1. Supplementary Materials

1.1. More Implementation Details

We disable the Thing-Class ImageNet Feature Distance(FD) [1] for all methods [1, 2] containing it. It is a regularization technique that uses ImageNet features which are trained from objects to provide guidance to segment object classes, which is inappropriate for segmenting semantics parts of object. In addition, we use the distribution-aware pixel contrast for SePiCo [3].

The tiger-like animals set contains the following classes from PartImageNet [4]: n02129604, n02125311, n02128385, n02130308. The horse-like animals set contains the following classes: n02403003, n02415577, n02423022, n02408429, n02412080, n02422699, n02437312, n02422106, n02417914.

1.2. Random Texture for SMAL Synthetic Data

In the CAD synthetic data, half of the synthetic images have real textures provided by the CAD models, which is shown in Figure 1, while the other half are using random textures(paste random real images on the 3D models without any UV mapping). However, we do not use real textures in SMAL synthetic data generation. We have the following two reasons: (1) SMAL models do not have textures and generating high quality textures for SMAL models with fitting algorithms like SMALR [5] requires a set of real images from comprehensive viewpoints which are hard to obtain in our case. (2) Table 1 shows training on synthetic tiger with random textures only has significant better performance. This implies the "real" textures provided with CAD models are not realistic enough and still have large domain gap with real tiger's and tiger-like animals' textures although they are visually more similar to real than random textures. For amazing performance achieved by random textures, we think one possible main reason is the domain randomization give the model better generalization ability to other domains. Another possible reason is that semantic parts of animals often have similar appearance which lessen the importance of realistic texture for part segmentation.

1.3. Visualization of SMAL Fitting Results and Synthetic Data

We have used more extreme viewpoints and larger range of camera distances than previous work [6,7] when rendering the synthetic data. Furthermore, we use transformations (translation and rotation) to produce more position variations to reduce the domain gap between real and synthetic. The visualization is shown below in Fig 2.

1.4. Visualization for Tail Ambiguity

The tail ambiguity refers to the difficulty of distinguishing between tails and horns in our model (shown in Fig 3). In our case, horns should belong to animal head class while our models often predict them as tail. This tail ambiguity problem makes our Class-Balanced Pseudo-Label Re-Weighting (CB) approach fail to be applied for tail class since we can not be confident enough in these tail predictions. As our synthetic data does not contain animals with horns, it also indicates that the part segmentation model currently lacks the capability to handle unseen parts.

References

[1] Lukas Hoyer, Dengxin Dai, and Luc Van Gool. Daformer: Improving network architectures and training strategies for domain-adaptive semantic segmentation, 2022. 1

[2] Lukas Hoyer, Dengxin Dai, and Luc Van Gool. Hrda: Context-aware high-resolution domain-adaptive semantic segmentation, 2022. 1

[3] Binhui Xie, Shuang Li, Mingjia Li, Chi Harold Liu, Gao Huang, and Guoren Wang. Sepico: Semantic-guided pixel contrast for domain adaptive semantic segmentation, 2023. 1

[4] Ju He, Shuo Yang, Shaokang Yang, Adam Kortylewski, Xiaoding Yuan, Jie-Neng Chen, Shuai Liu, Cheng Yang, Qihang Yu, and Alan Yuille. Partimagenet: A large, high-quality dataset of parts, 2022. 1

[5] Silvia Zuffi, Angjoo Kanazawa, and Michael J. Black. Lions and tigers and bears: Capturing non-rigid, 3D, articulated shape from images. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3955–3963. IEEE Computer Society, 2018. 1

[6] Jiteng Mu, Weichao Qiu, Gregory Hager, and Alan Yuille. Learning from synthetic animals, 2020. 1

[7] Qing Liu, Adam Kortylewski, Zhishuai Zhang, Zizhang Li, Mengqi Guo, Qihao Liu, Xiaoding Yuan, Jiteng Mu, Weichao Qiu, and Alan Yuille. Learning part segmentation through unsupervised domain adaptation from synthetic vehicles, 2022. 1

Table 1. **Ablation of Animal Texture.** The real textures refer to training on 5000 CAD synthetic tiger images with the real textures provided with CAD models. The random textures refer to training on 5000 CAD synthetic tiger images with random textures from real images. The test set are tiger-like set. Numbers are averaged over 3 random seeds.

method	texture		head	torso	leg	tail	bg	mIoU
	real	random						
SegFormer	✗	✓	71.44	46.94	35.01	22.29	83.86	51.91
	✓	✗	57.69	26.65	28.42	24.36	78.49	43.12

Figure 1. **Real Texture Provided in CAD Synthetic Data.**

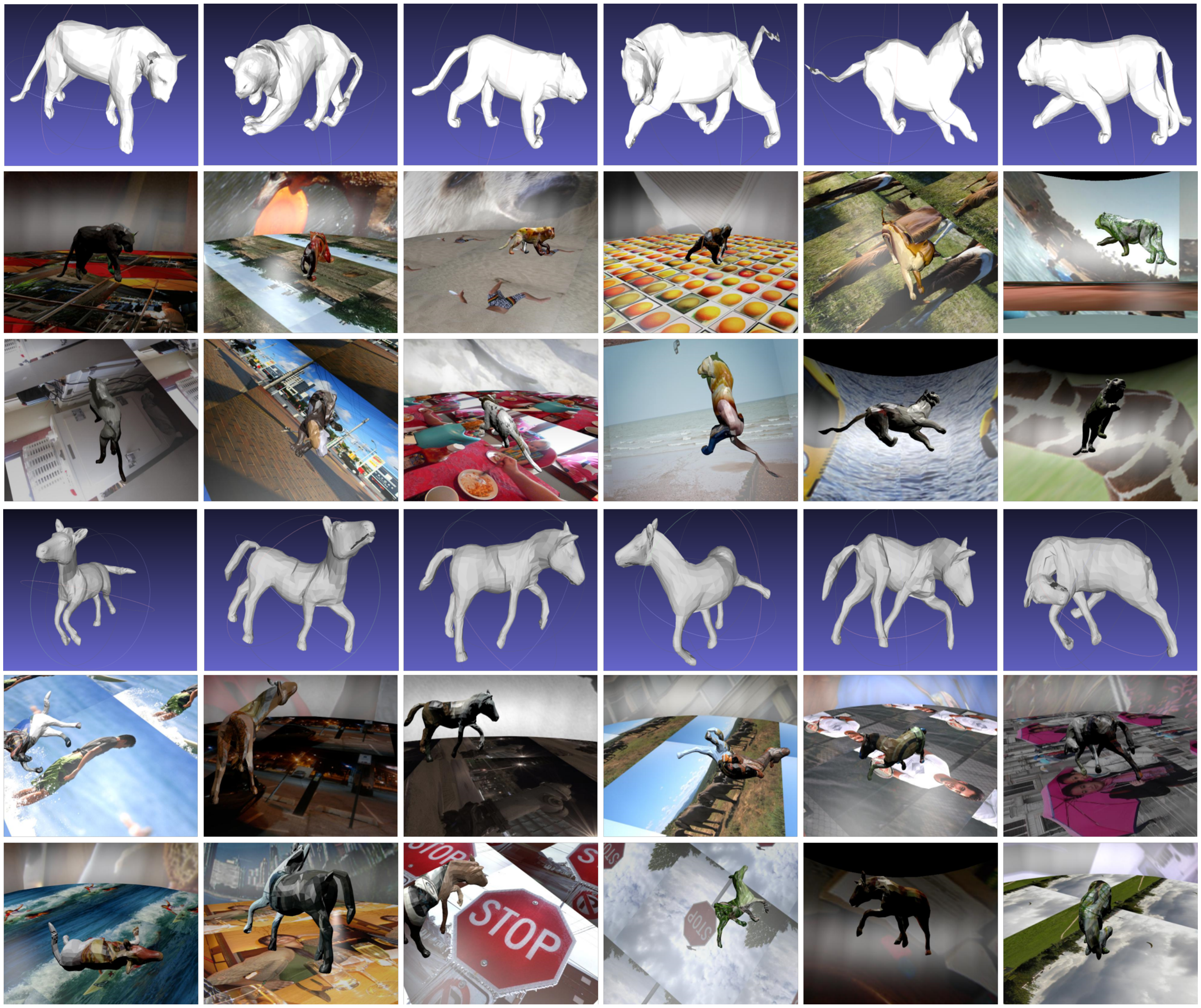

Figure 2. **More SMAL fitting results and their corresponding SMAL synthetic Data.**

Figure 3. **Failure Examples in Classification of Tails and Horns(belonging to Head Class).**