Syn-Tiger-360: Synthesizing 360° Biometric Representations for Tiger Re-Identification

Seven Song[1], Weijia Xu[1], Bin Zhu[2], Yichun Peng[3], Xi Guo[4], Lei Bao[5*], Jianping Ge[5*]

[1]School of Artificial Intelligence, Beijing Normal University
[2]School of Electronic and Computer Engineering, Peking University
[3]University of the Chinese Academy of Sciences
[4]Faculty of Geographical Science, Beijing Normal University
[5]College of Life Sciences, Beijing Normal University

Abstract

Individual tiger re-identification using camera traps is essential for effective, non-invasive wildlife monitoring. However, severe data scarcity and quality issues, such as sparse views, occlusions, and lighting variations, result in insufficient data for training robust re-identification models. While synthetic data generation offers a promising solution to training data scarcity, traditional 2D generative models (e.g., Stable Diffusion) fail to accurately capture both tiger pose and their surface-asymmetric stripe patterns. This leads to inconsistent biometric representations across different viewing angles. We introduce a novel framework leveraging image-to-360° video foundation models to synthesize rotation-consistent volumetric tiger biometrics. We present *Syn-Tiger-360* — the first synthetic dataset for animal re-identification, featuring 518 high-fidelity tiger videos with consistent stripe patterns. Extensive experiments demonstrate that Re-ID models trained on synthesized tiger data can be directly applied to real-world tiger re-identification. This work opens new perspectives that generative foundation models can be utilized to advance wildlife monitoring, highlighting promising avenues for future ecological applications.

1 Introduction

Wildlife conservation faces unprecedented challenges in the era of accelerating biodiversity disappearance. This is particularly true for endangered apex predators (Krofel and Jerina 2016; Stier et al. 2016), such as tigers (*Panthera tigris*), which require urgent, technologically advanced interventions (Seimon et al. 2013; Seidensticker 2010). Non-invasive monitoring, achieved through the use of camera traps and computer vision, has emerged as a promising approach to wildlife conservation (Caravaggi et al. 2017; Schneider et al. 2019; Christin, Hervet, and Lecomte 2019). Modern technology can now routinely automate the classification (Tabak et al. 2019; Norouzzadeh et al. 2021; Willi et al. 2019) and detection (Song et al. 2024; Tan et al. 2022) of different species. Animals like tigers can be well recognized with current state-of-the-art classification models (Tan et al. 2022). However, the successful monitoring of tiger populations relies critically on animal re-identification (Re-ID), a process that enables individual-level tracking (Li et al. 2019). This capability could provide essential data such as population dynamics, territory ranges, survival rates, and anthropogenic threats (Schneider et al. 2019; Čermák et al. 2024; Wahltinez and Wahltinez 2024).

While reliable individual-level Re-ID is paramount for accurate wildlife monitoring, achieving robust Re-ID in the wild remains critically challenging due to severe data limitations (Lou et al. 2019; Zhong et al. 2018). The central impediment is the extreme paucity of diverse, high-quality visual data collected under real-world field conditions via camera traps. Camera traps inherently operate in challenging environments characterized by uncontrolled lighting, frequent occlusions, vegetation obstructions, and severely limited field of view (Newey et al. 2015). Moreover, apex predators like tigers inherently exist at very low population densities in the wild, drastically reducing potential encounters (Seimon et al. 2013; Seidensticker 2010; Ordiz et al. 2021). Consequently, the scarcity of high-fidelity data directly undermines the performance of Re-ID models, as these models require abundant data to achieve robust performance and generalization.

To address the problem of data scarcity, the most common solution is to generate more training data, thereby enhancing the performance of trained models (Akkem, Biswas, and Varanasi 2024; Pezoulas et al. 2024) with synthesized data. Fingerprints serve as unique biometric representations for individual identification. However, samples for specific individuals remain scarce. Methods (Maltoni et al. 2009; Cappelli, Maio, and Maltoni 2002) are proposed to address this data limitation by synthesizing more fingerprints, thereby improving the capability of the model. Tiger stripes function as uniquely identifying biometric signatures, analogous to human fingerprints (Hiby et al. 2009). To address the challenge of data limitations, a natural and intuitive approach is to synthesize more tiger images with different stripe patterns as training data via image generative foundation models (e.g., Stable Diffusion (Rombach et al. 2022; Podell et al. 2024)).

However, this naive attempt fundamentally fails to capture the essential biometric traits of wildlife. Tiger stripe patterns simultaneously occupy two lateral planes; these identity-critical patterns are distributed asymmetrically across the left and right flanks, exhibiting non-mirror-symmetric configurations with intricate topological inter-

*Corresponding authors: Jianping Ge (gejp@bnu.edu.cn) and Lei Bao (baolei@bnu.edu.cn).

(a) Tigers captured by camera traps (b) Tigers generated by stable diffusion model

Figure 1: (a) Identified individual Amur tigers from the Northeast China Tiger and Leopard National Park, captured by camera traps (Wang et al. 2020b). (b) Tigers generated by the Stable Diffusion model. Unlike real-world tigers, these synthesized isolated images are impractical for training Re-Id models as they lack multi-view information of an individual tiger.

dependence (Reddy and Aravind 2012; Zuffi, Kanazawa, and Black 2018). As shown in Fig. 1, synthesizing *isolated 2D images* catastrophically simplifies this complex spatial reality. In the real world, stripes of tigers undergo continuous three-dimensional deformation. Naive 2D generative solutions (Podell et al. 2024; Rombach et al. 2022) are unable to simulate this spatially coherent transformation process. These isolated 2D images contain spurious stripe discontinuities, falsified topological connections, and perspective-inconsistent biometric representations of individuals (Burghardt and Campbell 2007). Consequently, Re-ID models trained on such synthetically augmented data exhibit poor real-world generalization due to the significant fidelity gap between synthetic images and real-world data.

To bridge this fundamental gap, we propose a paradigm shift from isolated 2D synthesis to comprehensive 3D representation. Unlike conventional methods that generate isolated viewpoints, our approach aims to create 360° biometric representations of tigers by leveraging recent advances in **image-to-video foundation models** (Singer et al. 2022; Yang et al. 2024; Wu et al. 2023). We adapt an image-to-video foundation model (Wan et al. 2025) using a pre-trained LoRA module (Hu et al. 2022) specialized for 360° rotation effects. This enables dynamic viewpoint synthesis without species-specific retraining. Our solution transforms 2D tiger images into volumetric biometric representations that preserve stripe pattern consistency across viewpoints and maintain topological relationships during transitions. The synthesized models naturally simulate dynamic occlusions/deformations during perspective changes, capturing continuous 3D spatial transformations of stripes – achieving pose coverage unattainable with 2D pipelines.

Building upon this framework, we further introduce **Syn-Tiger-360** - the first synthesized dataset comprising 518 360° tiger videos. Each video synthesizes complete stripe pattern variations across dorsal, lateral, and transitional body regions through full-body rotations. Experimental results demonstrate that models trained on Syn-Tiger-360 achieve significant gains in Re-ID, validating the efficacy of our 3D synthesis paradigm for wildlife biometrics. In summary, the contributions of this work can be summarized as:

- A novel data synthesis paradigm that integrates large generative foundation models, solving animal re-ID training data scarcity by providing high-fidelity training samples for robust re-identification models.
- The first synthetic dataset for wildlife biometrics - Syn-Tiger-360, featuring 518 high-fidelity 360° tiger videos capturing the complete stripe pattern across the entire body. This dataset fills the critical gap in high-quality training data for endangered species monitoring.
- Extensive experimental results demonstrate the capability of our framework in real-world tiger re-ID.

2 Background

2.1 Generative Foundation Models

Generative foundation models exhibit remarkable capabilities in creating diverse, high-fidelity data. Latent Diffusion Models (LDMs), pioneered by Stable Diffusion (Rombach et al. 2022), represent a paradigm shift in image synthesis. By operating efficiently in compressed latent spaces, it enables high-resolution image generation conditioned on semantic inputs like CLIP text embeddings (Radford et al. 2021), forming the backbone of modern text-to-image synthesis.

These generative capabilities extend robustly into the video domain. Pre-trained video foundation models (Blattmann et al. 2023; Wu et al. 2023; Lin et al. 2024a,b; Wan et al. 2025) demonstrate powerful spatiotemporal representation learning. By effectively leveraging techniques such as temporal attention layers and 3D convolutions on their LDM backbone, these models achieve compelling image-to-video synthesis, capable of generating diverse, dynamic scenes. Nonetheless, a key limitation persists: conventional approaches typically lack inherent, explicit 3D

scene understanding, which often manifests as visual inconsistencies during viewpoint changes, such as object rotations. Building upon this strong pre-trained representation capability, recent advances address the challenge of generating coherent multi-view sequences. 360° Synthesis Techniques harness the power of foundation models by applying specialized fine-tuning strategies. For instance, methods like (Cai et al. 2024) integrate geometric priors, leveraging the existing generative prowess of diffusion models to produce consistent rotation sequences from single images. Specifically exploiting the rich representations of pre-trained video diffusion models, our approach leverages Low-Rank Adaptation (LoRA) (Hu et al. 2022) fine-tuning on datasets of rotating 360° video clips. This lightweight adaptation tailors an existing 360°-capable video diffusion architecture, enabling precise control over continuous rotation patterns around subjects while maintaining topological coherence. This is crucial for generating biometrically consistent circumferential views of complex subjects (*i.e.*, tigers, zebras, and *etc.*).

2.2 Animal Re-Identification

Animal re-identification (Re-ID) constitutes a critical foundation for wildlife conservation by enabling individual-level monitoring (Li et al. 2019; Ye et al. 2024). Accurately tracking individuals facilitates research on animal movement, migration patterns, population dynamics, and social behavior, providing indispensable data for managing endangered species, habitats, and biodiversity (Schneider et al. 2019). Current Re-ID methods leverage species-specific visual biometrics, such as zebra stripes (Lahiri et al. 2011), penguin ventral spots (Noboru, Ozasa, and Tanaka 2024), and tiger flank patterns (Shukla et al. 2019; Liu, Zhang, and Guo 2019; Cheng et al. 2020; Liu et al. 2019).

Tiger Re-ID fundamentally depends on unique stripe patterns—biometric identifiers exhibiting asymmetric bilateral distribution with non-mirror-symmetric topological interdependence (Hiby et al. 2009; Zuffi, Kanazawa, and Black 2018). This contrasts sharply with human Re-ID, where standardized benchmarks provide dense multi-pose data (Zheng et al. 2015, 2017), while wild tigers inherently suffer from severe pose-limited data scarcity due to low population densities and camera-trap constraints (Newey et al. 2015). Consequently, pose variations catastrophically disrupt identification accuracy: Partial feature alignment networks (Liu, Zhang, and Guo 2019; Yu et al. 2019) fail without adequate training diversity, texture matching across poses (Shukla et al. 2019), and field 3D capture remains impractical (Hiby et al. 2009). Bridging this gap requires large-scale datasets that capture the real-world complexity of pose, which are essential for translating re-ID research into effective conservation practices. However, traditional synthetic data—including diffusion models (Rombach et al. 2022; Podell et al. 2024; Sun et al. 2025)—only compounds these issues by generating isolated 2D samples with spurious stripe discontinuities and perspective-inconsistent artifacts that violate topological integrity (Burghardt and Campbell 2007). This emphasizes the importance of a comprehensive 360° biometric representation as it delivers complete biological fidelity, resolving the fragmented approximations endemic to 2D synthesis paradigms. Therefore, we leverage recent generative foundation models to synthesize tiger samples with comprehensive 360° biometric representation, effectively bridging this fundamental gap.

3 Method

Our approach establishes a sequential text-to-image-to-3D pipeline(Fig. 2) that overcomes the viewpoint limitations of conventional 2D generation by leveraging state-of-the-art generative foundation models. The methodology consists of two integrated stages that transform textual descriptions of tiger stripes into comprehensive, 3D-consistent representations of individual tigers.

3.1 Synthesizing 2D Tiger Images

The initial stage utilizes Stable Diffusion v2.1 (Podell et al. 2024) to transform structured textual descriptions of tiger stripe patterns into 1024×1024 pixel reference images. Natural language prompts specify defining characteristics (e.g., *"A tiger stands still, captured in a standard profile side view, in wildlife photography. Complete body from tail to head. The body and head are horizontal."*). We generate images using DDIM sampling (Song, Meng, and Ermon 2021) with classifier-free guidance (Ho and Salimans 2021), formalized by the conditioned reverse diffusion process:

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \epsilon_\theta(\mathbf{x}_t, t, \tau_\theta(y)) \right) + \sigma_t \mathbf{z} \quad (1)$$

where ϵ_θ is the denoising network, $\tau_\theta(y)$ denotes text embeddings, $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$, σ_t controls stochasticity and $\mathbf{z} \sim \mathcal{N}(0, \mathbf{I})$. Topological consistency emerges from learned priors:

$$\nabla_{\mathbf{x}} \log p_\theta(\mathbf{x}_t|y) = -\frac{1}{\sqrt{1-\bar{\alpha}_t}} \epsilon_\theta(\mathbf{x}_t, t, \tau_\theta(y)) \quad (2)$$

ensuring coherent stripe structures without explicit constraints.

3.2 Synthesizing 360° Tiger Videos

To extend 2D representations into 3D-consistent sequences, we utilize a video diffusion foundation model designed explicitly for temporal synthesis. This approach learns implicit 3D consistency from large video datasets, preserving unique stripe identities across rotation angles while maintaining individual characteristics.

We apply a generative image-to-video foundation model, enhanced with rotation-specific LoRA adaptation. The model jointly synthesizes all video frames $\mathbf{V} = [\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(K)}]$ via:

$$\mathbf{V}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{V}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \epsilon_\theta(\mathbf{V}_t, t, \tau(\mathbf{I}_0)) \right) + \sigma_t \mathbf{Z}. \quad (3)$$

Viewpoint consistency is ensured through rotation-specific LoRA (Hu et al. 2022) adaptation:

$$\mathbf{W}' = \mathbf{W} + \mathbf{B}\mathbf{A}^\top, \quad \mathbf{A}, \mathbf{B} \in \mathbb{R}^{d \times 64}, \quad (4)$$

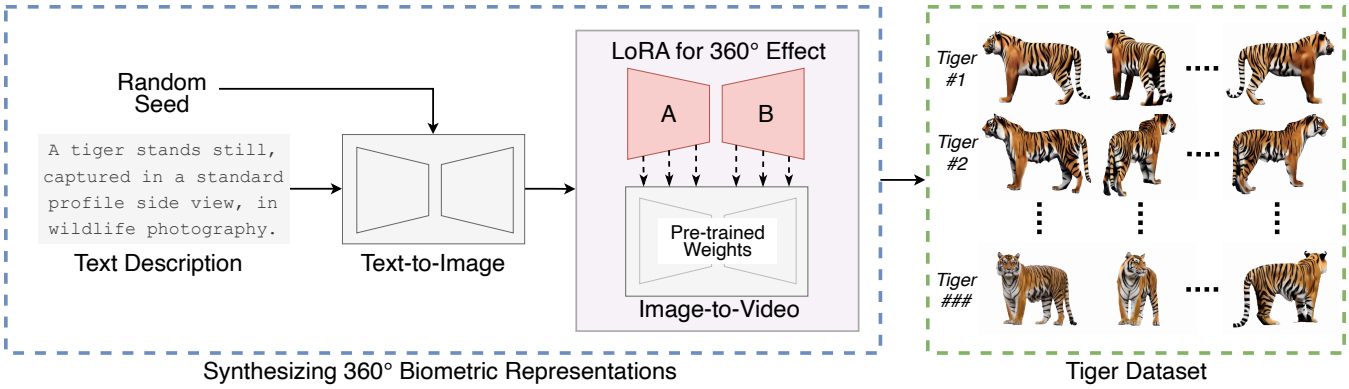

Figure 2: Creation Pipeline of Syn-Tiger-360 Dataset.

	Classification		Re-Identification				
	Train	Test	Train	Test		Real-world Test	
				Query	Gallery	Query	Gallery
Num. of Tigers	518 (syn)		466 (syn)	52 (syn)		12 (real)	12 (real) & 52 (syn)
Imgs. per Tigers	48	12	64	2	8	2	8

Table 1: Dataset splitting details of Syn-tiger-360.

which are trained on rotational sequences, this adaptation module enforces continuous stripe flow while preserving identity-specific patterns. The resulting 360° sequences maintain consistent anatomy and unique stripe configurations across all viewpoints, capturing the complete 3D structure of individual tigers.

3.3 Implementation Details

Our pipeline was executed within the ComfyUI workflow management framework[1], leveraging its node-based architecture to integrate specialized components. For the text-to-image stage, we implemented a custom parsing module that processes structured stripe descriptions, using Stable Diffusion v2.1 (Podell et al. 2024) at a resolution of 1024×1024 with 50-step DDIM sampling and classifier-free guidance (CFG) (Ho and Salimans 2021) scale 7.5 to maintain stripe fidelity.

The generative foundation base model `Wan2.1-14B`[2] is used for video generation. A 360 Degree Rotation Effect LoRA[3] is used to augment the base model. The LoRA was trained for 20 epochs in rotational video sequences and activated using the trigger in textual prompts. We applied the default parameters, including a LoRA strength of 1.0, an embedded guidance scale of 6.0, and a flow shift of 5.0, to optimize stripe consistency during rotation. The low-rank adaptation operated with weight matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{d\times 64}$ (where d denotes layer dimension), implementing the parameter transformation $\mathbf{W}' = \mathbf{W} + \Delta\mathbf{W} = \mathbf{W} + \mathbf{B}\mathbf{A}^\top$ while freezing original weights $\mathbf{W}$ during inference.

The workflow configuration utilized a modified version of Kijai's Wan Video Wrapper[4], which incorporated a dedicated LoRA node, sequentially chaining text prompt encoding (including rotation triggers), stripe-constrained image generation, and multi-view video synthesis. This arrangement outputted 16-frame sequences at a 512×512 resolution, maintaining temporal coherence through specialized attention blocks tuned to an attention scale of 1.2 for enhanced stripe persistence across rotation angles. All experiments were run on one NVIDIA A100 GPU, following the base model hyper-parameters except for temporal adjustments that ensured continuous pattern flow throughout 360° rotations. The workflow implementation directly incorporated the default configuration for seamless LoRA integration.

4 Experimental Results

We introduce a paradigm shift for wildlife data scarcity: generating 360° biometric representations using generative image-to-video models. Unlike traditional 2D methods, which fail to model complex stripe morphology, our approach produces consistent patterns essential for re-identifying wildlife animals. We further establish *Syn-Tiger-360* — the first synthetic dataset for 360° biometric representation of wildlife, thereby facilitating enhanced tiger monitoring and paving the way for more comprehensive future research.

In the subsequent subsections, we systematically examine two key dimensions of the introduced method: (1) enforcing

[1]https://github.com/comfyanonymous/ComfyUI
[2]https://huggingface.co/Wan-AI/Wan2.1-I2V-14B-480P
[3]https://huggingface.co/Remade-AI/Rotate
[4]https://github.com/kijai/ComfyUI-WanVideoWrapper

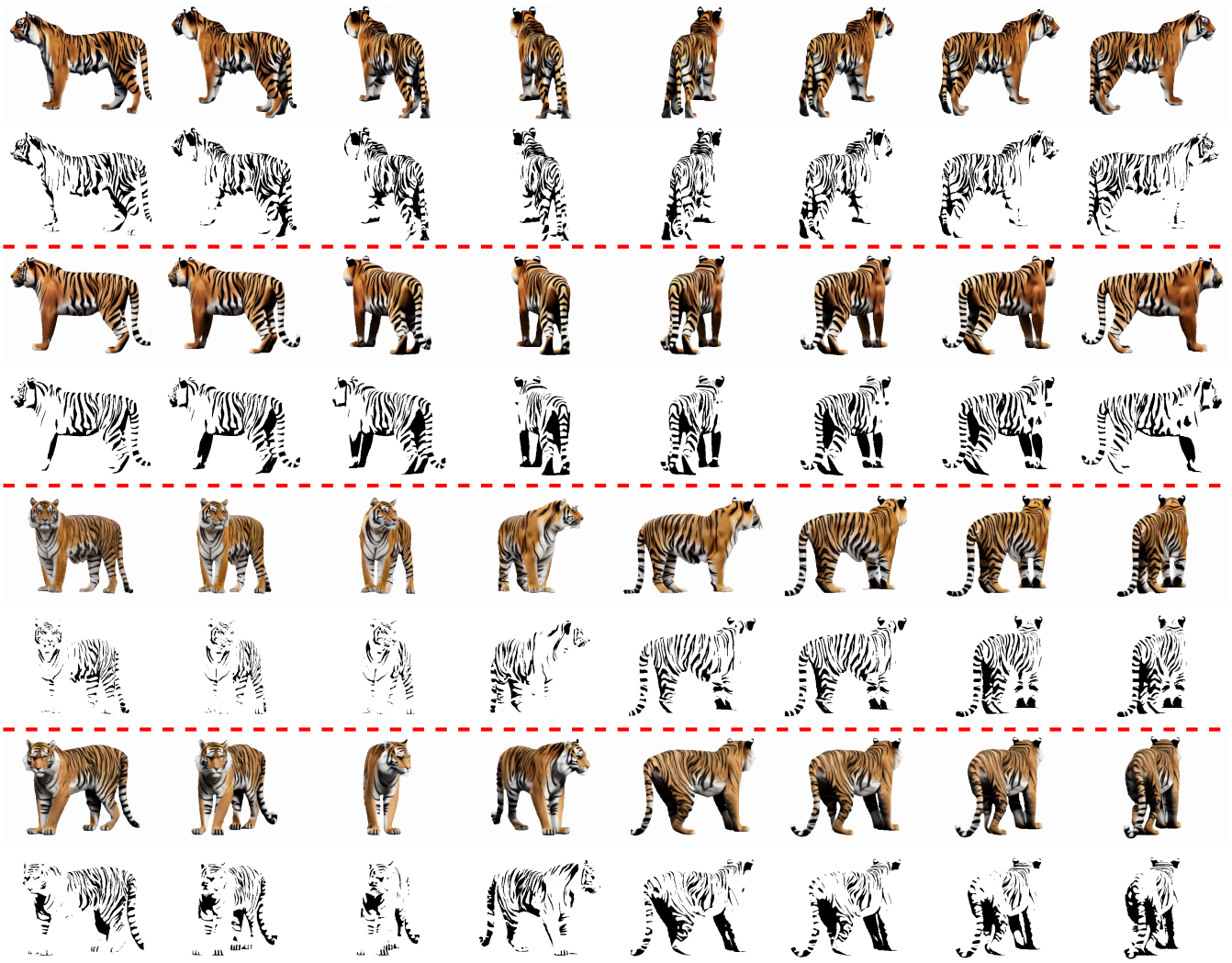

Figure 3: Synthesized 360° biometric representation of tigers. From top to bottom, each group contains multi-view images of a synthesized tiger.

Backbone	@Top1 Acc.	@Top5 Acc.	F1 Score	Precision	Recall
Resnet-50 (He et al. 2016)	0.9768	0.9966	0.9762	0.9809	0.9781
Resnet-101 (He et al. 2016)	0.9773	0.9971	0.9723	0.9754	0.9761
$\text{ConvNeXt}_{\text{base}}$ (Liu et al. 2022b)	0.9686	0.9957	0.9650	0.9727	0.9677
$\text{ConvNeXt}_{\text{large}}$ (Liu et al. 2022b)	0.9986	1.0000	0.9987	0.9989	0.9988
ViT_{base} (Dosovitskiy et al. 2021)	0.9532	0.9942	0.9421	0.9500	0.9485
$\text{ViT}_{\text{large}}$ (Dosovitskiy et al. 2021)	0.9860	0.9990	0.9781	0.9798	0.9815
$\text{Swin-T}_{\text{base}}$ (Liu et al. 2021)	0.9730	0.9957	0.9671	0.9745	0.9701
$\text{Swin-T}_{\text{large}}$ (Liu et al. 2021)	0.9747	0.9976	0.9678	0.9727	0.9692
$\text{BEiT}_{\text{base}}$ (Bao et al. 2022)	0.9513	0.9889	0.9429	0.9510	0.9489
$\text{BEiT}_{\text{large}}$ (Bao et al. 2022)	0.9363	0.9937	0.9302	0.9426	0.9377
DeiT (Touvron et al. 2021)	0.9667	0.9957	0.9646	0.9701	0.9685

Table 2: Classification results on syn-tiger-360 test-set.

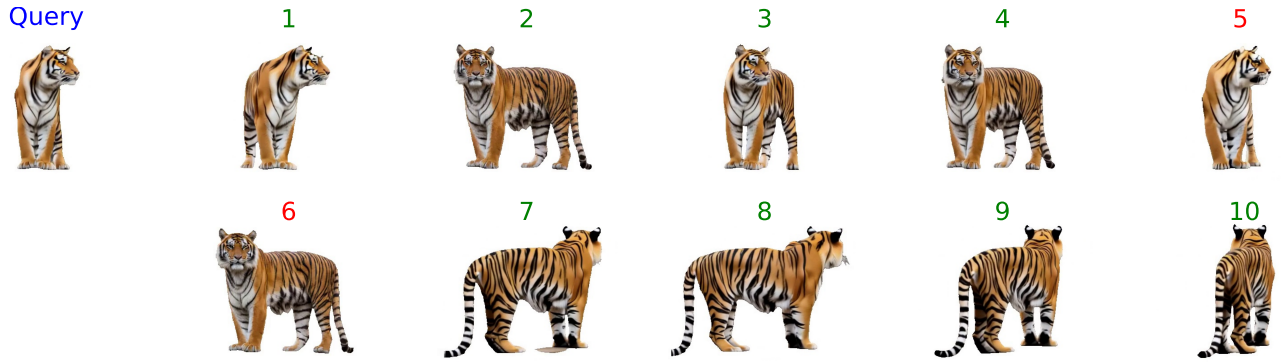

Figure 4: Re-identification results on syn-tiger-360 Re-id test-set. The green and red color denotes true and false match.

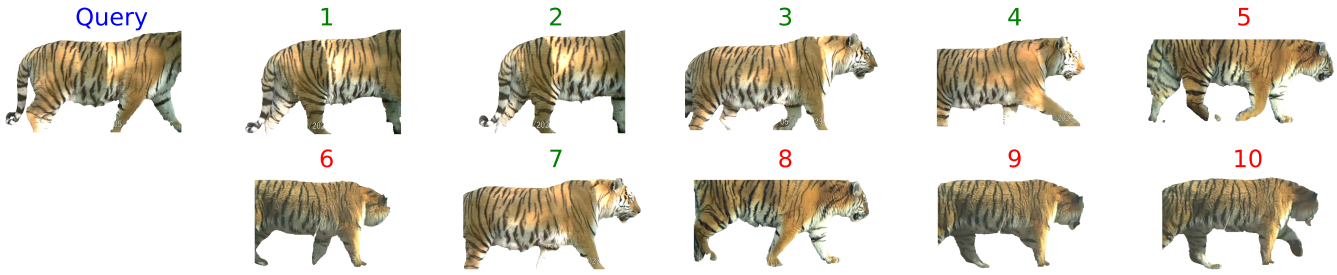

Figure 5: Re-identification results on Real-world test-set. The green and red color denotes true and false match.

360° topological consistency for biometric fidelity through validated stripe coherence (Sec. 4.1), (2) boosting real-world wildlife re-identification through training re-ID models on syn-tiger-360 dataset (Sec. 4.2).

4.1 Synthesized 360° Biometric Representation

Conventional 2D generative models (e.g., Stable Diffusion) fundamentally fail to model holistic stripe topology across the animal's form, yielding disconnected views with severe discontinuities and perspective artifacts (Fig. 1) that distort critical re-identification features and obscure true morphological understanding. Leveraging the generalization capabilities of large foundation models (Podell et al. 2024; Wan et al. 2025), we achieve zero-shot synthesis of 360° ecological signatures of tigers despite scarce 3D training data, producing full rotational videos (Fig. 3). This capability directly addresses the critical bottleneck of data scarcity in wildlife monitoring, generating high-fidelity, 3D-consistent representations essential for robust biometric analysis.

Through the generational prior of foundation models, Syn-Tiger-360 possesses the feature of simulating consistent stripe deformation across dorsal and lateral planes. This transcends technical improvement to address a biometric necessity: tiger stripes constitute intrinsic 3D signatures that require relational continuity across viewpoints (Reddy and Aravind 2012). Critically, these synthetically generated 360° representations provide detailed insights into the complex 3D morphology and genuine biometric variance of tiger pelage patterns, surpassing the limitations of sparse and fragmented real-world data.

4.2 Real-World Tiger Re-Identification

This research introduces a novel paradigm for generating reliable training data to build robust wildlife re-identification models. By synthesizing 3D-consistent volumetric representations of tiger, we overcome critical limitations in wildlife monitoring, where real-world data is inherently scarce and fragmented. The detailed splitting of the dataset is presented in Table 1. We comprehensively evaluate the utility of *Syn-Tiger-360* through three core experiments. First, classification tasks on synthetic tigers validate that deep networks can discern distinct biometric identities, as demonstrated in Table 2. Second, we evaluate the re-identification performance within the synthetic domain across diverse rotational viewpoints using multiple backbones (Dosovitskiy et al. 2021; He et al. 2016), with quantitative results detailed in Table 3. Third, for real-world validation, we experiment on 12 identified Amur tigers from Northeast China Tiger and Leopard National Park (Wang et al. 2020b), utilizing models pretrained on *Syn-Tiger-360* to extract discriminative features (Table 4).

Experimental results confirm the efficacy of our approach. Synthetic tigers exhibit exceptional biometric distinctness, evidenced by 99.86% top-1 classification accuracy (Table 2). Crucially, synthetic re-identification achieves 99% rank-1 accuracy under rotational variations (Table 3), confirming viewpoint robustness unattainable with 2D data. Models pretrained on *Syn-Tiger-360* substantially enhance real-world

Backbone	Rank-1	Rank-2	Rank-3	mAP	IDF1
ResNet (He et al. 2016)	0.9615	0.9903	1.0000	0.8358	0.7254
DensenNet (Huang et al. 2017)	0.9711	0.9903	1.0000	0.8376	0.7350
Swin-T (Liu et al. 2021)	0.9615	1.0000	1.0000	0.8588	0.7435
Swin-T-v2 (Liu et al. 2022a)	0.9903	1.0000	1.0000	0.8185	0.7168
HR-Net (Wang et al. 2020a)	0.9711	0.9903	1.0000	0.8333	0.7286
Convnext (Liu et al. 2022b)	0.9615	0.9711	0.9807	0.8186	0.7115

Table 3: Re-identification results on syn-tiger-360 dataset test-set. Please note that this evaluation is conducted on the tigers from the test-set, which are not used for training the Re-ID model.

Backbone	Rank-1	Rank-2	Rank-3	mAP	IDF1
ResNet (He et al. 2016)	0.9736	1.0000	1.0000	0.5735	0.4928
DensenNet (Huang et al. 2017)	1.0000	1.0000	1.0000	0.5830	0.4980
Swin-T (Liu et al. 2021)	0.9473	0.9473	0.9473	0.5912	0.4998
Swin-T-v2 (Liu et al. 2022a)	0.9210	0.9473	0.9473	0.5235	0.4380
HR-Net (Wang et al. 2020a)	0.9473	1.0000	1.0000	0.5593	0.4668
Convnext (Liu et al. 2022b)	0.9736	1.0000	1.0000	0.5845	0.5034

Table 4: Re-identification results on Real-world tiger re-id test-set (12 real tigers and 52 synthesized tigers). These tiger individuals from the test set were not used for training the Re-ID model.

tiger re-identification, yielding up to 100% rank-1 accuracy and 58.3% mAP improvement as quantified in Table 4.

To visually validate the reliability of our re-identification system, we present qualitative retrieval results in Fig. 4 for synthetic tigers and Fig. 5 for real-world tigers. These examples demonstrate consistent matching of query images to the correct gallery identities, even across challenging viewpoint changes and partial occlusions. It is important to note that all queries are performed on test-sets, which are strictly separated from the training data (as detailed in Table 1). This ensures the evaluation reflects the model's generalization to unseen individuals. These results further validate the ability to accurately match individuals against gallery databases, underscoring the operational utility of our approach in field conservation scenarios. The collective evidence confirms that re-identification models trained on our synthetic tiger data can be effectively and directly applied to real-world tiger conservation efforts.

5 Conclusion

This work addresses the critical data scarcity crisis in wildlife conservation by establishing a novel 360° generative paradigm for endangered species monitoring. By integrating image-to-360° video foundation models with specialized adaptation techniques, we generate volumetric tiger representations that preserve topological stripe consistency across all viewpoints, overcoming the fundamental limitation of 2D synthesis methods, which fail to capture complex spatial relationships in biometric patterns. The resulting *Syn-Tiger-360* dataset provides the first comprehensive resource for 360° wildlife biometrics, enabling robust re-identification models validated through real-world deployment on endangered Amur tigers. Crucially, this framework extends beyond tigers to other striped species, establishing a transformative approach for non-invasive population monitoring where viewpoint variation and topological integrity determine conservation efficacy.

Limitations and Future Work While this paper represents a novel method to advance wildlife monitoring, two core limitations warrant consideration. First, it generates rotation sequences in controlled settings, lacking natural behaviors like walking or crouching that reflect real-world conditions (Schneider et al. 2019). Second, performance gains shown in controlled benchmarks need validation with real camera-trap data, which often includes occlusions and environmental noise.

To address these, we propose three research directions. First, using text-to-video or 3D decoupling methods could enable synthesis of dynamic, behaviorally realistic sequences. Second, hybrid pipelines combining Syn-Tiger-360 with sparse real-world images could enhance ecological relevance while maintaining biometric accuracy. Third, collaboration with conservation societies could create challenge datasets from real tiger sightings, establishing standardized evaluation protocols to bridge the simulation-to-field gap. Future work could also incorporate species-specific priors, such as tailored LoRAs (Hu et al. 2022), to refine details like fur texture and movement patterns while preserving 360° viewpoint consistency.

Acknowledgment

This research is funded by the National Key Research and Development Program of China (grant number 2024YFF1307301).

References

Akkem, Y.; Biswas, S. K.; and Varanasi, A. 2024. A comprehensive review of synthetic data generation in smart farming by using variational autoencoder and generative adversarial network. *Engineering Applications of Artificial Intelligence*, 131: 107881.

Bao, H.; Dong, L.; Piao, S.; and Wei, F. 2022. BEiT: BERT Pre-Training of Image Transformers. In *International Conference on Learning Representations*.

Blattmann, A.; Dockhorn, T.; Kulal, S.; Mendelevitch, D.; Kilian, M.; Lorenz, D.; Levi, Y.; English, Z.; Voleti, V.; Letts, A.; et al. 2023. Stable video diffusion: Scaling latent video diffusion models to large datasets. *arXiv preprint arXiv:2311.15127*.

Burghardt, T.; and Campbell, N. 2007. Individual animal identification using visual biometrics on deformable coat patterns. In *International Conference on Computer Vision Systems: Proceedings (2007)*. Citeseer.

Cai, Z.; Mueller, M.; Birkl, R.; Wofk, D.; Tseng, S.-Y.; Cheng, J.; Stan, G. B.-M.; Lai, V.; and Paulitsch, M. 2024. L-MAGIC: Language Model Assisted Generation of Images with Coherence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7049–7058.

Cappelli, R.; Maio, D.; and Maltoni, D. 2002. Synthetic fingerprint-database generation. In *2002 International Conference on Pattern Recognition*, volume 3, 744–747. IEEE.

Caravaggi, A.; Banks, P. B.; Burton, A. C.; Finlay, C. M.; Haswell, P. M.; Hayward, M. W.; Rowcliffe, M. J.; and Wood, M. D. 2017. A review of camera trapping for conservation behaviour research. *Remote Sensing in Ecology and Conservation*, 3(3): 109–122.

Čermák, V.; Picek, L.; Adam, L.; and Papafitsoros, K. 2024. WildlifeDatasets: An open-source toolkit for animal re-identification. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 5953–5963.

Cheng, X.; Zhu, J.; Zhang, N.; Wang, Q.; and Zhao, Q. 2020. Detection features as attention (defat): A keypoint-free approach to amur tiger re-identification. In *2020 IEEE International Conference on Image Processing (ICIP)*, 2231–2235. IEEE.

Christin, S.; Hervet, É.; and Lecomte, N. 2019. Applications for deep learning in ecology. *Methods in Ecology and Evolution*, 10(10): 1632–1644.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. 2021. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

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

Ho, J.; and Salimans, T. 2021. Classifier-Free Diffusion Guidance. In *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*.

Hu, E. J.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; Chen, W.; et al. 2022. LoRA: Low-Rank Adaptation of Large Language Models. In *International Conference on Learning Representations*.

Huang, G.; Liu, Z.; Van Der Maaten, L.; and Weinberger, K. Q. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4700–4708.

Krofel, M.; and Jerina, K. 2016. Mind the cat: Conservation management of a protected dominant scavenger indirectly affects an endangered apex predator. *Biological Conservation*, 197: 40–46.

Lahiri, M.; Tantipathananandh, C.; Warungu, R.; Rubenstein, D. I.; and Berger-Wolf, T. Y. 2011. Biometric animal databases from field photographs: identification of individual zebra in the wild. In *Proceedings of the 1st ACM international conference on multimedia retrieval*, 1–8.

Li, S.; Li, J.; Tang, H.; Qian, R.; and Lin, W. 2019. ATRW: a benchmark for Amur tiger re-identification in the wild. *arXiv preprint arXiv:1906.05586*.

Lin, B.; Ge, Y.; Cheng, X.; Li, Z.; Zhu, B.; Wang, S.; He, X.; Ye, Y.; Yuan, S.; Chen, L.; et al. 2024a. Open-sora plan: Open-source large video generation model. *arXiv preprint arXiv:2412.00131*.

Lin, B.; Ye, Y.; Zhu, B.; Cui, J.; Ning, M.; Jin, P.; and Yuan, L. 2024b. Video-LLaVA: Learning United Visual Representation by Alignment Before Projection. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, 5971–5984.

Liu, C.; Zhang, R.; and Guo, L. 2019. Part-pose guided amur tiger re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 0–0.

Liu, N.; Zhao, Q.; Zhang, N.; Cheng, X.; and Zhu, J. 2019. Pose-guided complementary features learning for amur tiger re-identification. In *Proceedings of the IEEE/CVF international conference on computer vision workshops*, 0–0.

Liu, Z.; Hu, H.; Lin, Y.; Yao, Z.; Xie, Z.; Wei, Y.; Ning, J.; Cao, Y.; Zhang, Z.; Dong, L.; et al. 2022a. Swin transformer v2: Scaling up capacity and resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 12009–12019.

Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; and Guo, B. 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, 10012–10022.

Liu, Z.; Mao, H.; Wu, C.-Y.; Feichtenhofer, C.; Darrell, T.; and Xie, S. 2022b. A convnet for the 2020s. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11976–11986.

Lou, Y.; Bai, Y.; Liu, J.; Wang, S.; and Duan, L. 2019. Veri-wild: A large dataset and a new method for vehicle re-identification in the wild. In *Proceedings of the IEEE/CVF*

conference on computer vision and pattern recognition, 3235–3243.

Maltoni, D.; Maio, D.; Jain, A. K.; and Prabhakar, S. 2009. Synthetic fingerprint generation. *Handbook of fingerprint recognition*, 271–302.

Newey, S.; Davidson, P.; Nazir, S.; Fairhurst, G.; Verdicchio, F.; Irvine, R. J.; and Van Der Wal, R. 2015. Limitations of recreational camera traps for wildlife management and conservation research: A practitioner's perspective. *Ambio*, 44: 624–635.

Noboru, Y.; Ozasa, Y.; and Tanaka, M. 2024. Hyperspectral Image Dataset for Individual Penguin Identification. In *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*, 9383–9387. IEEE.

Norouzzadeh, M. S.; Morris, D.; Beery, S.; Joshi, N.; Jojic, N.; and Clune, J. 2021. A deep active learning system for species identification and counting in camera trap images. *Methods in ecology and evolution*, 12(1): 150–161.

Ordiz, A.; Aronsson, M.; Persson, J.; Støen, O.-G.; Swenson, J. E.; and Kindberg, J. 2021. Effects of human disturbance on terrestrial apex predators. *Diversity*, 13(2): 68.

Pezoulas, V. C.; Zaridis, D. I.; Mylona, E.; Androutsos, C.; Apostolidis, K.; Tachos, N. S.; and Fotiadis, D. I. 2024. Synthetic data generation methods in healthcare: A review on open-source tools and methods. *Computational and structural biotechnology journal*.

Podell, D.; English, Z.; Lacey, K.; Blattmann, A.; Dockhorn, T.; Müller, J.; Penna, J.; and Rombach, R. 2024. SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis. In *The Twelfth International Conference on Learning Representations*.

Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, 8748–8763. PmLR.

Reddy, K. P. K.; and Aravind, R. 2012. Measurement of asymmetry of stripe patterns in animals. In *2012 International Conference on Signal Processing and Communications (SPCOM)*, 1–5. IEEE.

Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; and Ommer, B. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10684–10695.

Schneider, S.; Taylor, G. W.; Linquist, S.; and Kremer, S. C. 2019. Past, present and future approaches using computer vision for animal re-identification from camera trap data. *Methods in Ecology and Evolution*, 10(4): 461–470.

Seidensticker, J. 2010. Saving wild tigers: a case study in biodiversity loss and challenges to be met for recovery beyond 2010. *Integrative zoology*, 5(4): 285–299.

Seimon, T. A.; Miquelle, D. G.; Chang, T. Y.; Newton, A. L.; Korotkova, I.; Ivanchuk, G.; Lyubchenko, E.; Tupikov, A.; Slabe, E.; and McAloose, D. 2013. Canine distemper virus: an emerging disease in wild endangered Amur tigers (Panthera tigris altaica). *MBio*, 4(4): 10–1128.

Shukla, A.; Sigh Cheema, G.; Gao, P.; Onda, S.; Anshumaan, D.; Anand, S.; Farrell, R.; et al. 2019. A hybrid approach to tiger re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 0–0.

Singer, U.; Polyak, A.; Hayes, T.; Yin, X.; An, J.; Zhang, S.; Hu, Q.; Yang, H.; Ashual, O.; Gafni, O.; et al. 2022. Make-a-video: Text-to-video generation without text-video data. *arXiv preprint arXiv:2209.14792*.

Song, J.; Meng, C.; and Ermon, S. 2021. Denoising Diffusion Implicit Models. In *International Conference on Learning Representations*.

Song, Q.; Guan, Y.; Guo, X.; Guo, X.; Chen, Y.; Wang, H.; Ge, J.; Wang, T.; and Bao, L. 2024. Benchmarking wild bird detection in complex forest scenes. *Ecological Informatics*, 80: 102466.

Stier, A. C.; Samhouri, J. F.; Novak, M.; Marshall, K. N.; Ward, E. J.; Holt, R. D.; and Levin, P. S. 2016. Ecosystem context and historical contingency in apex predator recoveries. *Science Advances*, 2(5): e1501769.

Sun, H.; Cao, Y.; Dong, H.; and Fink, O. 2025. Unseen Visual Anomaly Generation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 25508–25517.

Tabak, M. A.; Norouzzadeh, M. S.; Wolfson, D. W.; Sweeney, S. J.; VerCauteren, K. C.; Snow, N. P.; Halseth, J. M.; Di Salvo, P. A.; Lewis, J. S.; White, M. D.; et al. 2019. Machine learning to classify animal species in camera trap images: Applications in ecology. *Methods in Ecology and Evolution*, 10(4): 585–590.

Tan, M.; Chao, W.; Cheng, J.-K.; Zhou, M.; Ma, Y.; Jiang, X.; Ge, J.; Yu, L.; and Feng, L. 2022. Animal detection and classification from camera trap images using different mainstream object detection architectures. *Animals*, 12(15): 1976.

Touvron, H.; Cord, M.; Douze, M.; Massa, F.; Sablayrolles, A.; and Jégou, H. 2021. Training data-efficient image transformers & distillation through attention. In *International conference on machine learning*, 10347–10357. PMLR.

Wahltinez, O.; and Wahltinez, S. J. 2024. An open-source general purpose machine learning framework for individual animal re-identification using few-shot learning. *Methods in Ecology and Evolution*, 15(2): 373–387.

Wan, T.; Wang, A.; Ai, B.; Wen, B.; Mao, C.; Xie, C.-W.; Chen, D.; Yu, F.; Zhao, H.; Yang, J.; et al. 2025. Wan: Open and advanced large-scale video generative models. *arXiv preprint arXiv:2503.20314*.

Wang, J.; Sun, K.; Cheng, T.; Jiang, B.; Deng, C.; Zhao, Y.; Liu, D.; Mu, Y.; Tan, M.; Wang, X.; et al. 2020a. Deep high-resolution representation learning for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 43(10): 3349–3364.

Wang, T.; Feng, L.; Yang, H.; Bao, L.; Wang, H.; and Ge, J. 2020b. An introduction to Long-term Tiger-Leopard Observation Network based on camera traps in Northeast China. *biodiversity science*, 28(9): 1059.

Willi, M.; Pitman, R. T.; Cardoso, A. W.; Locke, C.; Swanson, A.; Boyer, A.; Veldthuis, M.; and Fortson, L. 2019. Identifying animal species in camera trap images using deep learning and citizen science. *Methods in Ecology and Evolution*, 10(1): 80–91.

Wu, J. Z.; Ge, Y.; Wang, X.; Lei, S. W.; Gu, Y.; Shi, Y.; Hsu, W.; Shan, Y.; Qie, X.; and Shou, M. Z. 2023. Tune-a-video: One-shot tuning of image diffusion models for text-to-video generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 7623–7633.

Yang, Z.; Teng, J.; Zheng, W.; Ding, M.; Huang, S.; Xu, J.; Yang, Y.; Hong, W.; Zhang, X.; Feng, G.; et al. 2024. Cogvideox: Text-to-video diffusion models with an expert transformer. *arXiv preprint arXiv:2408.06072*.

Ye, M.; Chen, S.; Li, C.; Zheng, W.-S.; Crandall, D.; and Du, B. 2024. Transformer for object re-identification: A survey. *International Journal of Computer Vision*, 1–31.

Yu, J.; Su, H.; Liu, J.; Yang, Z.; Zhang, Z.; Zhu, Y.; Yang, L.; and Jiao, B. 2019. A strong baseline for tiger re-id and its bag of tricks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 0–0.

Zheng, L.; Shen, L.; Tian, L.; Wang, S.; Wang, J.; and Tian, Q. 2015. Scalable person re-identification: A benchmark. In *Proceedings of the IEEE international conference on computer vision*, 1116–1124.

Zheng, L.; Zhang, H.; Sun, S.; Chandraker, M.; Yang, Y.; and Tian, Q. 2017. Person re-identification in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1367–1376.

Zhong, Z.; Zheng, L.; Zheng, Z.; Li, S.; and Yang, Y. 2018. Camstyle: A novel data augmentation method for person re-identification. *IEEE Transactions on Image Processing*, 28(3): 1176–1190.

Zuffi, S.; Kanazawa, A.; and Black, M. J. 2018. Lions and tigers and bears: Capturing non-rigid, 3d, articulated shape from images. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 3955–3963.