2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)
October 11-14, 2020. Toronto, Canada

Grouping Feature Learning for Giant Panda Face Recognition

Hai-Kun Wang[1, 2], Han Su[1], Peng Chen[2, 3, 4], Rong Hou[2, 3, 4], Zhihe Zhang[2, 3, 4]

[1]College of Computer Science, Sichuan Normal University, Chengdu, China
[2]Chengdu Research Base of Giant Panda Breeding, Chengdu, China
[3]Sichuan Key Laboratory of Conservation Biology for Endangered Wildlife, Chengdu, China
[4]Sichuan Academy of Giant Panda, Chengdu, China

***Abstract*—The giant panda (panda) has lived on the earth for at least eight million years, and as an endangered species, it has received extensive attention from scholars from all walks of life. As an important part of the panda population investigation, the individual identification of pandas can not only provide useful indications but also verify the effectiveness of protection measures. Some work has introduced image processing techniques and deep learning techniques to help researchers identify pandas using face images of a panda. In this paper, we proposed a grouping feature learning method for panda face recognition. In particular, we designed a feature mapping module which can be easily embedded into the existing feature extractors, and a grouping loss function is adopted to constrain the feature mapping, allows the learned similar features to be aggregated together and increase generalization. We use the open captive panda dataset to verify our method, and the results show that our method is effective.**

***Index Terms*—giant panda, individual identification, panda face recognition**

I. Introduction

Population size is an important factor that determines whether a species can survive in nature, and is also an important indicator of regional biodiversity. Accurate estimation of the population size is an important indicator for formulating effective protection and management plans. As one of the most endangered species in the world, panda has received extensive attention for its research and protection. Knowing the populations and individuals of pandas timely and accurately is of great significance to the protection of pandas.

The traditional survey methods, including direct counting method, route survey method, and distance-bite discrimination method, and the molecular biological methods based on the feces of pandas, such as DNA fingerprint detection technology and microsatellite analysis, were employed [1]. Although some methods can accurately distinguish between different pandas, while other methods cannot even distinguish them well, these methods are also difficult to implement and perform quickly because of the changeable field environment and the difficulty of sample collection. To reduce resource consumption and obtain individual information on panda effectively, some new methods have been proposed.

With the field camera deployed in the habitat of pandas, the image-based individual recognition method has entered people's attention. The previous work [17] confirmed the feasibility of image technology in panda face recognition and achieved gratifying results. Chen et al. and Hou et al. proposed the recognition frameworks and methods based on deep learning, and they have achieved good results in different collected captive panda datasets [4], [12]. However, they use complex models and additional manual features, as well as data augmentation, which is also the consumption of computing resources and manual labeling. To avoid complex models and extra data, and enable the model to learn effective features on a limited dataset, we proposed a grouping feature learning method, which includes a feature mapping module and a corresponding constraint function. As shown in Fig.

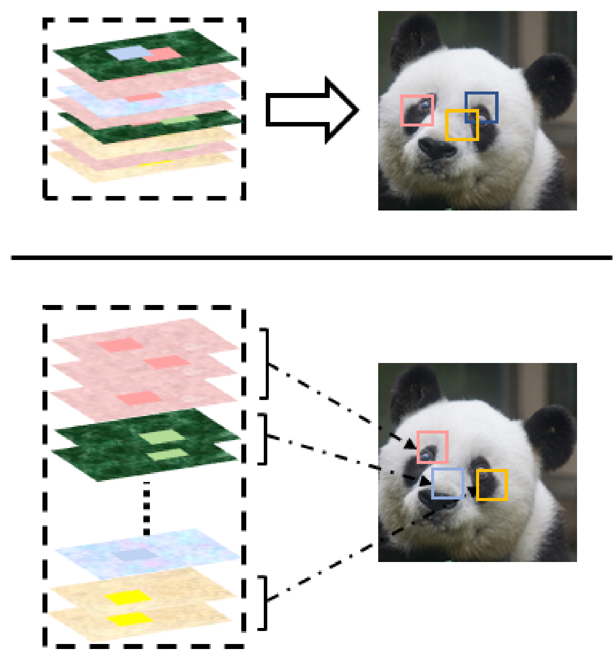
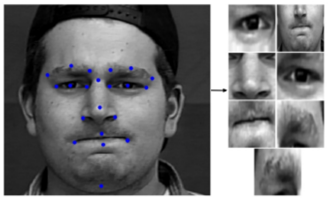
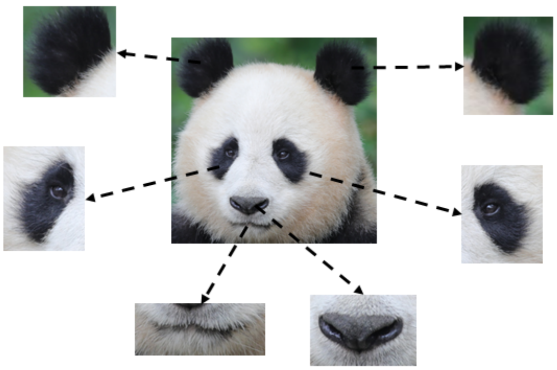


Fig. 1. The proposed grouping feature learning aims to make the originally cluttered features orderly and diverse. Above the black dividing line are the original features, they may have redundancy and duplication, and they are highly relevant. The picture below the dividing line is our goal, different feature groups are associated with different parts.



(a)



(b)

Fig. 2. Face images of human and panda. ((a) shows facial patches of a human face, which is from [7], and (b) is a panda face with its component.)

1, the features extracted by the original model are messy, disordered, and cross together. Our goal is to group these features according to similarity and distinguish the different features. Our experiment shows that these grouped features represent different facial regions, and the redundancy of features is reduced, while the diversity and distinguishability of features will be improved.

II. Related Work

A. Image-based Animal Recognition

The rapid development of deep learning has solved problems in many computer vision tasks [13], many effective models have been proposed [9], [10], [26]. These methods have been employed to analyze animal images for conservation, wildlife biology, and zoology applications. Research on tigers [15], whales [22], red pandas [11], and zebras [14] shows that we can distinguish animals based on their unique streak or spot characteristics. Freytag et al. proposed a series of chimpanzee facial attribute recognition algorithms, which can obtain individual, gender, and age information from facial images [6]. Schofield et al. and Bain et al. studied the methods of chimpanzee face detection, tracking, and recognition for field video [2], [24].

In the field of panda research, a small dataset has been used for panda face recognition [17], verified the feasibility of the image-based panda individual recognition method. Subsequently, deep learning-based methods were also applied. Hou et al. proposed an individual recognition method based on a single VGG [26] model, and SPP [8] was used for multi-scale feature fusion, the dataset they used contains 65,000 images of 25 individuals with rich poses for experiments, while they cropped the facial features manually [12]. And Chen et al. proposed a panda individual recognition framework used a series of cascade models for automatic face detection, landmark alignment, and recognition, while data augmentation and extra supervised label are used, the dataset includes 6,441 images of 218 individuals which include 39.78% of captive pandas in the world, and it has been released [4]. Although their methods are effective, they either use large amounts of data or complex models. In this paper, we proposed a new method that will be easily used and cost less, and we will use the released panda images dataset [20] for our study.

B. Facial Feature Learning

Facial feature learning has always been a research hotspot in face recognition [16], [27], [29]. It is an effective method to use facial features for face recognition and facial expressions recognition [7], [19]. However, the sparse representation of face recognition also shows that not all facial features are effective, and it has certain redundancy, Zeng et al. proposed deep sparse autoencoders for facial expression recognition [30]. Focusing on effective and diverse features of the face will be effective for recognition. Earlier works on face recognition are based on manual features, which are always cropping the facial components for feature extract, the facial components usually indicate eyes, nose, ears, and mouth, as shown in Fig. 2a. Dadi et al. used the gridding facial features for face recognition [5], and Polikovsky et al. proposed 3D HOG for facial micro-expression recognition, also divided faces into some specific areas [21]. Hazour et al. aggregates multi facial patches to recognize the facial expression [7].

However, panda faces are different from human faces, and the recognition of panda faces is much more difficult. It can be observed from Fig. 2, the facial features of human faces are more planar, and each facial component has obvious raised features, and each face has unique facial texture features, while the panda faces are more three-dimensional, due to the hair coverage, it lacks facial landmarks, texture, and expression and other information. The faces of different pandas are even similar because almost all panda faces are in black and white, (e.g. black eyes, ears, and white faces), which brings great difficulties to the identification of panda. Therefore, we tend to use deep features to recognize the pandas, and here is an example to classify the gender of panda via deep feature learning [28]. It has been observed in some works that feature extracted by convolutional channels in high-level layers will represent specific semantic patterns [3]. As shown in Fig. 1, the original features may be disordered and redundant, so we associate similar semantic features together to reduce feature redundancy and cluttery. We use these patterns to identify individual pandas and introduce a special feature mapping module to map the features of the last convolutional layer, we divide the features into similar groups to represent the whole panda face, the unnecessary features will not be mapped, and a loss function is used to constraint the module.

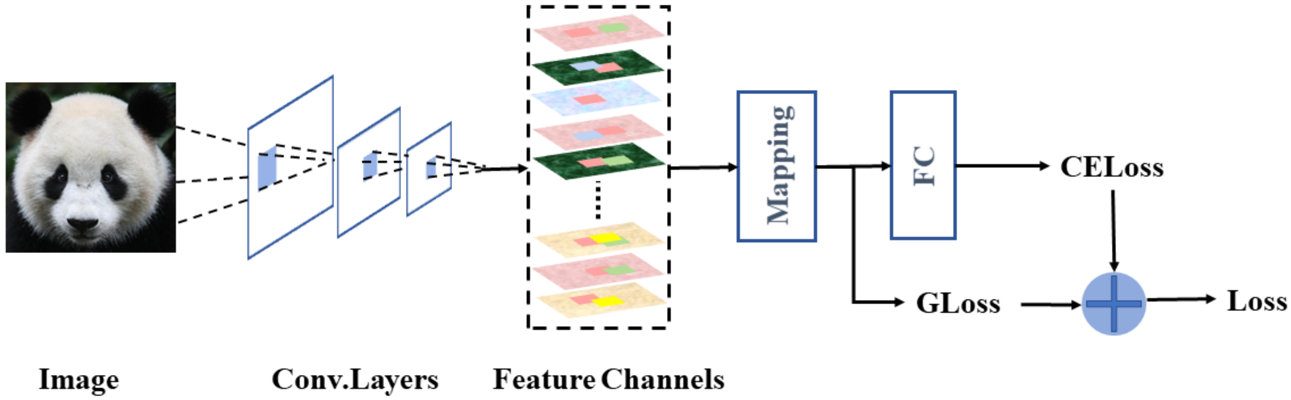


Fig. 3. The framework of the model we used with the proposed grouping feature learning. Conv. Layers represent the convolution layer, which is the basic part of feature extractor, which can be replaced by some existing models, such as VGG [26] and ResNet [9]. Mapping is the proposed module, embedded behind the last convolutional layer. And CELoss denotes the cross-entropy loss and GLoss denotes the grouping loss, The final loss is obtained by adding the two losses.

III. Proposed Method

In this section, we further explain the proposed grouping feature learning method and corresponding components. The model architecture we used in this paper is shown in Fig. 3.

Assume that there is an input image X_i of panda face, we first extract the feature maps by feeding it into a basic feature extractor, (e.g. VGGNet [26] and ResNet [9]). The extracted feature maps of the last convolutional layer can be represented as $\mathcal{F} \in \mathbb{R}^{C\times H\times W}$, where C is the feature dimensions, with height H and width W. The feature extraction process can be expressed as:

$$\mathcal{F} = fi(...f_0(X_i;\theta_0);\theta_i) \tag{1}$$

where f_i indicates the convolution operation of the ith convolutional layer, and θ_i indicates the parameters. To be more detailed, feature $\mathcal{F}$ can be expressed as a combination of multiple features, as:

$$\mathcal{F} = [m_1, m_2, ..., m_i, ..., m_C] \tag{2}$$

where m_i is the ith feature, which is expressed as a two-dimensional matrix, the shape of m_i is $H \times W$.

As the convolution parameters are agnostic and there are no other constraints to distinguish these parameters, the features will be repeatedly extracted, which will cause redundant semantic information, making the features indistinguishable and lacking in diversity. Therefore, we propose to group features according to similar semantics, and a feature mapping module is adopted to map the extracted features to the semantically similar feature groups. The mapping operation can be expressed as:

$$\widetilde{\mathcal{F}} = f_{mapping}(\mathcal{F};w) \tag{3}$$

where $f_{mapping}$ indicates the mapping function and w indicates the parameters.

We pre-defined the number of groups as G, then each semantic group contains C/G feature maps, and we set G to be divisible by C to ensure that each semantic group has the same number of feature maps. The group index of the ith mapped feature G_i is computed as $\lfloor i/G \rfloor$. Features with the same index are divided into a group, and we hope that the semantics represented by each semantic group is similar. To constraint the relationship of the mapping function, the grouping loss is used.

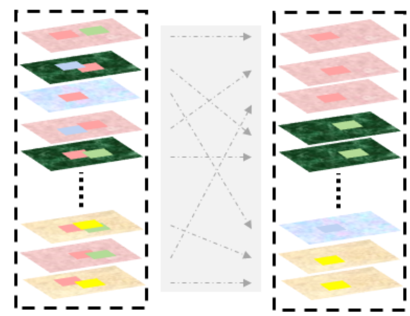


Fig. 4. The feature mapping operation is shown as the gray area in the figure. The order and distribution of features have changed after this operation.

A. The Feature Mapping Module

As mentioned above, the feature mapping module is used to map the extracted features into groups with similar semantics. And we do not want to change the size of the input feature maps, and this mapping function should be a linear operation, as shown in Fig. 4. This transformation can be expressed as:

$$\begin{aligned}\widetilde{\mathcal{F}} &= A^T\mathcal{F} \\ &= A^T[m_1, m_2, ..., m_i, ..., m_C] \\ &= [\widetilde{m}_1, \widetilde{m}_2, ..., \widetilde{m}_i, ..., \widetilde{m}_C]\end{aligned} \tag{4}$$

where $A \in \mathbb{R}^{C\times C}$ is the mapping matrix. This kind of process is similar to convolution operation, and a 1×1

convolutional layer without bias is adopted, the input channel and output channel are both C, to keep the feature dimension unchanged. And the parameters in (3) of the mapping function can be replaced by the convolution parameters. The 1×1 convolutional layer can be easily embedded into any existing models, and the parameters are learnable, which is exactly what we need.

We also introduced CWA proposed in [3] which is a short notation for the Channel-Wise Attention and is actually a randomly channel-wise drop function. This operation can reduce overfitting to some extent, and force the model to learn more effective features. We set a proportional coefficient of $p \in (0, 1)$ to determine the number of random zeroes for each group of features. The features will be $M \cdot \widetilde{\mathcal{F}}$, where $M \in \mathbb{R}^{C\times1\times1}$ is a binary mask.

B. The Grouping Loss Function

Since the mapping module is actually a linear transformation, it may not have any effect on the original features. The original features will be mapped to the final features directly, both valid and invalid. To force the mapping function to map valid features with similar semantics to corresponding semantic groups, additional constraints are needed. We hope that the semantic features of the same group are as similar as possible, and the features of different groups are as different as possible, the expected loss function for the ith feature map can be expressed as:

$$\begin{aligned} L_i &= -L_{within-group} + L_{inter-group} \\ &= -\sum_{\substack{0<j<C \\ (G_i=G_j)}} s_{ij}^2 + \sum_{\substack{0<j<C \\ (G_i\neq G_j)}} s_{ij}^2 \end{aligned} \tag{5}$$

where s_{ij} denotes the similarity between $\widetilde{m}_i$ and $\widetilde{m}_j$, G_i is the index of the group to which the ith feature belongs. To express the similarity of these semantic features, cosine similarity [18] is used, which can measure the similarity of two space vectors. The formula is:

$$s_{ij} = \frac{\widetilde{m}_i^T \widetilde{m}_j}{\|\widetilde{m}_i\|_2 \|\widetilde{m}_i\|_2} \tag{6}$$

note that $\widetilde{m}$ has been flattened into a one-dimensional vector before calculating the similarity. As our goal is to maximize the within-group similarity and minimize the inter-group similarity, thus there is a minus sign front $L_{within-group}$ in (5). Since the grouping loss function compulsorily maps similar semantic features, the group independent features can also be filtered out. The whole grouping loss is computed as:

$$L_G = -\frac{1}{C}\sum_{i}^{C} L_i \tag{7}$$

The final loss function of the whole model can be defined as:

$$L = L_{CE} + \lambda L_G \tag{8}$$

where L_{CE} indicates the cross-entropy loss, is a commonly used loss function for classification models, and λ is a trade-off parameter to balance the influence of the two losses.

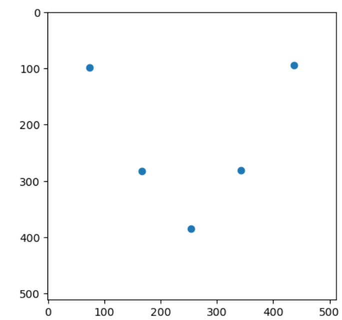


Fig. 5. The standard facial key points.

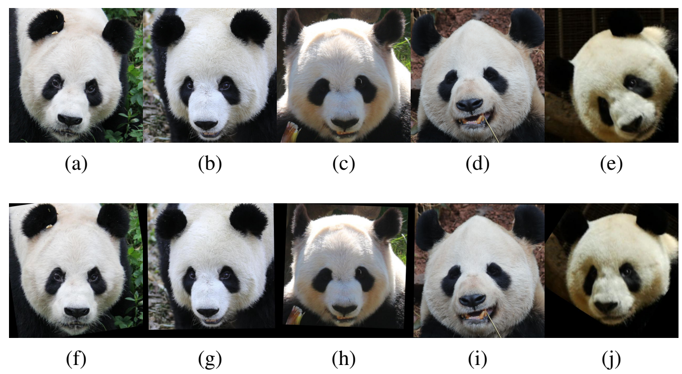


Fig. 6. Samples of panda images. (Images in the first row are the original cropped faces, and the second row is aligned faces.)

IV. Experiment and Results

A. Dataset

We validate our method on the released panda images dataset [20], which contains 6,441 images of 218 captive panda individuals. The dataset also contains some labeling information such as bounding boxes and facial landmark masks. We use the bounding boxes labels to crop out the head image and resize it to 512×512. It should be noted that during the cropping process, we use reduction or increase the side length to ensure that the aspect ratio is 1: 1. We also selected 200 facial images with the facial landmark annotations, to obtain a unified facial key point for face alignment, because we can not get panda faces under constraints, even though they are all frontal faces. The standard facial key points obtained are shown in Fig. 5, and the statistics of the dataset is provided in Table I.

During the training phase, the images are resized to 224×224, and a random crop method with padding 4 is used to crop the images to 224×224 to fit the input of the network. In the test phase, just resize the image to 224×224 pixels. We also use affine transformation to align the input images, and the shadow is filled with 0. Some samples of the original images and the aligned images are shown in Fig. 6.

B. Experimental Setup and Results

We conduct experiments on a 64-bit Ubuntu 18.04 computer with an Intel 3.5 GHz CPU and an NVIDIA RTX 2080 Ti

TABLE I
STATISTICS OF THE PANDA IMAGES DATASET

Dataset	#Individuals	#Training	#Testing
GiantPanda	218	6039	402

TABLE II
RESULTS OF DIFFERENT METHODS FOR PANDA FACE RECONGNITION. (GFL DENOTES GROUPING FEATURE LEARNING.)

Methods	Accuracy
ResNet18+STN+mask [4]	95.02
ResNet50+STN+mask [4]	96.27
ResNet18+GFL(ours)	94.52
ResNet50+GFL(ours)	95.02
ResNet18+GFL(ours)+align	95.77
ResNet50+GFL(ours)+align	**97.26**

GPU. The model is implemented with Python 3.6 and PyTorch 1.3.0. In detail implementation Stochastic Gradient Descent (SGD) optimizer with momentum 0.9 and weight decay 5e-4 is used in the training phase, The initial learning rate is 0.01, and it decays by 0.1 every 30 epochs, the batch size is set to 32 and total epoch is 90.

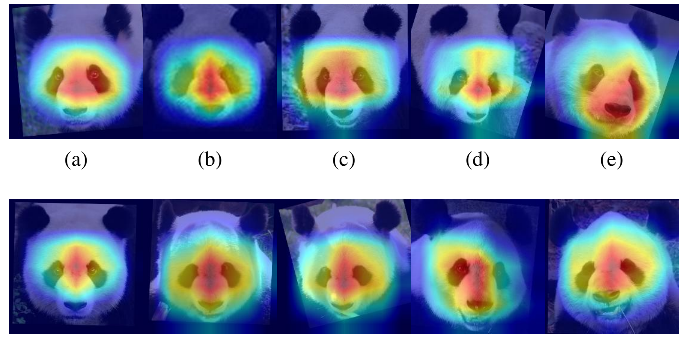
We use the pre-trained ResNet18 and ResNet50 [9] as feature extractors, and a feature normalized method is used before the last fully connected layer to reduce the influence of image resolution and make training more stable [23], features are normalized and multiplied by a scale parameter 12. Moreover, as the purpose of λ in (8) is to balance the two losses, so we scale them to the same order of magnitude, for example, if the cross-entropy loss is optimized from a single digit, then the grouping loss is also scaled to less than 10, and we set λ to 0.0005 in our experiments. And the drop probability p mentioned in CWA is 0.5, same with [3]. The number of groups is 64 for ResNet18 and 128 for ResNet50, the results are shown in Table II. As the results in the table show, face alignment is needed as the result using aligned faces get a much better result because the dataset does not have enough data for the model to learn [12]. We also use Grad-Cam [25] to visualize the regions which the model focuses on. It can be seen from Fig. 7 that the model mainly focuses on the center areas of faces, which is consistent with the results in previous work [4]. However, the panda faces are more three-dimensional, and the deviation of the angle will have a great impact on facial feature extraction, the results also confirm this fact.

TABLE III
INFLUENCE OF GROUP NUMBERS ON PANDA IMAGES DATASET.(TRAINED FROM SCRATCH)

Methods	GroupNums	Accuracy
ResNet18 + GFL	1	85.82
ResNet18 + GFL	8	85.57
ResNet18 + GFL	16	85.82
ResNet18 + GFL	64	86.82
ResNet18 + GFL	128	85.32
ResNet18 + GFL	256	84.82

TABLE IV
QUANTITATIVE COMPARISONS OF THE PROPOSED MODULES

Methods	GroupNums	Accuracy
ResNet18	None	92.78
ResNet18 + Mapping	64	93.03
ResNet18 + Mapping + GLoss	64	94.22
ResNet18 + Mapping + GLoss + CWA	64	94.52



(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

Fig. 7. Visualization of the localized regions returned from Grad-CAM [25].

C. Discussion

We did additional experiments based on ResNet18 to study the impact of the number of groups, and the models are trained from scratch. As shown in Table III, when the group number is 256, we get the lowest result, the group is too much that the model will be overwhelmed, and get the best result while the number of groups is 64. It is obvious that when the group number is 256, there are only two channels in each group, as feature channels of ResNet18 is 512, so more groups will be separated and less similar, but the facial features are limited, which will lead to degeneracy. Therefore, it is necessary to select the number of groups appropriately.

We also made some quantitative comparisons of the proposed modules, using the pre-trained ResNet18 as the feature extractor. As we mentioned earlier, the mapping module not only groups the features but also discards some unnecessary features. The pre-trained model trained on the ImageNet dataset, which will make the convolutional layers respond to features other than the faces. And useless features will be filtered out after the mapping module, and constraints are added to make the mapping more effective, results in Table IV can support these claims. On the other hand, CWA does have some enhancement effects, and it reduces overfitting to some extent.

V. CONCLUSION

In this paper, we propose a grouping feature learning method for panda face recognition, which includes a feature mapping operation and a constraint function and can be easily embedded into the existing feature extractors without adding much burden of inference. We have verified on the released panda image dataset [20], and through visualization, we can find that our method is effective. In future work, we will focus on the diversity and explainability of specific semantic features, to find more effective features to avoid the impact of face alignment, and apply the proposed method in other fields.

ACKNOWLEDGEMENT

This research is supported by Chengdu Research Base of Giant Panda Breeding (NO. CPB2018-02, CPB2018-01), the National Natural Science Foundation of China (61403266, 61403196, and 31300306), Chinese overseas returnees science and technology activities project funding of the Ministry of Human Resources and Social Security, and the Sichuan Science and Technology Program (2018JY0096).

REFERENCES

[1] State Forestry Administration. The third national survey report on giant panda in china, 2006.

[2] Max Bain, Arsha Nagrani, Daniel Schofield, and Andrew Zisserman. Count, crop and recognise: Fine-grained recognition in the wild. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 0–0, 2019.

[3] Dongliang Chang, Yifeng Ding, Jiyang Xie, Ayan Kumar Bhunia, Xiaoxu Li, Zhanyu Ma, Ming Wu, Jun Guo, and Yi-Zhe Song. The devil is in the channels: Mutual-channel loss for fine-grained image classification. *IEEE Transactions on Image Processing*, 29:4683–4695, 2020.

[4] Peng Chen, Pranjal Swarup, Michal Matkowski Wojciech, Adams Wai Kin Kong, Su Han, Zhihe Zhang, and Hou Rong. A study on giant panda recognition based on images of a large proportion of captive pandas. *Ecology and Evolution*.

[5] Harihara Santosh Dadi and GK Mohan Pillutla. Improved face recognition rate using hog features and svm classifier. *IOSR Journal of Electronics and Communication Engineering*, 11(04):34–44, 2016.

[6] Alexander Freytag, Erik Rodner, Marcel Simon, Alexander Loos, Hjalmar S Kühl, and Joachim Denzler. Chimpanzee faces in the wild: Log-euclidean cnns for predicting identities and attributes of primates. In *German Conference on Pattern Recognition*, pages 51–63. Springer, 2016.

[7] Ahmed Rachid Hazourli, Amine Djeghri, Hanan Salam, and Alice Othmani. Deep multi-facial patches aggregation network for facial expression recognition. *arXiv preprint arXiv:2002.09298*, 2020.

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 37(9):1904–1916, 2015.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *European conference on computer vision*, pages 630–645. Springer, 2016.

[11] Qi He, Qijun Zhao, Ning Liu, Peng Chen, Zhihe Zhang, and Rong Hou. Distinguishing individual red pandas from their faces. In *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, pages 714–724. Springer, 2019.

[12] Jin Hou, Yuxin He, Hongbo Yang, Thomas Connor, Jie Gao, Yujun Wang, Yichao Zeng, Jindong Zhang, Jinyan Huang, Bochuan Zheng, et al. Identification of animal individuals using deep learning: A case study of giant panda. *Biological Conservation*, 242:108414, 2020.

[13] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[14] Santosh Kumar and Sanjay Kumar Singh. Visual animal biometrics: survey. *IET Biometrics*, 6(3):139–156, 2016.

[15] Shuyuan Li, Jianguo Li, Weiyao Lin, and Hanlin Tang. Amur tiger re-identification in the wild. *arXiv preprint arXiv:1906.05586*, 2019.

[16] Iacopo Masi, Yue Wu, Tal Hassner, and Prem Natarajan. Deep face recognition: A survey. In *2018 31st SIBGRAPI conference on graphics, patterns and images (SIBGRAPI)*, pages 471–478. IEEE, 2018.

[17] Wojciech Michal Matkowski, Adams Wai Kin Kong, Han Su, Peng Chen, Rong Hou, and Zhihe Zhang. Giant panda face recognition using small dataset. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 1680–1684. IEEE, 2019.

[18] Hieu V Nguyen and Li Bai. Cosine similarity metric learning for face verification. In *Asian conference on computer vision*, pages 709–720. Springer, 2010.

[19] Omkar M Parkhi, Andrea Vedaldi, and Andrew Zisserman. Deep face recognition. 2015.

[20] Chen Peng, Pranjal Swarup, Wojciech Michal Matkowski, Wai Kin Kong Adams, Su Han, Zhang Zhihe, and Hou Rong. Panda images dataset, 2020.

[21] Senya Polikovsky, Yoshinari Kameda, and Yuichi Ohta. Facial micro-expressions recognition using high speed camera and 3d-gradient descriptor. 2009.

[22] Andrei Polzounov, Ilmira Terpugova, Deividas Skiparis, and Andrei Mihai. Right whale recognition using convolutional neural networks. *arXiv preprint arXiv:1604.05605*, 2016.

[23] Rajeev Ranjan, Carlos D Castillo, and Rama Chellappa. L2-constrained softmax loss for discriminative face verification. *arXiv preprint arXiv:1703.09507*, 2017.

[24] Daniel Schofield, Arsha Nagrani, Andrew Zisserman, Misato Hayashi, Tetsuro Matsuzawa, Dora Biro, and Susana Carvalho. Chimpanzee face recognition from videos in the wild using deep learning. *Science advances*, 5(9):eaaw0736, 2019.

[25] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017.

[26] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[27] Daniel Sáez Trigueros, Li Meng, and Margaret Hartnett. Face recognition: from traditional to deep learning methods. *arXiv preprint arXiv:1811.00116*, 2018.

[28] Hongnian Wang, Han Su, Peng Chen, Rong Hou, Zhihe Zhang, and Weiyi Xie. Learning deep features for giant panda gender classification using face images. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 0–0, 2019.

[29] Nannan Wang, Xinbo Gao, Dacheng Tao, Heng Yang, and Xuelong Li. Facial feature point detection: A comprehensive survey. *Neurocomputing*, 275:50–65, 2018.

[30] Nianyin Zeng, Hong Zhang, Baoye Song, Weibo Liu, Yurong Li, and Abdullah M Dobaie. Facial expression recognition via learning deep sparse autoencoders. *Neurocomputing*, 273:643–649, 2018.