Prior-Aware Synthetic Data to the Rescue: Animal Pose Estimation with Very Limited Real Data

Le Jiang Shuangjun Liu Xiangyu Bai Sarah Ostadabbas
Electrical and Computer Engineering Department, Northeastern University

BMVC 2022

Problem

Human share the world with and learn the behavior from with billions of animals. **However, low cooperativity and high species diversity prevent us from building comprehensive large-scale datasets to know our neighbors.** Further, the small data training is like cultivating on a tundra.

- Small data training makes the model lack of robustness and thus difficult to cope with more free-ranging movements, occlusions and environments.
- The synthetic animal data used as a supplement lacks realism of poses and it is difficult to be blended into the real background.

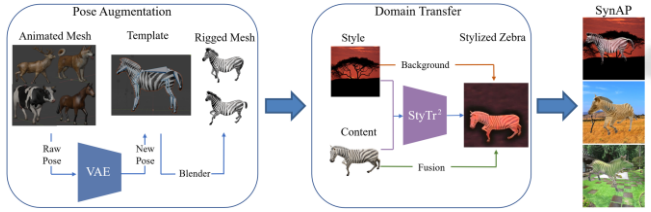
Approach

We present a cost-effective and generic prior-aware synthetic data generation pipeline, called **PASyn**, for animals pose estimate tasks that suffer from severe data scarcity.

- A novel variational autoencoder (VAE)-based synthetic animal data generation pipeline PASyn to generate probabilistically-valid pose data
- A style transfer strategy to militate the inconsistency between synthetic animal and real background
- A synthetic animal pose (SynAP) dataset, containing 3,000 zebra images and 3,600 images of six common quadrupeds

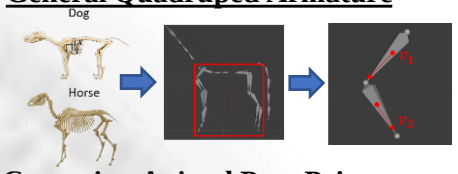
Prior-Aware Synthetic Data Generation

Our Prior-Aware Synthetic Data Generation (PASyn) Pipeline includes three main steps:



Pose Augmentation

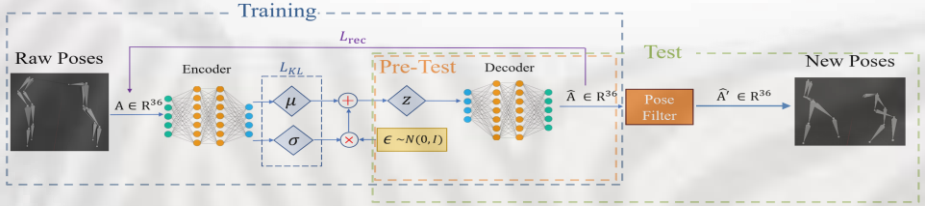
General Quadruped Armature



Training

$$\mathcal{L}_{\text{total}} = w_1\mathcal{L}_{KL} + w_2\mathcal{L}_{\text{rec}},$$
$$\mathcal{L}_{KL} = KL(q(z|A)||\mathcal{N}(0,I)),$$
$$\mathcal{L}_{\text{rec}} = \|A - \hat{A}\|_2^2$$

Capturing Animal Pose Prior

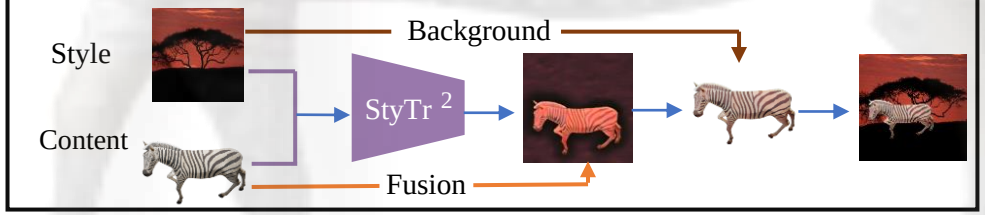


Stylization

Why style transfer

- Enrich the texture diversity in a reasonable way
- Blend the synthetic animal image into the real backgrounds

Architecture



Synthetic Animal Pose (SynAP) Dataset

Size

SynAP contains **3,000** synthetic zebra images and SynAP+ extends the SynAP with **3,600** images of horses, cows, sheep, dogs, giraffes and deer.

Background

300 grass, savanna, and forest real scenes are collected from Internet to stylize the synthetic animal.

Results

- Evaluation Over SynAP

Method	Backbone	Training Set	PCK@0.05 Pose Estimation Accuracy on Zebra-300 Set										
			Eye	Nose	Neck	Shoulders	Elbows	F-Paws	Hips	Knees	B-Paws	RoT	Average
MMPose [0]	HRNet-w32	R(99)	97.3	95.8	**83.2**	78.8	77.1	62.6	86.0	74.9	59.8	82.4	78.7
		R(99)+ S(3K)	**97.8**	**98.3**	81.1	**94.0**	**93.5**	**92.0**	93.7	93.5	89.0	**87.6**	**92.4**
		R(99)+ S(5K)	97.5	96.9	81.8	89.6	91.3	90.7	**94.1**	**94.1**	**90.4**	86.0	91.6
DeepLabCut [■]	EfficientNet-B6	R(99)	93.7	96.2	**82.5**	**91.4**	80.8	67.4	88.1	84.5	71.8	83.2	83.6
		R(99)+ S(3K)	**95.1**	**97.9**	81.5	90.1	83.3	75.5	**93.2**	89.3	83.9	86.8	87.6
		R(99)+ S(5K)	94.1	92.6	80.8	90.8	**87.0**	**85.7**	90.5	**93.3**	**88.3**	**86.8**	**89.2**
MMPose [0]	ResNet-50	R(99)	96.2	96.9	**80.8**	59.0	71.3	71.2	88.5	78.2	59.3	85.2	76.9
		R(99)+ S(3K)	95.6	95.8	69.9	**87.3**	**84.6**	84.3	90.8	91.2	84.4	77.2	86.7
		R(99)+ S(5K)	**97.0**	**97.9**	74.5	85.3	84.2	**84.5**	**94.6**	**91.4**	**88.0**	**85.6**	**88.4**

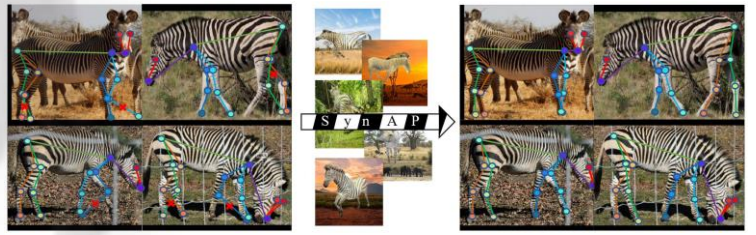
- The effect of SynAP with large real data

Training Set	Real Zebra	PCK@0.05 Pose Estimation Accuracy on Zebra-300 Set										
		Eye	Nose	Neck	Shoulders	Elbows	F-Paws	Hips	Knees	B-Paws	RoT	Average
R(8K) (SOTA)	✓	**97.5**	97.2	79.4	87.8	90.3	93.8	95.3	94.1	89.5	86.4	91.4
R(8K) + S(3K)	✓	97.3	**98.3**	79.0	93.1	94.9	96.0	95.3	**96.7**	93.3	**89.6**	93.8
R(8K) + S(5K)	✓	97.3	97.6	**81.1**	**93.7**	**95.7**	**96.0**	**96.6**	96.0	**94.3**	87.6	**94.2**
R(8K)	✗	79.7	87.7	37.4	77.6	80.0	87.6	82.0	86.4	81.3	67.2	78.3
R(8K) + S(3K)	✗	94.8	96.2	**67.1**	90.8	87.9	90.2	87.6	91.6	89.7	77.6	88.2
R(8K) + S(5K)	✗	**97.5**	**96.2**	66.1	**91.6**	**89.5**	**93.8**	**93.9**	**93.5**	**91.1**	**77.6**	**90.2**

- Ablation Study

Index	Training Set	VAE	Style Transfer	σ^2	Zoo Zebra Set	Zebra-300 Set
a	R(99)	✗	✗	✗	76.0	78.7
b	S(3K)	✗	✗	2I	38.7	30.0
c	S(3K)	✓	✗	2I	44.2	36.7
d	S(3K)	✓	✓	2I	42.9	46.6
e	R(99)+S(3K)	✗	✗	2I	89.8	88.0
f	R(99)+S(3K)	✓	✗	2I	90.4	89.8
g	R(99)+S(3K)	✓	✓	I	90.5	91.1
h	R(99)+S(3K)	✓	✓	2I	**91.5**	**92.4**

- Visualized Results



In conclusion, our synthetic animal pose dataset SynAP and its extended version SynAP+, and th of real data is verified on different backbones and achieves state-of-the positive effect of them on pose estimation task of animals with a small amounte-art performance.