



Article

Enhancing Endangered Feline Conservation in Asia via a Pose-Guided Deep Learning Framework for Individual Identification

Weiwei Xiao ¹, Wei Zhang ^{1,*} and Haiyan Liu ²¹ College of Science, North China University of Technology, Beijing 100144, China; xiaoww@ncut.edu.cn² Department of Statistics, University of Leeds, Leeds LS2 9JT, UK; h.liu1@leeds.ac.uk

* Correspondence: weizz99@mail.ncut.edu.cn

Abstract

The re-identification of endangered felines is critical for species conservation and biodiversity assessment. This paper proposes the Pose-Guided Network with the Adaptive L2 Regularization (PGNet-AL2) framework to overcome key challenges in wild feline re-identification, such as extensive pose variations, small sample sizes, and inconsistent image quality. This framework employs a dual-branch architecture for multi-level feature extraction and incorporates an adaptive L2 regularization mechanism to optimize parameter learning, effectively mitigating overfitting in small-sample scenarios. Applying the proposed method to the Amur Tiger Re-identification in the Wild (ATRW) dataset, we achieve a mean Average Precision (mAP) of 91.3% in single-camera settings, outperforming the baseline PPbM-b (Pose Part-based Model) by 18.5 percentage points. To further evaluate its generalization, we apply it to a more challenging task, snow leopard re-identification, using a dataset of 388 infrared videos obtained from the Wildlife Conservation Society (WCS). Despite the poor quality of infrared videos, our method achieves a mAP of 94.5%. The consistent high performance on both the ATRW and snow leopard datasets collectively demonstrates the method's strong generalization capability and practical utility.

Keywords: adaptive regularization; Amur tiger; snow leopard; re-identification; deep learning; wildlife conservation



Academic Editor: Michael Wink

Received: 21 November 2025

Revised: 11 December 2025

Accepted: 11 December 2025

Published: 12 December 2025

Citation: Xiao, W.; Zhang, W.; Liu, H. Enhancing Endangered Feline Conservation in Asia via a Pose-Guided Deep Learning Framework for Individual Identification. *Diversity* **2025**, *17*, 853. <https://doi.org/10.3390/d17120853>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Wildlife constitutes a vital component of Earth's biodiversity, playing an irreplaceable role in preserving ecological balance and supporting other essential functions [1]. However, with the intensification of human activities, global wildlife now faces unprecedented threats to its survival [2]. The International Union for Conservation of Nature (IUCN) Red List of Threatened Species indicates that over 48,600 species currently face extinction risks [3].

The Amur tiger, listed as endangered on the Red List, is one of the world's largest felines. It was once widely distributed across Northeast China, the Russian Far East, and the Korean Peninsula [4]. Affected by multiple threats, the Amur tiger's wild population once approached the brink of extinction. In recent years, through joint efforts by China and Russia, its population has shown signs of recovery and growth [5]. The snow leopard was classified as Vulnerable (VU) in the 2017 IUCN Red List assessment [6]. The same year, the International Snow Leopard Foundation launched the Population Assessment of the World's Snow Leopards (PAWS) initiative to systematically evaluate snow leopard

population dynamics using scientific methods [7]. Wildlife conservation is not merely about preserving individual species; it is crucial for maintaining global ecological balance and ensuring sustainable human development [8].

Individual identification plays a crucial role in conservation efforts for Amur tiger and snow leopard populations. Although various methods and tools are currently available for animal identification and activity recording [9,10], not all methods are suitable for all animal species. For Amur tigers and snow leopards, the commonly used individual identification methods primarily include DNA identification [11,12], footprint identification [13,14], scent identification [15], and fur pattern identification. With advancements in information technology and artificial intelligence, recognition based on the unique stripes of the Amur tiger and the distinctive spots of the snow leopard has gained increasing attention [16,17]. These stripe and spot patterns exhibit lifelong consistency, analogous to human fingerprints, and possess high individual specificity. Compared to other methods, surface pattern recognition offers advantages such as operational simplicity and reliable results, making it currently the most practical method for individual identification of Amur tigers and snow leopards.

The primary objective of wildlife re-identification is to automatically recognize and match the same individual across different images. The earliest research direction for re-identification tasks was pedestrian re-identification. In 2005, Zajdel first proposed the concept of “pedestrian re-identification”, after which numerous scholars are interested in this field [18]. Benefiting from mature developments in pedestrian re-identification, re-identification technology is gradually being applied to the field of animal re-identification [19]. The first application of computer vision in individual animal recognition dates back to 1990. In a pioneering study by Whitehead et al. [20], researchers used custom software to scan projected slides of sperm whale tail flukes onto digital plates. Then they manually annotated unique feature points, such as notches and scars, on the tail flukes to establish a feature database. For each query image, the software calculates similarity scores with all gallery images and returns the individual with the highest score as the matching result, thus achieving individual re-identification. Ravela and Gamble employed features such as Taylor approximations of local color intensity, multiscale histograms, and curvature to perform individual recognition on overhead images of spotted salamanders from Massachusetts [21]. With advances in machine learning, neural networks have increasingly been applied to animal re-identification. Carter et al. laid the groundwork for neural network applications in this field by extracting green sea turtle shell patterns, vectorizing the data as input, and training 50 distinct simple feedforward neural networks for turtle re-identification [22].

The aforementioned work holds significant importance for research in the field of animal re-identification. However, the process of establishing the dataset relies heavily on manual labor, and the data is overly standardized and uniform, resulting in limitations when applied in practice. In particular, re-identification studies of Amur tigers and snow leopards face additional challenges. As endangered and vulnerable species in the wild, they have small sample sizes, posing risks of overfitting. Furthermore, infrared photography often results in blurred images and underexposure, while limb movements cause significant pose variations. These factors make it difficult to establish large, uniform datasets, substantially increasing the complexity of re-identification research for these wild, endangered felines.

To address the challenge of re-identification for Amur tigers in wild environments, Li et al. constructed a large-scale Amur Tiger Re-identification (ATRW) dataset and proposed two benchmark methods for tiger re-identification: PPbM-a and PPbM-b, achieving a mAP accuracy of 77.1% under single-camera conditions [23]. This method employs

ResNet-50 as the backbone network and incorporates precise pose-local information, effectively addressing the significant pose variation issue in Amur tigers. Building upon this foundation, researchers explored various network architecture enhancement strategies to improve recognition performance. To better integrate global and local features, Liu et al. proposed the PPGNet architecture, comprising two major modules: a Global Stream and a Local Stream, achieving a single-shot accuracy of 90.6% [24]. Xu et al. further designed a global inverted pyramid multi-scale feature fusion module and a local dual-domain attention feature enhancement module. By leveraging multi-scale feature extraction and attention mechanisms to enhance key region features, they achieved a mAP accuracy of 78.7% under single-shot conditions [25]. With the successful application of transformer architectures in computer vision, researchers have been exploring their potential for wildlife re-identification. Bai et al. evaluated multiple deep learning architectures on the ATRW dataset, demonstrating that a combined ViT-MGN model achieved 83.4% mAP, proving the significant advantage of Transformer architectures in wildlife individual recognition tasks [26].

In 2023, researchers Bohnett et al. conducted an identification study using a curated dataset of free-ranging snow leopards photographed in Afghanistan between 2012 and 2019, along with data from captive individuals in zoos across Finland, Sweden, Germany, and the United States. By integrating convolutional neural networks, pose-invariant embedding, and the HotSpotter algorithm, their method achieved 85% Rank-1 accuracy and 99% Rank-20 accuracy [27]. In 2025, Solari et al. successfully developed an innovative genetic detection technique based on a multiplex PCR-SNP panel, enabling high-accuracy individual identification from snow leopard fecal samples. This method demonstrated exceptional performance in validation studies using paired zoo samples, achieving an allele call accuracy rate of 96.7%. The team further validated the technique with field-collected fecal samples from various regions of Pakistan, confirming that the SNP panel maintained reliable individual identification even for aged or low-quality wild samples. This breakthrough provides a powerful molecular biology tool for snow leopard conservation research [28].

Real-world applications impose stringent demands on both recognition accuracy and generalization capability of models. However, existing literature remains insufficient to fully meet these challenges. Current methods still suffer from limited recognition accuracy under complex pose variations and weak cross-species generalization capability, making them difficult to transfer directly to re-identification tasks involving other endangered feline species. To address these challenges, we hypothesize that integrating pose-guided semantic alignment with adaptive regularization can improve feature representation quality and model generalization for endangered feline re-identification under limited data conditions. Based on this hypothesis, this paper proposes PGNet-AL2 (Pose-Guided Network with Adaptive L2 Regularization), a novel framework designed specifically for this task.

The proposed framework achieves outstanding accuracy on the ATRW dataset and successfully transfers to the snow leopard re-identification domain, demonstrating its effectiveness for constructing re-identification systems for endangered wild felines.

2. Materials and Methods

2.1. Dataset

2.1.1. Wild Amur Tiger Dataset

ATRW [23] serves as the benchmark dataset for individual recognition of Amur tigers, characterized by its large sample size and rich annotation information. The data was collected under diverse natural environmental conditions. Figure 1 shows three representative scenarios: (a) the same individual in different poses captured by a single camera, (b) the

same individual with significant variation across cameras, and (c) different individuals with similar stripe patterns. The dataset contains 3649 bounding box annotations across 92 distinct Amur tigers. These are further divided into 182 recognition entities, where the left and right body sides of a tiger are treated as separate entities due to their independent stripe patterns. Given that tigers are non-rigid deformable objects, pose-invariant features are critical for recognition. Consequently, in addition to individual ID labels, the ATRW dataset provides spatial coordinate annotations for 15 key joints per tiger, as detailed in Figure 2 and Table 1. Unlike the Market1501 dataset [29] in the pedestrian re-identification field, not all entities in this dataset originate from multi-camera scenes. Approximately 70 entities were captured across camera views, while the remaining entities were sourced from different time-series frames of a single camera.

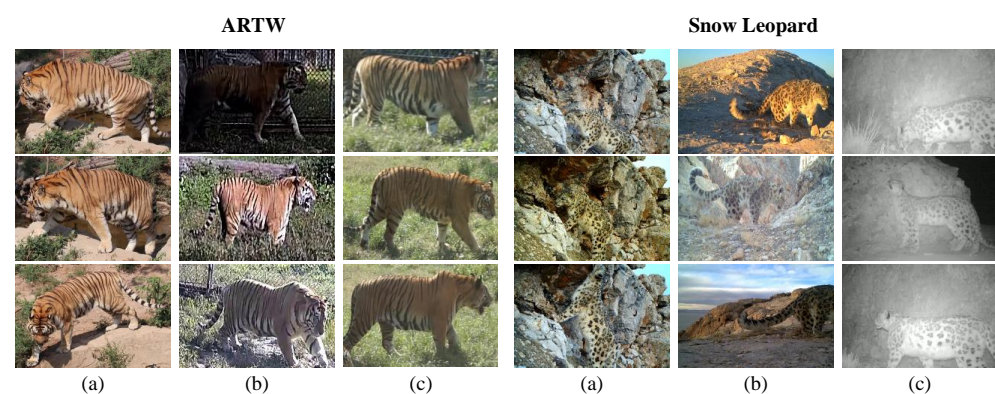


Figure 1. Challenges in wildlife re-identification. **Left:** Amur Tiger Re-identification (ATRW) dataset; **Right:** Snow Leopard dataset. (a) Same individual with different poses under single camera; (b) Same individual with significant variations across cameras; (c) Different individuals with similar stripe patterns.

This study adopts the standard data partition for the ATRW dataset, utilizing all 3649 images. The training set contains 1887 images from 107 entities, and the test set contains 1762 images from 75 entities. For evaluation, the entire test set is used as both the query and the gallery set.

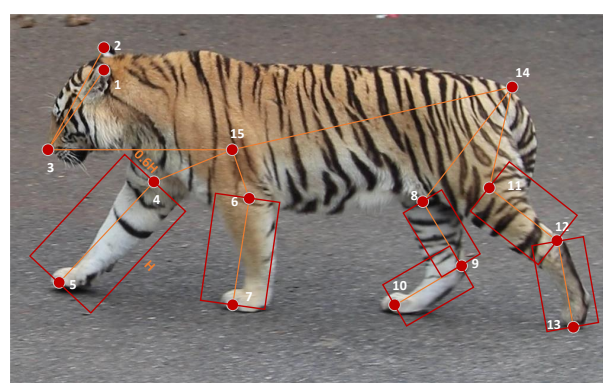


Figure 2. Key-point annotation examples from the ATRW dataset.

Table 1. Definition of the 15 key-points for pose annotation in the ATRW dataset.

Key-Point	Definition	Key-Point	Definition
1	left ear	9	right knee
2	right ear	10	right back paw
3	nose	11	left hip
4	right shoulder	12	left knee

Table 1. *Cont.*

Key-Point	Definition	Key-Point	Definition
5	right front paw	13	left back paw
6	left shoulder	14	root of tail
7	left front paw	15	center, mid point of 3 & 14
8	right hip		

2.1.2. Wild Snow Leopard Dataset

The wild snow leopard dataset, provided by the Wildlife Conservation Society (WCS), originates from 388 infrared camera videos featuring 30 individuals. From these videos, we extracted a total of 2013 images for re-identification, each with complete individual identity annotations. As shown in Figure 1, the infrared camera images in the dataset generally exhibit significant background interference. To address this, object detection bounding box annotations were applied to 2013 images in the dataset. Figure 3 displays some object detection results. The data partitioning scheme is as follows: the training set comprises 1680 images and 30 entities, while the test set includes 433 images and 15 entities. For evaluation, the entire test set is used as both the query and the gallery.

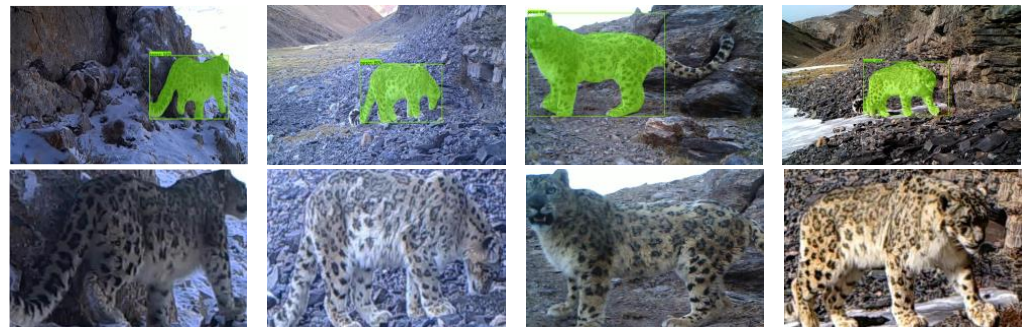


Figure 3. Examples of object detection results on the Snow Leopard dataset.

2.2. Method

Animal re-identification, as an image retrieval task, employs neural network methods to first extract discriminative global and local feature representations of individual animals. Subsequently, metric learning is utilized to optimize the embedding space, bringing the feature distances of identical individuals closer together while increasing the distances between different individuals. This ultimately enables cross-view individual matching and retrieval.

The dense and intricate stripes on the flanks of Amur tigers exhibit unique and stable characteristics, making them the primary identification region in tiger re-identification [17]. Traditional re-identification networks typically use photographs capturing the tiger's entire body as input, which often include the shooting background, head, and limb areas. However, this design method may lead to two issues: First, in field-collected images, the lateral stripes of different individual tigers often exhibit high similarity, while the background environments vary significantly [30]. This data distribution pattern, characterized by "similar foregrounds and diverse backgrounds", can cause models to learn spurious background-related features while neglecting the true discriminative details of the lateral stripes. This over-reliance on background features severely weakens the model's generalization ability, causing recognition performance to decline when backgrounds change. Second, tigers are mostly in motion within their natural habitats, resulting in significant differences in limb postures across images. This inconsistency in spatial relationships between body parts makes traditional global feature extraction methods ill-suited to adapt to pose variations,

thereby affecting re-identification accuracy. Furthermore, as endangered and vulnerable species, respectively, the wild populations of Amur tigers and snow leopards are scarce, resulting in extremely limited available image samples. This small-sample data often leads to overfitting during deep neural network training, restricting the model's generalization capability [31].

2.2.1. Pose-Guided and Adaptive Regularization-Based Re-Identification Network

To simultaneously address the challenges of background interference, pose variations, and data scarcity, this paper proposes a Pose-Guided Network with Adaptive L2 Regularization (PGNet-AL2) featuring a dual-branch architecture, inspired by the work of Liu et al. [24]. This network constructs an architecture where global and pose-guided branches collaborate synergistically. Both branches perform feature extraction based on ResNet, enabling progressive multi-level feature learning from coarse to fine. The overall network architecture is illustrated in Figure 4.

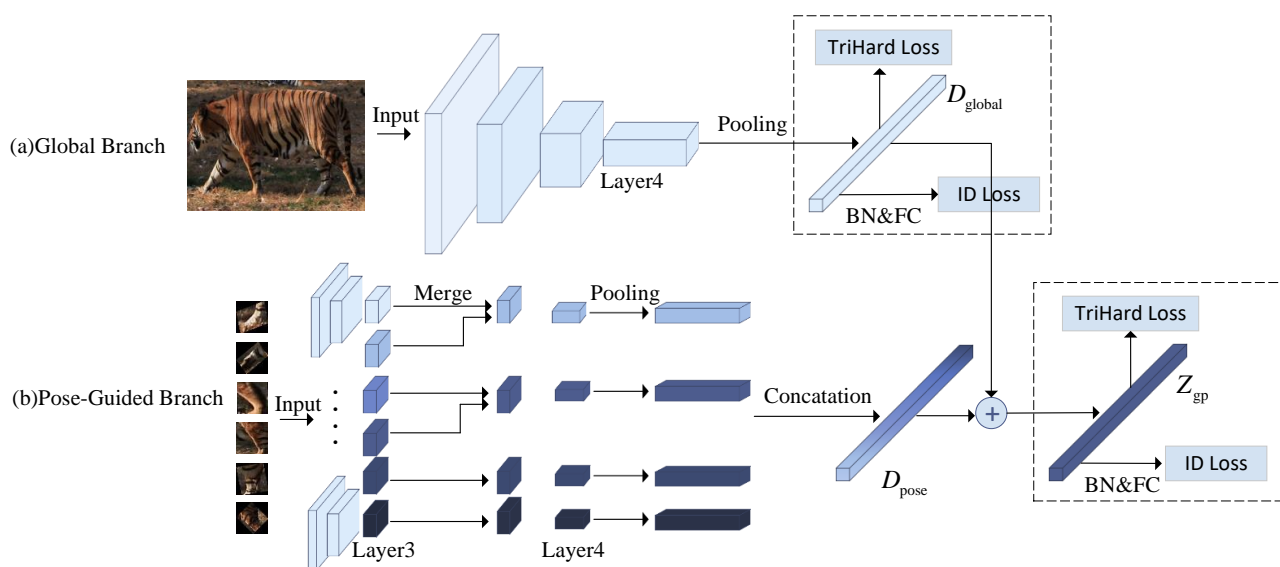


Figure 4. Structure of the proposed Pose-Guided Network with the Adaptive L2 Regularization (PGNet-AL2).

Global Branch: This branch takes the original image as input and extracts global features using ResNet-152. The global branch outputs a 2048-dimensional feature vector D_{global} . As mentioned earlier, the original image contains excessive background information that may cause the model to develop a certain dependence on the background. However, in practical applications, some contextual information (such as habitat type and vegetation characteristics) contains valuable discriminative clues that can help improve recognition accuracy. Therefore, this paper does not eliminate all background information but retains a small amount of scene information.

Posture-Guided Branch: This branch employs pose keypoint information to achieve fine-grained feature extraction at the body part level. Through a semantic alignment mechanism, it maps corresponding body parts across different individuals into a unified feature space, guiding the network to focus on the lateral stripes of Amur tigers while suppressing background interference. Specifically, based on the detected 10 limb keypoints, we crop 6 local regions: 4 regions from the 2 hind limbs and 2 regions from the 2 forelimbs. Each region undergoes feature extraction via an independent ResNet-34. To reduce number of parameters and enhance feature representation, after feature extraction up to layer 3, we fuse semantically related feature maps: combining features from regions 1 and 2, and

regions 3 and 4, yielding two sets of fused features. Subsequently, these two sets of fused features and the features from regions 5 and 6 are processed through the final convolutional layer of four separate networks, yielding four 512-dimensional vectors. These are finally concatenated to form the 2048-dimensional feature vector D_{pose} .

Accordingly, the PGNet-AL2 network consists of three core modules: the global branch, the pose-guided branch, and the adaptive L2 regularization. For the ATRW dataset, we employ the complete PGNet-AL2 network. Given the relatively small size of the snow leopard dataset (30 individuals), adaptive L2 regularization was incorporated into the re-identification network design for snow leopards, ultimately yielding a 2048-dimensional feature vector D_{global} .

2.2.2. Loss Function

To fully leverage the complementarity of multi-branch features and prevent the network from degenerating into reliance on a single branch, this paper proposes a joint supervision strategy. The global branch feature D_{global} is fused with the local branch feature D_{pose} to obtain the fused feature Z_{gp} . This yields two levels of feature representations: the global feature D_{global} and the fused feature Z_{gp} . For these two feature representations, the following supervision loss function is designed:

ID Loss: Assign an independent classifier to supervise each feature representation. Specifically, for features D_{global} and Z_{gp} , they are each mapped to k -dimensional logits vectors through a fully connected layer:

$$z = Wf + b \quad (1)$$

where $W \in R^{k \times d}$ is the weight matrix, $b \in R^k$ is the bias vector, f represents the feature vector (D_{global} or Z_{gp}), and k is the number of classes.

Subsequently, the Softmax function is applied to transform the logits vector into a probability distribution: $\hat{y} = \text{Softmax}(z) = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_k)$. Finally, the cross-entropy loss is computed between the predicted probability distribution \hat{y} and the true labels $y = [y_1, y_2, \dots, y_k]$:

$$L_{CE}(w) = - \sum_{i=1}^k y_i \log(\hat{y}_i) = - \sum_{i=1}^k y_i \log \left(\frac{e^{z_i}}{\sum_{j=1}^k e^{z_j}} \right) \quad (2)$$

where w contains all learnable parameters of the network.

Measuring learning loss: Apply a hard triplet loss to D_{global} and Z_{gp} respectively:

$$L_{tri}(w) = \max(0, d_p(w) - d_n(w) + \text{margin}) \quad (3)$$

where d_p is the Euclidean distance between the anchor sample and the most difficult positive sample (i.e., the same-identity sample that is farthest from the anchor) in the batch, d_p is the Euclidean distance between the anchor sample and the closest negative sample (i.e., the different-identity sample that is nearest to the anchor) in the batch.

The margin is a predefined threshold that defines the minimum required separation between d_p and d_n (Figure 5). Without this margin constraint, the model may trivially satisfy the loss by compressing all embeddings into a small region where d_n is only marginally larger than d_p , failing to learn meaningful feature representations. By enforcing a sufficient margin, the model is compelled to learn discriminative embeddings with adequate inter-class separability in the feature space.

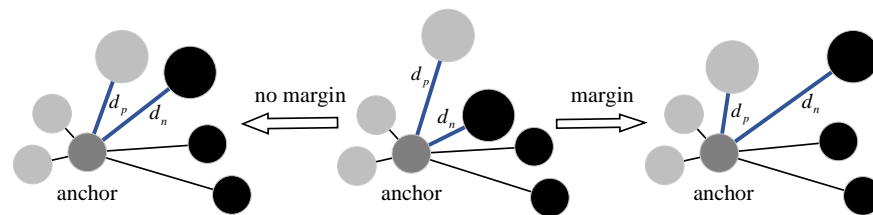


Figure 5. Illustration of batch hard triplet loss with and without margin. White and black circles denote positive and negative samples, respectively. Larger circles highlight the most difficult positive sample (white) and the closest negative sample (black).

The overall objective function comprises four loss terms to ensure that each branch learns collaboratively rather than relying solely on a single factor, formulated as follows:

$$L_{\text{total}}(w) = L_{\text{CE_D}_{\text{global}}} + L_{\text{CE_Z}_{\text{gp}}} + 2L_{\text{tri_D}_{\text{global}}} + 2L_{\text{tri_Z}_{\text{gp}}} \quad (4)$$

Given that individual re-identification is fundamentally a metric learning task, we assign a higher weight to the triplet loss to emphasize learning discriminative embeddings in the feature space. The cross-entropy loss provides auxiliary supervision for feature learning through classification.

2.2.3. Introduction of Adaptive Regularization

During network training, L2 regularization effectively enhances model generalization by constraining neural network parameters and introducing a penalty term to the loss function. The scarcity of available samples for Amur tigers and snow leopards makes it challenging to construct large-scale datasets comparable to those used in pedestrian re-identification. Therefore, employing regularization techniques to prevent overfitting is crucial when training neural networks on such small-sample data.

L2 regularization achieves parameter constraints by introducing a penalty term into the objective function:

$$L_{\text{regularized}}(w) = L_{\text{total}}(w) + \lambda \sum_{i=1}^n \|w_i\|_2^2 \quad (5)$$

Here, n denotes the number of distinct parameters in the neural network, and $L_{\text{regularized}}(w)$ represents the complete objective function with the addition of an L2 regularization term. The set $w = \{w_i | i = 1, \dots, n\}$ encompasses all learnable parameters of the model, where each element w_i takes the form of a scalar, vector, matrix, or 4D tensor. For instance, the weights of a convolutional layer constitute a 4D tensor, while the bias represents a 1D vector, each corresponding to distinct w_i within the set w . $\|w_i\|_2^2$ denotes the square of the L2 norm of parameter w_i , with parameter $\lambda \in R_+$ controlling the regularization strength.

Standard L2 regularization imposes a uniform constraint on all parameters in the network, with the parameter λ remaining constant throughout training. The optimal value of the hyperparameter λ relies on manual tuning for determination. In practical applications, different layers of a network often require varying degrees of regularization intensity. For instance, weaker regularization may be applied to shallow layers, while stronger regularization is applied to deep layers. To achieve this, the hyperparameter λ can be generalized into dedicated coefficients for each weight w_i in the network:

$$L_{\text{regularized}}(w) = L_{\text{total}}(w) + \sum_{i=1}^n (\lambda_i \|w_i\|_2^2) \quad (6)$$

However, ResNet152 comprises 152 layers with multiple parameter types in each layer (such as convolutional kernels, batch normalization parameters, and fully connected layer weights). Since each parameter type in each layer requires an independent regularization factor λ_i , manual adjustment of hundreds of λ_i values is clearly infeasible. Therefore, this paper adopts the adaptive L2 regularization proposed in [32], treating all λ_i as learnable parameters that adapt to w_i :

$$L_{\text{regularized}}(w) = L_{\text{total}}(w) + \sum_{i=1}^n (A f(\theta_i) \|w_i\|_2^2) \quad (7)$$

The hyperparameter $A \in R_+$ provides global control over the regularization strength, preventing excessively large individual coefficients from disrupting the training process. $\theta_i \in R_+ \{i = 1, 2, \dots, n\}$ represents trainable scalar variables. The hard sigmoid activation function $f(\theta_i)$ ensures that regularization coefficients remain non-negative, expressed as:

$$f(\theta_i) = \begin{cases} 0, & \theta_i \leq c \\ 1, & \theta_i > c \\ \theta_i / (2c) + 0.5, & \text{otherwise} \end{cases} \quad (8)$$

2.2.4. Optimization Methods

Data Augmentation: (1) This paper performs horizontal flipping on training set photos to generate new instances, doubling the dataset size to twice the original quantity; (2) An augmentation strategy combining geometric transformations with color space transformations is applied. Random rotation enhances the model's robustness to pose variations, while color dithering simulates diverse lighting conditions and camera settings by randomly adjusting brightness, contrast, saturation, and hue. These augmentation techniques significantly enhance training data diversity, effectively preventing model overfitting.

Label smoothing: The true label $y = [y_1, y_2, \dots, y_k]$ is a one-hot vector. During the process of minimizing the loss function LCE, it guides the network to converge the probability toward 1 for correctly predicted categories and toward 0 for incorrectly predicted categories. This ultimately leads the network to produce extreme logit values, where the logit for the correct category methods positive infinity and the logit for the incorrect category methods negative infinity. Extreme logit values cause the model to become overconfident and lack generalization ability. Therefore, this paper introduces a label smoothing operation with hyperparameter $\alpha \in R_+$, where the true label is transformed as follows:

$$y_i = (1 - \beta)y_i + \frac{\alpha}{k}. \quad (9)$$

Dropout: Deep neural networks contain a large number of trainable parameters, while the training samples for wildlife re-identification tasks are relatively limited. This mismatch between parameter scale and sample size can easily lead to model overfitting on the training data. To mitigate this issue, this paper introduces Dropout regularization. During the forward propagation process, Dropout randomly sets the activation outputs of neurons to zero with probability p , preventing the network from becoming overly dependent on specific combinations of neurons.

The learning rate scheduling employs a two-stage “warm-up and decay” strategy: First, during the warm-up phase, the learning rate linearly increases from 0.00025 to 0.0025 over the initial 25 training epochs. This method stabilizes gradient updates during early training, preventing parameter oscillations caused by an excessively high initial learning rate. Subsequently, the decay phase commences. After the warm-up concludes, the learning rate is multiplied by a decay factor of 0.5 every 80 training epochs. By periodically reducing

the learning rate, the model progressively refines its parameter configuration, accelerating training convergence.

3. Results

This experiment was implemented using the PyTorch 1.13.1 deep learning framework on a workstation equipped with two NVIDIA GeForce RTX 3090 GPUs (24 GB memory each). Model training was conducted over 210 epochs with a batch size of 64, utilizing DataParallel for distributed training across both GPUs. The average training time per epoch was 127.262 s, achieving a processing speed of 27.2 samples per second. For adaptive L2 regularization, the hard sigmoid function parameters were set to $c = 1.0$ and $A = 0.005$.

3.1. Evaluation Metrics

This paper employs two standard re-identification evaluation metrics:

1. Cumulative Matching Characteristics (CMC). Suppose the query set contains N samples, and the gallery set contains M samples. For a given query sample q , there are m ground truth matches (i.e., samples with the same ID) in the gallery. The retrieval results are ranked in descending order of similarity as: $\{g_1, g_2, \dots, g_M\}$. For each query q , the Rank- k indicator is defined as:

$$\text{CMC}_q(k) = \begin{cases} 1, & \text{if a correct match appears within top-}k \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

The Rank- k accuracy is computed by averaging over all queries:

$$\text{CMC}(k) = \frac{1}{N} \sum_{q=1}^N \text{CMC}_q(k) \quad (11)$$

We report Rank-1, Rank-5, and Rank-10 accuracies as standard evaluation metrics.

2. Mean Average Precision (mAP). We first define a binary indicator function $\delta(k)$:

$$\delta(k) = \begin{cases} 1, & \text{if the item at rank } k \text{ is a correct match} \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

The precision at rank k is defined as:

$$P(k) = \frac{1}{k} \sum_{i=1}^k \delta(i) \quad (13)$$

The Average Precision (AP) for a single query is computed as:

$$\text{AP}_q = \frac{1}{m} \sum_{k=1}^M P(k) \cdot \delta(k) \quad (14)$$

Finally, the mean Average Precision (mAP) is obtained by averaging AP over all N queries:

$$\text{mAP} = \frac{1}{N} \sum_{q=1}^N \text{AP}_q \quad (15)$$

CMC measures whether any correct match appears within the top- k results, treating the query as successful if at least one same-identity sample is ranked in the top- k posi-

tions. contrast, mAP is more stringent, as it evaluates the overall quality by considering the ranking positions of all correct matches in the retrieval list.

3.2. Comparison with Existing Methods

Table 2 shows the performance comparison results of applying the proposed PGNet-AL2 model against other existing methods into the ATRW dataset. It can be seen that the proposed PGNet-AL2 achieves optimal performance in single-camera scenarios, with a mAP of 91.4%, surpassing PPGNet by 0.7 percentage points and PPbM-b by 18.5 percentage points. Rank-1 and Rank-5 accuracy rates reach 98.9% and 99.7%, respectively, demonstrating the model's strongest recognition capability under a single viewpoint. In cross-scene scenarios, PGNet-AL2 achieves an mAP of 71.3%, with Rank-1 and Rank-5 accuracy rates of 95.4% and 97.7%, respectively. Its performance in this setting also stands out among numerous methods, demonstrating enhanced robustness to variations in viewpoint, lighting, and background. Figure 6 illustrates four randomly selected query samples and their retrieval results.

Table 2. Performance comparison with state-of-the-art methods on the ATRW dataset (%). Bold values indicate the best performance.

Methods	Single-Cam			Cross-Cam		
	mAP	Top-1	Top-5	mAP	Top-1	Top-5
CE [23]	59.1	78.6	92.7	38.1	69.7	87.8
Aligned-reID [23]	64.8	81.2	92.4	44.2	73.8	90.5
PPbM-a [23]	74.1	88.2	96.4	51.7	76.8	91.0
PPbM-b [23]	72.8	89.4	95.6	47.8	77.1	90.7
ResNet50+IFPM+LAEM [25]	78.7	96.3	98.9	—	—	—
ResNet50+ViT+MGN [26]	83.4	92.3	94.9	43.6	79.4	85.7
PPGNet(re-rank) [24]	90.6	97.7	99.1	72.6	93.6	96.7
PGNet-AL2(ours)	91.3	98.9	99.7	71.3	95.4	97.7

For snow leopard recognition, mAP reached 94.5%, with Rank-1 and Rank-5 accuracy at 98.6% and 98.9%, respectively. This demonstrates that the proposed PGNet-AL2 model achieves strong recognition performance even on datasets with poor image quality (e.g., motion blur, exposure issues, low resolution, severe occlusions), proving its excellent transferability.

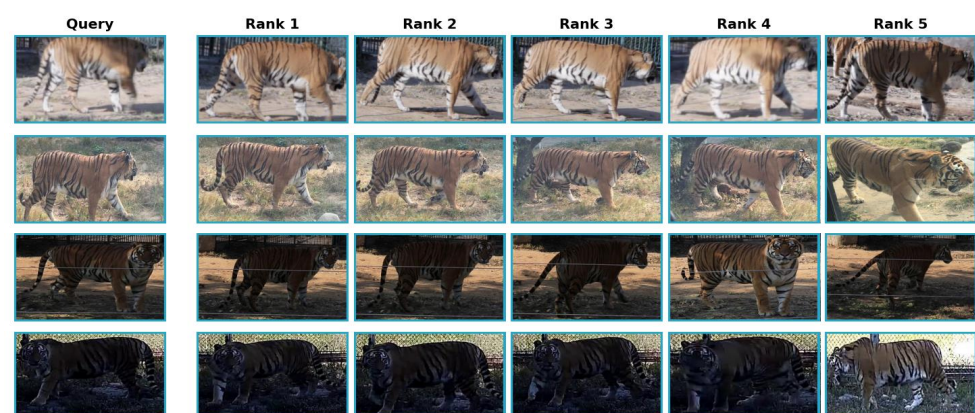


Figure 6. Visualization of retrieval results for four randomly selected query samples.

Notably, our method demonstrates advantages in both computational efficiency and theoretical contribution. The streamlined dual-branch architecture allows inference using only the global branch, reducing computational cost. Beyond empirical gains, this work

systematically introduces adaptive regularization to wildlife re-identification, providing a principled approach to mitigating overfitting under limited data—a critical challenge for endangered species. While the mAP improvement appears modest (0.7 percentage points), it represents a 7.4% relative error reduction (from 9.4% to 8.7%). In endangered species monitoring where each correct identification impacts conservation decisions, this improvement offers significant practical value.

The exceptional performance of the proposed re-identification method stems from the synergistic effects of the following factors: (1) Dual-branch fusion enhances feature discriminability. The global branch captures overall appearance features, providing global semantic information and complete contour characteristics to model macroscopic features such as body shape and overall stripe patterns. In the pose-guided branch, we fully leverage the pose point information provided by the ARTW dataset to achieve local region alignment, extracting pose-invariant, fine-grained discriminative features that enhance learning of the animal’s intrinsic texture. This multi-level feature representation and fusion enables the network to simultaneously learn discriminative information from macro to micro scales. (2) Due to the scarcity of samples in wildlife datasets, networks are prone to memorizing specific patterns in training data. The adaptive L2 regularization proposed in this paper effectively addresses model overfitting and enhances generalization capabilities.

3.3. Ablation Experiment

To validate the effectiveness of each module, this paper designed systematic ablation experiments. Using the global branch with standard L2 regularization as the baseline model, we progressively added components and analyzed their contributions. Tables 3 and 4 present detailed ablation results on the ATRE and Snow Leopard datasets, respectively. To further analyze the model’s decision-making basis, this paper introduces Gradient-weighted Class Activation Mapping (Grad-CAM) technology [33] to generate attention heatmaps, visualizing the key regions the model focuses on during the feature extraction stage. Additionally, histograms of adaptive regularization factor distributions across parameters post-training were presented, confirming the adaptive mechanism successfully implemented differential constraints on varying parameters.

Table 3. Ablation experiments with the global branch as baseline on the ARTW dataset (%). Bold values indicate the best performance.

Methods	Single-Cam			Cross-Cam		
	mAP	Top-1	Top-5	mAP	Top-1	Top-5
Baseline	88.9	95.1	97.7	69.7	92.0	97.1
+Pose Guided	89.9	97.7	99.1	69.5	88.0	96.0
+Pose Guided+AL2	91.3	98.8	99.7	71.3	95.4	97.7

Table 4. Ablation experiments with the global branch as baseline on the Snow Leopard dataset (%). Bold values indicate the best performance.

Methods	Single-Cam			
	mAP	Top-1	Top-5	Top-10
Baseline	92.7	97.4	98.4	98.6
+AL2	94.5	98.6	98.9	99.5

Figure 7a, 7b, and 7c show the attention heatmaps generated by the Baseline, Baseline+Pose Guided, and Baseline+Pose Guided+AL2 network architectures, respectively. As observed in Figure 7a, the baseline model primarily focuses on background areas rather

than the target individuals themselves. As mentioned earlier, due to the significant variability in natural shooting environments, the model tends to rely on background cues for recognition. However, the key feature for individual Amur tiger recognition should be the stripe pattern on its flanks. This excessive reliance on background cues results in weak generalization capabilities. As demonstrated in Figure 1c, distinguishing different individuals becomes extremely challenging when they appear in similar backgrounds.

Figure 7b demonstrates that after incorporating pose information, the model successfully redirects attention toward stripe features, significantly reducing reliance on background information. However, the model's focus remains overly concentrated within small local regions, failing to fully utilize stripe information across the entire body side. This limitation constrains its generalization capabilities.

Figure 7c demonstrates that introducing adaptive L2 regularization (AL2) significantly expands the model's attention scope, enabling it to capture lateral stripe features more comprehensively. This improvement effectively enhances the model's generalization capability, as evidenced by the corresponding increase in mAP scores.

The ablation experiments validated the effectiveness of each module from both quantitative and qualitative perspectives. Quantitatively, the incremental addition of components consistently improved performance. Qualitatively, heatmaps visually demonstrated the evolution of the model's focus areas across different module combinations. Specifically, the introduction of local branches successfully redirected the model's attention from background areas to the lateral stripes. Building upon this, the addition of adaptive L2 regularization further expanded the model's focus on stripe features, enabling it to utilize lateral stripe information more comprehensively. This mechanism of shifting feature attention from local to global effectively enhanced the model's generalization capability, ultimately reflected in a significant improvement in recognition accuracy.

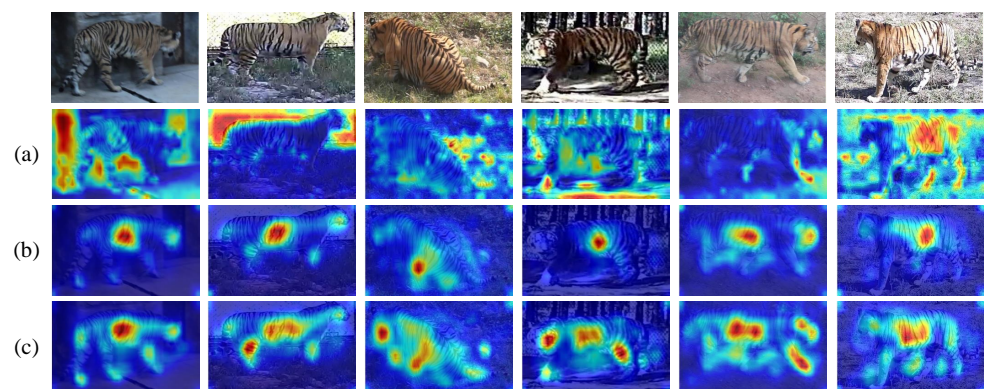


Figure 7. Gradient-weighted Class Activation Mapping (Grad-CAM) attention heatmaps for ablation study. (a) Baseline: attention concentrated on background. (b) +Pose Guided: attention shifted to stripe patterns but overly localized. (c) +Pose Guided+AL2: expanded attention coverage capturing comprehensive body stripe features.

4. Discussion

To objectively evaluate the performance of the proposed PGNet-AL2 network, this section conducts a systematic comparison with representative methods employing different technical approaches. Experimental results demonstrate that the proposed method achieves 91.3% mAP on the ATRW dataset and 94.5% mAP on our custom snow leopard dataset, exhibiting notable advantages over mainstream works in recent Amur tiger and snow leopard re-identification research.

4.1. Comparison with Traditional CNN Methods

In the ATRW benchmark established by Li et al. [23], their PPbM method obtained 77.1% mAP. Liu et al. [24] proposed the three-branch PPGNet, improving performance to 90.6%. Building upon their pose-guided concept, this study implements critical improvements: streamlining to a two-branch architecture and introducing an adaptive L2 regularization mechanism. Unlike PPGNet's fixed regularization parameters, PPGNet-AL2 dynamically adjusts regularization strength according to training progression, enabling the model to better adapt to pose variations in wild environments. The 0.7 percentage point improvement over PPGNet (91.3% vs. 90.6%) suggests that in wildlife scenarios with limited samples, the adaptive mechanism can contribute to improved generalization compared to traditional fixed architectures.

4.2. Comparison with Transformer-Based Methods

Bai et al. [26] applied Transformers to wildlife re-identification, with their ViT-MGN model achieving 83.4% mAP on ATRW. The proposed CNN-based method leads by 7.9 percentage points (91.3% vs. 83.4%), a gap that can be explained from two perspectives:

The first is the data efficiency issue. The moderate scale of the ATRW dataset makes it difficult to fully exploit the representational capacity advantages of Transformers. In contrast, the pose-guided CNN architecture provides inductive biases more suitable for small-sample scenarios through explicit modeling. The second is the difference in modeling approaches. The proposed method explicitly incorporates pose keypoints to achieve semantic-level alignment, while Transformers rely on self-attention mechanisms to implicitly learn spatial relationships. Under the pose diversity and occlusion scenarios characteristic of wildlife data, the explicit modeling strategy demonstrates stronger robustness.

4.3. Effectiveness Analysis of Adaptive Regularization

The aforementioned comparative methods share a common limitation: they adopt fixed regularization schemes that cannot dynamically adapt to the inherent heterogeneity of wildlife data. The adaptive L2 regularization mechanism proposed in this study fills this gap. This mechanism assigns differentiated constraint strengths to individual network weights based on training status, achieving fine-grained control at the parameter level.

Figure 8 presents a histogram of regularization factor distribution at the end of training (bin width: 0.0001). Taking convolutional layer parameters as an example, their regularization factors exhibit a notably dispersed distribution within the range of 0.0004–0.0016: the peak occurs at 0.0012 (33 parameters), while the intervals 0.001 and 0.0011 contain 20 and 27 parameters, respectively. This distribution pattern indicates that even for structurally identical parameters, the adaptive mechanism can independently optimize regularization strength according to their distinct roles in feature learning. This phenomenon provides empirical support for the necessity and effectiveness of dynamically adjusting regularization strength across different network layers and parameter types.

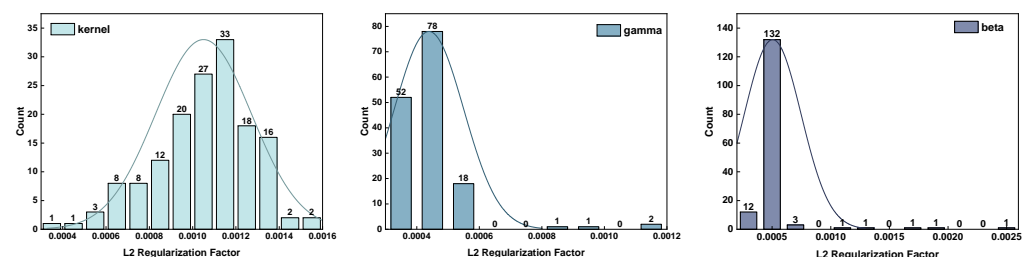


Figure 8. Distribution histogram of adaptive L2 regularization factors at the end of training (bin width = 0.0001). The dispersed distribution demonstrates that the adaptive mechanism successfully assigns differentiated regularization strengths to individual parameters based on their characteristics.

In summary, the performance improvements achieved by PGNet-AL2 stem from the synergistic effect of two aspects: first, the inherited and refined pose-guided two-branch architecture provides effective structural priors; second, the adaptive regularization strategy specifically designed for wildlife data characteristics enhances the model's generalization capability. This combined strategy offers an effective solution for wildlife re-identification tasks characterized by limited samples and high pose variation.

5. Conclusions

This study has proposed PGNet-AL2, a pose-guided network with adaptive L2 regularization, to address key challenges in endangered feline re-identification under limited data conditions, including pose variations, background interference, and sparse samples. The research hypothesis—that integrating pose-guided semantic alignment with adaptive regularization can improve feature representation quality and model generalization—has been validated through comprehensive experiments on both ATRW and snow leopard datasets.

A parallel network architecture comprising a global branch and a pose-guided branch was designed. The global branch captures overall appearance features through full-image convolutions; the pose-guided branch performs semantic alignment of body parts based on key point information, eliminating background interference to focus on the animal's intrinsic texture and extract local features. Feature fusion from both branches generates a representation that integrates both global and local details. To address the issue of weak network generalization due to sparse samples in wild endangered feline datasets, an adaptive L2 regularization strategy was proposed. This mechanism dynamically adjusts the regularization factor based on training progress, assigning appropriate constraint strengths to each network weight. To further improve robustness, we employ data augmentation strategies including horizontal flipping, random rotation, and color dithering, combined with label smoothing and dropout regularization techniques.

Experimental results on the ATRW dataset demonstrate that this method effectively mitigates network overfitting: mAP reaches 91.3%, surpassing the existing state-of-the-art by 0.7 percentage points. The successful transfer to the snow leopard re-identification domain, achieving 94.5% mAP, validates the effectiveness of the proposed approach for constructing re-identification systems for endangered wild felines. Ablation studies confirm the contribution of each component. Heatmap visualizations clearly illustrate how the model's attention shifts across different patches: the local branch guides focus to stripe regions on the body side, while the adaptive L2 regularization expands the stripe attention area, thereby further enhancing the model's discrimination capability and recognition accuracy.

Furthermore, using only the global branch during inference significantly speeds up processing, making the method more suitable for practical applications. This study provides an effective technical solution for wildlife re-identification, offering significant practical value for monitoring and conserving Amur tiger and snow leopard populations.

Author Contributions: Conceptualization, W.X. and W.Z.; methodology, W.X., W.Z. and H.L.; software, W.X. and W.Z.; validation, W.X., W.Z. and H.L.; formal analysis, W.X. and W.Z.; investigation, W.X. and W.Z.; data curation, W.X. and W.Z.; writing—original draft preparation, W.X. and W.Z.; writing—review and editing, W.X., W.Z. and H.L.; visualization, W.X. and W.Z.; supervision, W.X. and H.L.; project administration, W.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The ATRW dataset is publicly available at DOI: 10.1145/3394171.3413569. The snow leopard dataset is available from the corresponding author upon reasonable request due to third-party ownership and conservation security concerns. The source code will be made

publicly available at <https://github.com/weizz99/PGNet-AL2> (accessed on 20 November 2025) upon publication.

Acknowledgments: We are grateful to the creators of the ATRW dataset for making their data publicly available. We thank the researcher who provided the snow leopard dataset for this study. We also express our gratitude to the anonymous reviewers for their valuable comments and suggestions that significantly improved this manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Gupta, S.; Kumaresan, P.R.; Saxena, A.; Mishra, M.R.; Upadhyay, L.; Arul Sabareeswaran, T.A.; Alimudeen, S.; Magrey, A.H. Wildlife Conservation and Management: Challenges and Strategies. *Uttar Pradesh J. Zool.* **2023**, *44*, 280–286. [CrossRef]
2. Hiby, L.; Lovell, P.; Patil, N.; Kumar, N.S.; Gopalaswamy, A.M.; Karanth, K.U. A Tiger Cannot Change Its Stripes: Using a Three-Dimensional Model to Match Images of Living Tigers and Tiger Skins. *Biol. Lett.* **2009**, *5*, 383–386. [CrossRef] [PubMed]
3. IUCN Red List of Threatened Species. Available online: <https://www.iucnredlist.org/> (accessed on 23 February 2025).
4. Jeong, D.; Hyun, J.Y.; Marchenkova, T.; Matiukhina, D.; Cho, S.; Lee, J.; Kim, D.Y.; Li, Y.; Darman, Y.; Min, M.S.; et al. Genetic Insights and Conservation Strategies for Amur Tigers in Southwest Primorye Russia. *Sci. Rep.* **2024**, *14*, 29985. [CrossRef] [PubMed]
5. Wang, T.; Andrew Royle, J.; Smith, J.L.; Zou, L.; Lü, X.; Li, T.; Yang, H.; Li, Z.; Feng, R.; Bian, Y.; et al. Living on the Edge: Opportunities for Amur Tiger Recovery in China. *Biol. Conserv.* **2018**, *217*, 269–279. [CrossRef]
6. International Union for Conservation of Nature. *IUCN Annual Report 2017*; International Union for Conservation of Nature (IUCN): Gland, Switzerland, 2017.
7. Sharma, K.; Alexander, J.S.; Durbach, I.; Kodi, A.R.; Mishra, C.; Nichols, J.; MacKenzie, D.; Ale, S.; Lovari, S.; Modaqiq, A.W.; et al. PAWS: Population Assessment of the World’s Snow Leopards. In *Snow Leopards*; Elsevier: Amsterdam, The Netherlands, 2024; pp. 437–447.
8. Singh, M.; Vallarasu, K. Environmental Conservation and Sustainability: Strategies for a Greener Future. *Int. J. Multidimens. Res. Perspect.* **2023**, *1*, 185–200. [CrossRef]
9. Andreychev, A.V. A New Methodology for Studying the Activity of Underground Mammals. *Biol. Bull.* **2018**, *45*, 937–943. [CrossRef]
10. Tang, X.; Tang, S.; Li, X.; Menghe, D.; Bao, W.; Xiang, C.; Gao, F.; Bao, W. A Study of Population Size and Activity Patterns and Their Relationship to the Prey Species of the Eurasian Lynx Using a Camera Trapping Approach. *Animals* **2019**, *9*, 864. [CrossRef] [PubMed]
11. Caragiulo, A.; Pickles, R.S.A.; Smith, J.A.; Smith, O.; Goodrich, J.; Amato, G. Tiger (*Panthera tigris*) Scent DNA: A Valuable Conservation Tool for Individual Identification and Population Monitoring. *Conserv. Genet. Resour.* **2015**, *7*, 681–683. [CrossRef]
12. Dou, H.; Yang, H.; Feng, L.; Mou, P.; Wang, T.; Ge, J. Estimating the Population Size and Genetic Diversity of Amur Tigers in Northeast China. *PLoS ONE* **2016**, *11*, e0154254. [CrossRef] [PubMed]
13. Alibhai, S.K.; Gu, J.; Jewell, Z.C.; Morgan, J.; Liu, D.; Jiang, G. ‘I Know the Tiger by His Paw’: A Non-Invasive Footprint Identification Technique for Monitoring Individual Amur Tigers (*Panthera tigris altaica*) in Snow. *Ecol. Inform.* **2023**, *73*, 101947. [CrossRef]
14. Sharma, S.; Jhala, Y.; Sawarkar, V.B. Identification of Individual Tigers (*Panthera Tigris*) from Their Pugmarks. *J. Zool.* **2005**, *267*, 9–18. [CrossRef]
15. Kerley, L.L.; Salkina, G.P. Using Scent-Matching Dogs to Identify Individual Amur Tigers from Scats. *J. Wildl. Manag.* **2007**, *71*, 1349–1356. [CrossRef]
16. Shi, C.; Liu, D.; Cui, Y.; Xie, J.; Roberts, N.J.; Jiang, G. Amur Tiger Stripes: Individual Identification Based on Deep Convolutional Neural Network. *Integr. Zool.* **2020**, *15*, 461–470. [CrossRef] [PubMed]
17. Shi, C.; Xu, J.; Roberts, N.J.; Liu, D.; Jiang, G. Individual Automatic Detection and Identification of Big Cats with the Combination of Different Body Parts. *Integr. Zool.* **2023**, *18*, 157–168. [CrossRef] [PubMed]
18. Zajdel, W.; Zivkovic, Z.; Krose, B. Keeping Track of Humans: Have I Seen This Person Before? In Proceedings of the 2005 IEEE International Conference on Robotics and Automation, Barcelona, Spain, 18–22 April 2005; pp. 2081–2086.
19. Schneider, S.; Taylor, G.W.; Linquist, S.; Kremer, S.C. Past, Present and Future Approaches Using Computer Vision for Animal Re-identification from Camera Trap Data. *Methods Ecol. Evol.* **2019**, *10*, 461–470. [CrossRef]
20. Whitehead, H. Computer Assisted Individual Identification of Sperm Whale Flukes. In *Individual Recognition of Cetaceans: Use of Photo-Identification and Other Techniques to Estimate Population Parameters*; Hammond, P.S., Mizroch, S.A., Donovan, G.P., Eds.; Report of the International Whaling Commission: Cambridge, UK, 1990; pp. 71–77.

21. Ravela, S.; Gamble, L. On Recognizing Individual Salamanders. In Proceedings of the 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '04), Montreal, QC, Canada, 17–21 May 2004; Volume 5, pp. 742–747.
22. Carter, S.J.; Bell, I.P.; Miller, J.J.; Gash, P.P. Automated Marine Turtle Photograph Identification Using Artificial Neural Networks, with Application to Green Turtles. *J. Exp. Mar. Biol. Ecol.* **2014**, *452*, 105–110. [[CrossRef](#)]
23. Li, S.; Li, J.; Tang, H.; Qian, R.; Lin, W. ATRW: A Benchmark for Amur Tiger Re-identification in the Wild. In Proceedings of the 28th ACM International Conference on Multimedia, Seattle, WA, USA, 12–16 October 2020; pp. 2590–2598.
24. Liu, C.; Zhang, R.; Guo, L. Part-Pose Guided Amur Tiger Re-Identification. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Republic of Korea, 27–28 October 2019; pp. 315–322.
25. Xu, N.; Ma, Z.; Xia, Y.; Dong, Y.; Zi, J.; Xu, D.; Xu, F.; Su, X.; Zhang, H.; Chen, F. A Serial Multi-Scale Feature Fusion and Enhancement Network for Amur Tiger Re-Identification. *Animals* **2024**, *14*, 1106. [[CrossRef](#)] [[PubMed](#)]
26. Bai, X.; Islam, T.; Bin Azhar, M.A.H. Transformer-Based Models for Enhanced Amur Tiger Re-Identification. In Proceedings of the 2024 IEEE 22nd World Symposium on Applied Machine Intelligence and Informatics (SAMI), Stará Lesná, Slovakia, 18–20 January 2024; pp. 000411–000416.
27. Bohnett, E.; Holmberg, J.; Faryabi, S.P.; Li, A.; Ahmad, B.; Rashid, W.; Ostrowski, S. Determining Snow Leopard (*Panthera uncia*) Occupancy in the Pamir Mountains of Afghanistan. *Ecol. Inform.* **2023**, *77*, 102214. [[CrossRef](#)]
28. Solari, K.A.; Ahmad, S.; Armstrong, E.E.; Campana, M.G.; Ali, H.; Hameed, S.; Ullah, J.; Khan, B.U.; Nawaz, M.A.; Petrov, D.A. Next-Generation Snow Leopard Population Assessment Tool: Multiplex-PCR SNP Panel for Individual Identification from Faeces. *Mol. Ecol. Resour.* **2025**, *25*, e14074. [[CrossRef](#)] [[PubMed](#)]
29. Zheng, L.; Shen, L.; Tian, L.; Wang, S.; Wang, J.; Tian, Q. Scalable Person Re-identification: A Benchmark. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1116–1124.
30. Salman, S.; Liu, X. Overfitting Mechanism and Avoidance in Deep Neural Networks. *arXiv* **2019**, arXiv:1901.06566. [[CrossRef](#)]
31. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Li, F.-F. ImageNet: A Large-Scale Hierarchical Image Database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
32. Ni, X.; Fang, L.; Huttunen, H. Adaptive L2 Regularization in Person Re-Identification. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 9601–9607.
33. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 618–626.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.