Learning Deep Features for Giant Panda Gender Classification using Face Images

Hongnian Wang
Sichuan Normal University
Chengdu, Sichuan
hongnianwang@gmail.com

Han Su
Sichuan Normal University
Chengdu, Sichuan
jkxy_sh@scinu.edu.cn

Peng Chen, Rong Hou, Zhihe Zhang
Chengdu Research Base of Giant Panda Breeding
Chengdu, Sichuan
capricorncp@163.com, 405536517@qq.com, zzh@panda.org.cn

Weiyi Xie
Sichuan Normal University
Chengdu, Sichuan
xwylove@gmail.com

Abstract

Giant panda (panda) has lived on earth for at least eight million years and is known as the living fossil. It is also a vulnerable species which requires urgent protection. It is essential to conduct population survey collecting information of their population, density, age structure, and gender ratio so as to design protection schemes and measure their effectiveness. However, it is challenging to accurately and timely obtain gender ratio of pandas because their pelage lacks distinguishable gender patterns and panda is sparsely distributed population in large habitats. All current approaches rely heavily on manual collection of samples in the wild, which are time consuming, costly, or even dangerous. With the widely deployed camera traps, if the gender of pandas can be determined from images, it is possible to monitor panda gender ratio in different regions in real-time. However, no such study was done. In this paper, a deep learning method is developed to study the distinctiveness of panda face for gender classification, in which the largest panda image dataset with 6,549 panda face images collected from 100 male and 121 female pandas is established. The experimental results show that panda faces contain some gender information, although they look very similar to human vision.

1. Introduction

Panda is a national treasure of China, living in 6 mountain ranges divided into 33 subpopulations [1]. The Chinese government and international organizations such as World Wildlife Fund (WWF) are putting significant effort for its survival. Monitoring their population dynamics, including gender ratio, helps identify and predict population changes and other habitat problems. Gender ratio referring to the ratio of the number of males to females in a population is a critical factor for population growth and decline [20]. For example, if male pandas are much more than female pandas in a region, it implies that some male pandas are hard to find mating partners. If ecologists and wildlife researchers are able to know the gender ratio of panda in time, they can utilize emergency mechanisms through manual intervention to prevent population decline. However, it is not easy to estimate gender ratio of panda because panda is not sexual dimorphism, meaning that two sexes have no externally displayed different characteristics beyond the differences in their sexual organs [15]. Even for panda breeders, it is hard to discriminate their genders.

Currently, there are two primary non-invasive methods, bite-size and DNA-based approaches, to identify individual giant pandas and obtain their gender information. Bite-size approach using bamboo in feces is simple and practical, but it lacks accuracy [27]. The DNA-based approach utilizing sex-specific gene is reliable, but its effectiveness highly depends on the freshness of the sample [28]. All these two approaches require huge manpower resources to collect the samples from panda habitats, which are expensive, time consuming or even dangerous. Furthermore, these approaches cannot provide real time information. Currently, Chinese government can only perform panda population survey for every ten years. Once the panda population declines, information obtained from the survey is probably too late for implementing protection measure. Thus, new techniques for effectively estimating panda gender ratio are still in-demand.

Camera traps have been increasingly installed in the

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

Figure 1. Animal with genderual dimorphism patterns. ((a) and (b) are lions; (c) and (d) are mandrills; (e) and (f) are peacocks. Red and blue underline in images mean the gender of animal is female or male, respectively. Images from https://www.mnn.com)

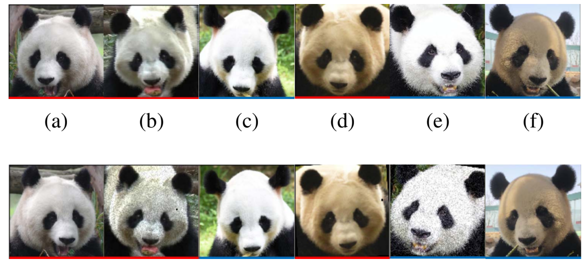
wild, including the habitats of panda [6, 21]. It provides a great opportunity to apply computer vision approaches for wildlife conservation [16, 5]. To finally leverage the potential of the camera traps and computer vision methods for panda conservation, the distinctiveness of panda face for gender classification is investigated. Comparing with the panda body, panda face has less motion that makes gender classification easier. It should be noted that, this study does not use images captured by camera traps because the number of qualified images received from the camera traps is extremely small up to now [19]. Furthermore, there is no guarantee to get accurate gender information for supervised learning. Our contributions are as follows:

- A large panda face dataset with 6,549 panda images from 100 male and 121 female pandas is established.
- The study shows that panda face contains gender discriminative information. According to the best knowledge of the authors, no one studied gender classification via panda face images before.
- The regions containing gender discriminative information are highlighted for ecologists and wildlife researchers for further investigation.

The rest of this paper is organized as follows. In the next section, related work is briefly reviewed. Section 3 presents the dataset and methodology. Section 4 provides experimental results. Section 5 gives some conclusive remarks.

2. Related Work

Researchers have utilized computer vision technology for animal studies for many years. A survey of visual animal biometrics is presented in[13]. Recently, remarkable advances in deep learning drive researchers to apply it to various domains, including wildlife and zoology studies [12, 16]. Endangered species, including panda, tiger and lion, are important for biodiversity and environmental health. There are some work based on deep learning technology to protect them, including whale recognition[22], red panda identification[10], and tiger trajectory tracking[16]. In the case of gender classification, some animals have the manifestation of sexual dimorphisms, such as lion's mane, peacock and mandrills' coloration, as shown in Fig. 1, the gender of which are easy to recognize by human vision. However, the gender of some animals is harder to be differentiated, such as panda and polar bear. The existing methods for animal gender classification based on computer vision mainly focus on footprint images. They used handcrafted features, such as geometric and Gabor features [25], and traditional classifiers to recognize animals gender, including tiger [7], puma [2] and panda [15]. However, the footprint collection extremely depends on the substrate of the environment and is still dangerous. It is also expensive because the collection is still manual, instead of setting the camera traps. Comparing with traditional computer vision methods, deep learning has the ability to learn high-level features from massive datasets and offer accurate results in most of cases [17, 26]. Using deep learning to classify human gender based on face images has achieved very accurate results [3, 14, 4]. It relies heavily on the availability of large-scale dataset, but there was no such database available for applying deep learning for panda study [19]. For this study, the largest panda face image dataset with gender labels is established. Note that its size is still much smaller than human face datasets for gender classification. As shown in the first row of Fig. 2, the pandas have a similar appearance, like white hair and dark circles. For these reasons, we consider this problem as a fine-grained binary classification, which is that input face images to output the probability of male or female, under the high intra-class vs. low inter-class variance challenge.



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

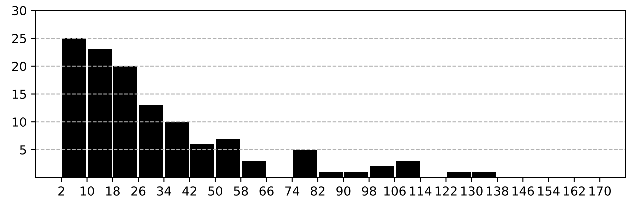
(g) (h) (i) (j) (k) (l)

Figure 2. Example of raw and augmentation images. (The first row is example images of raw images, and the second row is example images of augmentation image.)

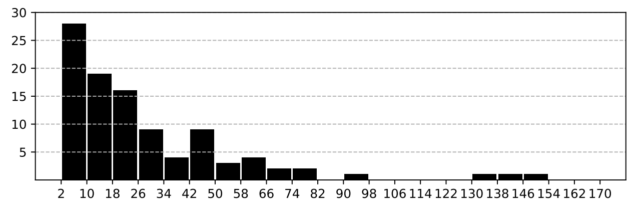
3. Dataset and Methodology

3.1. Dataset

Panda images and gender information used in this study were collected by the Chengdu Research Base of Giant Panda Breeding. Their detailed information, such as genealogy, gender, and date of birth are strictly recorded. The dataset consists of 6,549 images from 100 male pandas and 121 female pandas. The different pandas have differ-



(a)



(b)

Figure 3. Distribution of individual data. (The X axis is the number of images, and the Y axis is the number of individual pandas. (a) and (b) refer to histograms of female and male, respectively.)

ent numbers of images ranging from 2 images to 151 images and the histogram of images is shown in Fig 3. It can be seen that the pandas mostly have only about 20 images and the females and males have roughly the same data distribution. The photoshoots experience an extensive time span within recent years and are taken by the professional panda photographer. The cameras used include Panasonic dvx200 video camera, Canon 1DXmarkII camera, Canon 5DmarkIII camera, and Panasonic Lumix DMC-GH4 camera.

The frontal face regions were manually cropped and resized to 512 by 512 pixels. In order to avoid individual panda information influencing the results, the dataset is split to a training set and a testing set based on panda identity. The pandas are first ranked in descending order, according to the number of their images. Then, the first two pandas are put in the training set, the next one is placed in the testing set, and the last two are put in the training set. The selection process is repeated for the next five pandas and the rest of the pandas. This ensures that the ratios of the training set and the testing set in terms of both images and pandas are roughly 8 to 2. Detailed statistics about the training and testing datasets are listed in Table 1.

Although the dataset is the largest panda face dataset, it is still small for training deep learning method. To alleviate this problem, as with training other deep networks, data augmentation is applied. More clearly, random noise, rotation, cropping, erasing, horizontal and vertical flips, translations, and color distortions are applied. The second row in Fig. 2 shows some augmented data.

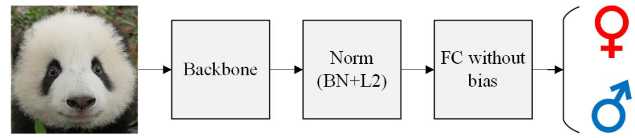


Figure 4. Architecture for gender classification (Backbone refers to convolutional neural network that removes the classification layer. BN, L2 Norm and FC are respectively the short forms of batch normalization, L2 normalization and fully connected layer.)

3.2. Methodology

As shown in the figures above, panda gender cannot be distinguished by human vision. To learn high-level features that may not be obvious to human vision, Convolutional Neural Network (CNN) with residual learning is employed. To enhance the stability of the training, Batch Normalization (BN) and L2 normalization are adopted as norm layer. Finally, the final result is outputted by a Fully Connected (FC) layer. The network architecture is shown in Fig. 4.

CNN is effective in exploiting spatial dependencies by using the local connection, which has more advantages than other types of neural networks in CV tasks [12]. The idea of residual learning is to use an additional skip connection as a shortcut, and deeper layers can directly access the features from previous layers. This method has provided state-of-the-art results for multiple challenging CV tasks, including image classification, object detection, segmentation, and localization [9]. Formally, a residual unit with an identical mapping [9] is defined as:

$$\boldsymbol{X}^{l} = \mathcal{F}(\boldsymbol{X}^{l-1}; \theta^{l}) + \boldsymbol{X}^{l-1} \tag{1}$$

where $\boldsymbol{X}^{l-1}$ and $\boldsymbol{X}^{l}$ are the input and output of the l-th residual unit, respectively. $\mathcal{F}(;\theta^{l})$ denotes the residual function, which implements the composition of two convolutions parameterized by weights θ^{l} and the application of the ReLU [8]. The central idea of the residual learning is to learn the additive residual function $\mathcal{F}(;\theta^{l})$ with respect to $\boldsymbol{X}^{l}$ [9]

For this study, an ImageNet pre-trained ResNet [8] model is modified. Its last classification layer is removed and the backbone storing the information from ImageNet is retained. Batch Normalization (BM) is added at the end backbone network. BN is a technique for improving training speed, performance, and stability of neural networks via imposing Gaussian distribution on intermediate batch features [11]. It can be described by the equation:

$$\hat{X}_{avg} = \frac{X_{avg} - \mathrm{E}[X_{avg}]}{\sqrt{\mathrm{Var}[X_{avg}] + \epsilon}} * \gamma + \beta \tag{2}$$

where X_{avg} is the features outputted by the backbone, $E(\bullet)$ represents the expected value and $Var(\bullet)$ represents the variance. The expected value and the variance are calculated per-dimension over the mini-batches. γ and β are learnable parameters. L2 normalization is used to enhance the stability of the high-level features. A FC layer with 2 neurons, but without bias is added to the last layer, which outputs the probability of male or female. For data augmentation, the input augmentation images are randomly resized and cropped to 448 by 448 pixels.

To visualize the results, Gradient-weighted Class Activation Mapping (Grad-CAM) and t-distributed Stochastic Neighbor Embedding (t-SNE) are used. Grad-CAM is a class-discriminative localization method. It assigns a score to each class (e.g, female and male) using the backpropagation-based filter gradient and convolution activation values [23]. It provides a way to look into what particular parts of an image influencing the model decision. Mathematically, the class-discriminative localization map Grad-CAM L^c is computed as follows:

$$L^c = f(\sum_k a_k^c A^k) \tag{3}$$

where A^k refers to the k-th feature map of a CNN layer, a_k^c is obtained by global average pooling the gradient of the score for class c as y_c with respect to feature maps A^k. $f(\bullet)$ is ReLU to highlight features with positive influence on the class of interest.

t-SNE is a nonlinear dimensionality reduction technique through constructing the distribution of the samples [18]. The results are obtained by minimizing the Kullback-Leibler (KL) divergence, which can be written as follows:

$$C = \sum_i KL(P_i||Q_i) = \sum_i \sum_j p_{j|i} log \frac{p_{j|i}}{q_{j|i}} \tag{4}$$

Suppose that there are two raw data points x_j and x_i, y_j and y_i are the corresponding points of x_j and x_i in a low dimensional space (e.g, two-dimensional space). $p_{j|i}$ represents the similarity of x_j and x_i, and $q_{j|i}$ represents the corresponding similarity in a low-dimensional space. In the equation, P_i and Q_i denote the conditional probability distribution over all other data points given data points x_i, and Q_i represents the conditional probability distribution over all other map points given map point y_i.

4. Experiment

4.1. Settings

We conduct experiments on a 64-bit Ubuntu 16.04 computer with an Intel 3.40 GHz and an NVIDIA GTX 1080Ti GPU. The model is implemented with Python 3.6 and Pytorch 0.4.1. In detail implementation, Stochastic Gradient Descent (SGD) with momentum is adopted as the optimization method and cross entropy with L2 regularization as the loss function. The learning rate is initialized to 0.003 and we half it in every 5 epochs. The learnable parameters are initialized using the default parameters in Pytorch and the batch size set to 8. To avoid overfitting, the first two block parameters of all ResNet are frozen, and we only train the last two blocks and the additional classification layer with the fixed epochs. To obtain the reliable experimental result, we split the dataset based on the panda individual, as shown in Table 1.

Gender	Train		Test	
	#IDs	#Images	#IDs	#Images
Female	97	3,030	24	748
Male	80	2,201	20	570

Table 1. Distribution of dataset (#ID means the number of panda individual.)

Methods	Accuracy
HOG+SVM	58.7%
VGG11+BN	73.4%
ResNet18	74.9%
ResNet50+Norm	72.5%
ResNet34+Norm	73.1%
ResNet18+Norm	**77.2%**

Table 2. Gender classification experiment results

4.2. Results

We conduct the experiments on support vector machine (SVM) with linear kernel, VGG11 with BN [24] and different depth ResNet, including ResNet18, ResNet34, and ResNet50.

The results are given in Table 2. The accuracy of all learning-based model outperforms the traditional method. Specifically, The accuracy of ResNet18 is 1.5% higher than VGG11 with BN, because the deep network can learn feature representation better via residual learning. The ResNet18 with BN and L2 is better than ResNet18, implying that BN and L2 normalization is beneficial. It is also better than ResNet34 and ResNet50 as backbone. The Receiver Operating Characteristic (ROC) curves of ResNet18 as backbone model are given in Fig. 5. The results show the mean Area Under the Curve (AUC) for two-class is 0.84, indicating that panda face contains some gender information.

One unanticipated finding was that most of the error cases are from young pandas when we analyze misjudgment. In other words, the old pandas are easier to be classified than the young ones. It is true for human as well. We also do ablation study on gender information, the experimental steps are as follows: we remove all male images,

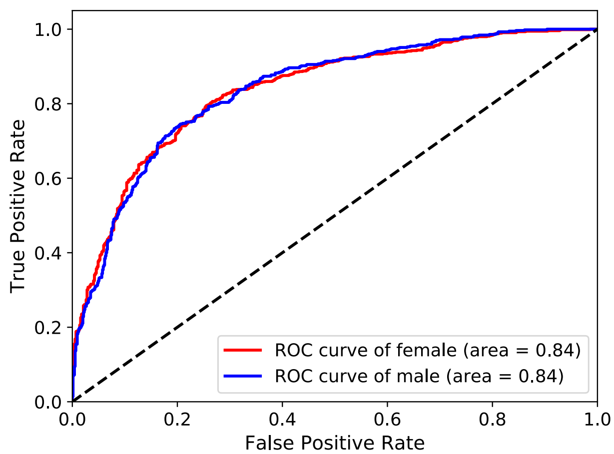


Figure 5. Receiver operating characteristic curve of ResNet18 as backbone.

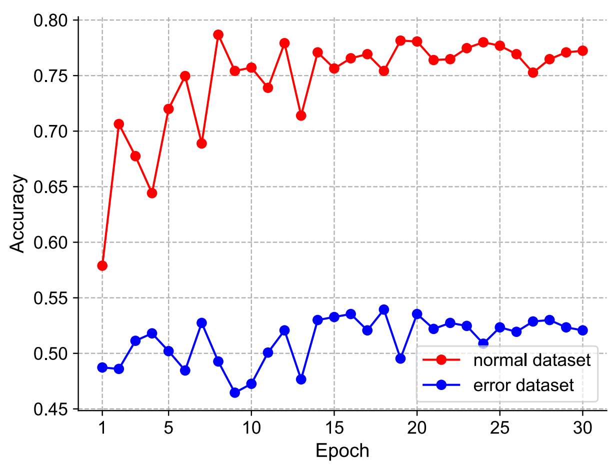


Figure 6. Testing accuracy for model on normal dataset and error dataset.

randomly reverse label of half female images, called error dataset. Then we redo the experiment with error dataset. Fig. 6 plots the testing accuracy of the model on normal dataset and error dataset. This contrast support further the idea of supervised learning can learn what we expected, like gender in this study.

To better understand the model, we employed Grad-CAM on the final convolutional layer to localize and highlight the discriminative regions. Some examples are as shown in Fig. 7. It is possible to see that the highlighted regions of male pandas are around their noses, while the highlighted regions of female pandas are somewhere else. This finding, while preliminary, is highly attended by panda experts and needs further exploration. We also visualized the features X_{avg} outputted by the backbone in a two-dimensional space by t-SNE algorithm. The results are present in Fig. 8, where the red and blue dots represent females and males respectively. From the Fig. 8, we can see that the features of gender distribution seem to be different, but some pandas whose learned gender characteristics are still confused.

5. Conclusion and Future Work

Gender ratio is a critical factor for population growth and decline. In this paper, we present the first study on panda gender classification based on face images. The deep convolutional network is adopted for learning features from the panda face image dataset, which contains 6,549 images from 221 pandas. The experimental results show that panda face contains some gender information and male and female pandas have different features on faces. These new findings should be further investigated by ecologists and wildlife researchers, and we will investigate a better algorithm for gender classification, including using other non-facial features.

Acknowledgment

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 Ministry of Human Resources and Social Security, and the Sichuan Science and Technology Program (2018JY0096).

References

[1] Report of the fourth national giant panda census. Technical report, State Forestry Administration, P.R.C, 2015.

[2] S. Alibhai, Z. Jewell, and J. Evans. The challenge of monitoring elusive large carnivores: An accurate and cost-effective tool to identify and sex pumas (puma concolor) from footprints. *PloS one*, 12(3):e0172065, 2017.

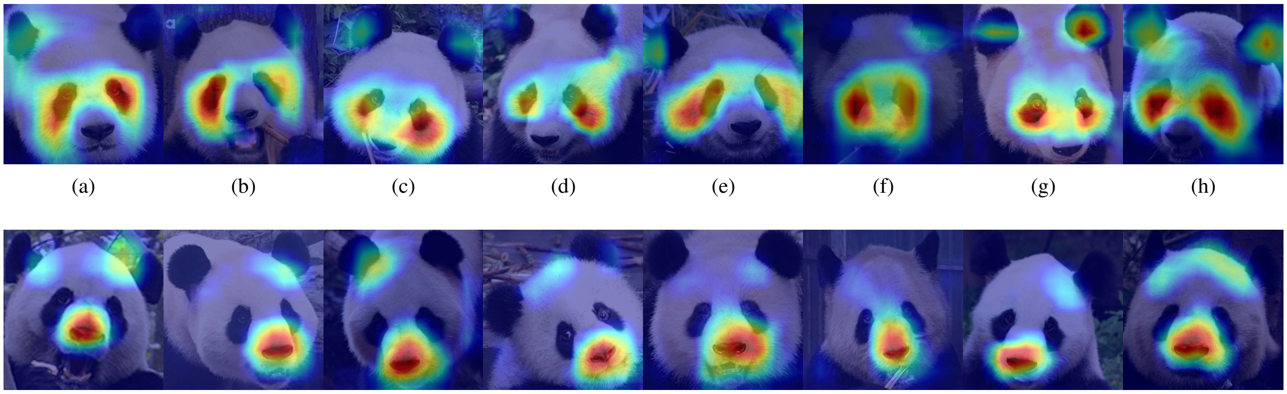
[3] M. Castrillón-Santana, J. Lorenzo-Navarro, and E. Ramón-Balmaseda. On using periocular biometric for gender classification in the wild. *Pattern Recognition Letters*, 82:181–189, 2016.

[4] A. Dehghan, E. G. Ortiz, G. Shu, and S. Z. Masood. Dager: Deep age, gender and emotion recognition using convolutional neural network. *arXiv preprint arXiv:1702.04280*, 2017.

[5] A. Freytag, E. Rodner, M. Simon, A. Loos, H. S. Kühl, and J. Denzler. Chimpanzee faces in the wild: Log-euclidean cnns for predicting identities and attributes of primates. In B. Rosenhahn and B. Andres, editors, *Pattern Recognition*, pages 51–63, Cham, 2016. Springer International Publishing.

[6] A. Gomez Villa, A. Salazar, and J. Vargas-Bonilla. Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks. *Ecological Informatics*, 41, 03 2016.

[7] J. Gu, S. K. Alibhai, Z. C. Jewell, G. Jiang, and J. Ma. Sex determination of amur tigers (panthera tigris altaica) from



(a) (b) (c) (d) (e) (f) (g) (h)

(i) (j) (k) (l) (m) (n) (o) (p)

Figure 7. Visualization results through class activated maps overlaid on input images together with the raw photographs. (Each subfigure is a different panda individual, the first row is female pandas and the second row is male pandas.)

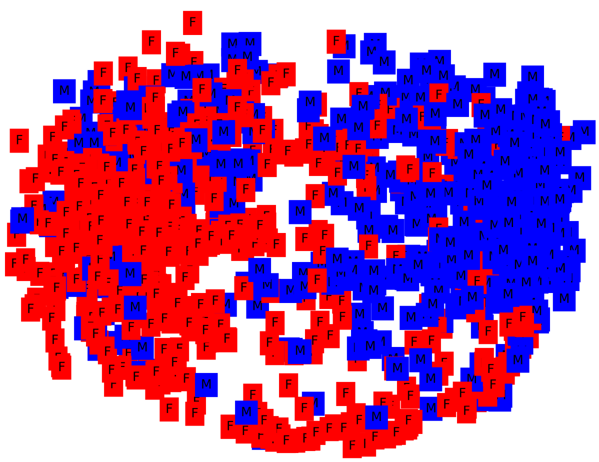


Figure 8. Visualization results through t-distributed stochastic neighbor embedding (red and blue dots represent female and male, respectively.)

footprints in snow. *Wildlife Society Bulletin*, 38(3):495–502, 2014.

[8] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, June 2016.

[9] K. He, X. Zhang, S. Ren, and J. Sun. Identity mappings in deep residual networks. In *Proceedings of the 14th European Conference on Computer Vision (ECCV)*, pages 630–645, Sept 2016.

[10] Q. He, Q. Zhao, N. Liu, P. Chen, Z. Zhang, and R. Hou. Distinguishing individual red pandas from their faces. *arXiv preprint arXiv:1908.03391*, 2019.

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

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems*, pages 1097–1105, 2012.

[13] S. Kumar and S. K. Singh. Visual animal biometrics: survey. *IET Biometrics*, 6(3):139–156, 2016.

[14] G. Levi and T. Hassncer. Age and gender classification using convolutional neural networks. In *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 34–42, June 2015.

[15] B. Li, S. Alibhai, Z. Jewell, D. Li, and H. Zhang. Using footprints to identify and sex giant pandas. *Biological Conservation*, 218:83–90, 02 2018.

[16] S. Li, J. Li, W. Lin, and H. Tang. Amur tiger re-identification in the wild. *arXiv preprint arXiv:1906.05586*, 2019.

[17] Y. Liu, A. Wai Kin Kong, and C. Keong Goh. A constrained deep neural network for ordinal regression. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 831–839, 2018.

[18] L. v. d. Maaten and G. Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(Nov):2579–2605, 2008.

[19] W. M. Matkowski, A. W.-K. Kong, H. Su, P. Chen, R. Hou, and Z. Zhang. Giant panda face recognition using small dataset. *ArXiv*, abs/1905.11163, 2019.

[20] T. P. Moorhouse and D. W. MacDonald. Indirect negative impacts of radio-collaring: sex ratio variation in water voles. *Journal of Applied Ecology*, 42(1):91–98, 2005.

[21] H. Nguyen, S. J. Maclagan, T. D. Nguyen, T. Nguyen, P. Flemons, K. Andrews, E. G. Ritchie, and D. Phung. Animal recognition and identification with deep convolutional neural networks for automated wildlife monitoring. In *2017 IEEE international conference on data science and advanced Analytics (DSAA)*, pages 40–49. IEEE, 2017.

[22] A. Polzounov, I. Terpugova, D. Skiparis, and A. Mihai. Right whale recognition using convolutional neural networks. *arXiv preprint arXiv:1604.05605*, 2016.

[23] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *2017 IEEE*

International Conference on Computer Vision (ICCV), pages 618–626, 2017.

[24] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[25] H. Su and A. W. K. Kong. A study on low resolution androgenic hair patterns for criminal and victim identification. *IEEE Transactions on Information Forensics and Security*, 9(4):666–680, 2014.

[26] H. Wang and H. Su. STAR: A concise deep learning framework for citywide human mobility prediction. In *2019 20th IEEE International Conference on Mobile Data Management (MDM)*, pages 304–309, June 2019.

[27] Z. Xiangjiang, T. Yi, L. Ming, Z. Zejun, B. Goossens, C. Youping, W. Hongjia, M. Bruford, and W. Fuwen. Accurate population size estimates are vital parameters for conserving the giant panda. *Ursus*, 20(1):56–63, 2009.

[28] X. Zhan, Z. Zhang, H. Wu, B. Goossens, M. Li, S. Jiang, M. W. Bruford, and F. Wei. Molecular analysis of dispersal in giant pandas. *Molecular Ecology*, 16(18):3792–3800, 2007.