proposed method on the SCU-Tiger dataset when pose classification task and complementary learning module are enabled or disabled.

Pose classification task	Complementary learning	mAP (%)	top-1 (%)	top-5 (%)
×	×	60.5	73.1	82.1
✓	×	74.4	82.1	**92.1**
×	✓	63.3	74.3	87.4
✓	✓	**76.4**	**84.8**	91.6

Table 3. Effectiveness of base features, complementary features and unified features for re-ID on the SCU-Tiger dataset by using the PGCFL model trained on the ATRW dataset.

Features	mAP (%)	top-1 (%)	top-5 (%)
Base (f_{11})	67.3	73.8	87.9
Complementary (f_{21})	69.0	77.4	86.3
Unified (f_{32})	**76.4**	**84.8**	**91.6**

Table 4. Impact of binarizing activation maps with different threshold values on the re-ID performance on the SCU-Tiger dataset with the PGCFL model trained on the ATRW dataset.

Threshold	mAP (%)	top-1 (%)	top-5 (%)
0.2	75.4	81.6	91.6
0.3	76.0	83.7	91.0
0.4	73.4	81.6	90.5
0.5	75.9	83.2	91.0
0.6	76.4	**84.8**	**91.6**
0.7	**76.6**	84.2	**91.6**
0.8	74.3	80.1	91.6

Impact of Thresholding Activation Maps. As introduced in Section 3.3, the image-erasure module locates the image regions to be masked out for the complementary module based on the binarized activation map of the base module. In this experiment, we use different threshold values to binarize the activation map and evaluate their impact on the re-ID performance. Table 4 gives the obtained results, according to which we observe that a threshold value in between 0.6 and 0.7 is better. In the other experiments in this paper, we set the threshold value as 0.6.

Table 5. Computational complexity of the proposed model in terms of involved number of parameters, required number of floating point operations per second (FLOPs), and the time taken to process one batch of samples during training and testing phases, respectively, when using CPU or GPU.

Phase	Parameters (M)	FLOPs (G)	CPU (s)	GPU (s)
Train	60.54	321.96	23.71	0.50
Test	60.54	321.90	9.27	0.15

4.4. Computational Complexity

To evaluate the computational complexity of the proposed method, we ran our model in batch mode on a PC with one CPU of Intel core i7-8700K and two GPUs of NVIDIA GeForce GTX 1080 Ti. We counted the seconds taken by our model for processing one batch of samples (batch size is set to 10, and image size is 288×488) during training phase and testing phase, respectively, when using CPU or GPU. The results are given in Table 5. As can be seen, our model can re-identify an Amur tiger within tens of milliseconds. The number of parameters involved in our model and the number of floating point operations per second (FLOPs) required by our model are also reported in Table 5.

5. Conclusion

In this paper, we propose a novel method to discover discriminative diverse features for Amur tiger re-identification. In order to deal with large pose variation of tiger body, unlike existing state-of-the-art methods, we employ a simplified definition of tiger pose (i.e., right-headed or left-headed) and use an auxiliary task of pose classification to supervise the re-ID feature learning. Moreover, we introduce for the first time complementary feature learning into tiger re-ID. By adaptively erasing partial regions on tiger images, our method can force different feature extraction branches to focus on different parts of tiger images and thus learn complementary features for re-identifying tigers. Extensive evaluation experiments on the ATRW dataset and our self-constructed SCU-Tiger datasets show that our proposed method can substantially improve state-of-the-art in-the-wild tiger re-ID accuracy. In the future, we are going to further extend our proposed method to other species, e.g., red pandas.

Acknowledgements

The authors would like to thank Zhicong Feng for his constructive discussion on the proposed model. All correspondences should be directed to Q. Zhao at qjzhao@scu.edu.cn.

References

[1] T. Burghardt, B. Thomas, P. J. Barham, and J. Calic. Automated visual recognition of individual african penguins. In *Fifth International Penguin Conference*, 2004.

[2] T. Chehrsimin, T. Eerola, M. Koivuniemi, M. Auttila, R. Levänen, M. Niemi, M. Kunnasranta, and H. Kälviäinen. Automatic individual identification of saimaa ringed seals. *IET Computer Vision*, 12(2):146–152, 2018.

[3] D. Crouse, R. L. Jacobs, Z. Richardson, S. Klum, A. Jain, A. L. Baden, and S. R. Tecot. Lemurfaceid: A face recognition system to facilitate individual identification of lemurs. *Bmc Zoology*, 2(1):2, 2017.

[4] D. Deb, S. Wiper, S. Gong, Y. Shi, C. Tymoszek, A. Fletcher, and A. K. Jain. Face recognition: Primates in the wild. In *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–10, 2018.

[5] M. F. Hansen, M. L. Smith, L. N. Smith, M. G. Salter, E. M. Baxter, M. Farish, and B. Grieve. Towards on-farm pig face recognition using convolutional neural networks. *Computers in Industry*, 98:145–152, 2018.

[6] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016.

[7] Q. He, Q. Zhao, N. Liu, P. Chen, Z. Zhang, and R. Hou. Distinguishing individual red pandas from their faces. In *The 2nd Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, 2019.

[8] T. He, Z. Zhang, H. Zhang, Z. Zhang, J. Xie, and M. Li. Bag of tricks for image classification with convolutional neural networks. In *CVPR*, pages 558–567, 2019.

[9] A. Hermans, L. Beyer, and B. Leibe. In defense of the triplet loss for person re-identification. *arXiv:1703.07737*, 2017.

[10] L. Hiby, P. Lovell, N. Patil, N. S. Kumar, A. M. Gopalaswamy, and K. U. Karanth. A tiger cannot change its stripes: Using a three-dimensional model to match images of living tigers and tiger skins. *Biology Letters*, 5(3):383–386, 2009.

[11] J. Hu, L. Shen, and G. Sun. Squeeze-and-excitation networks. In *CVPR*, pages 7132–7141, 2018.

[12] B. Hughes and T. Burghardt. Automated visual fin identification of individual great white sharks. *International Journal of Computer Vision*, 122(3):542–557, 2016.

[13] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv:1502.03167*, 2015.

[14] H. Jin, Z. Bochuan, L. I. Yujie, B. Wenke, Q. I. Guilan, D. Junfei, Y. Ze-Jing, and Z. Jindong. Facial recognition of giant pandas based on developmental network recognition model. *ACTA Theriologica Sinicamode*, 39(1):43–51, 2019.

[15] S. Li, J. Li, W. Lin, and H. Tang. Amur tiger re-identification in the wild. In *ICCV workshop*, 2019.

[16] W. Li, Z. Ji, L. Wang, C. Sun, and X. Yang. Automatic individual identification of holstein dairy cows using tailhead images. *Computers and Electronics in Agriculture*, 142:622–631, 2017.

[17] M. Lin, Q. Chen, and S. Yan. Network in network. *ICLR*, 2013.

[18] J. Martin, W. M. Kitchens, and J. E. Hines. Importance of well-designed monitoring programs for the conservation of endangered species: Case study of the snail kite. *Conservation Biology*, 21(2):472–481, 2010.

[19] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and F.-F. Li. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.

[20] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[21] A. Tharwat, T. Gaber, A. E. Hassanien, H. A. Hassanien, and M. F. Tolba. Cattle identification using muzzle print images based on texture features approach. *Proceedings of the Fifth International Conference on Innovations in Bio-Inspired Computing and Applications IBICA 2014*, pages 217–227, 2014.

[22] W. Yang, H. Huang, Z. Zhang, X. Chen, K. Huang, and S. Zhang. Towards rich feature discovery with class activation maps augmentation for person re-identification. In *CVPR*, pages 1389–1398, 2019.

[23] P. Zhang. Study on northeast tiger skin texture extraction and recognition based on bp network. *Doctor, Northeast Forestry University*, 2008.

[24] X. Zhang, Y. Wei, J. Feng, Y. Yang, and T. S. Huang. Adversarial complementary learning for weakly supervised object localization. In *CVPR*, pages 1325–1334, 2018.