Common3D: Self-Supervised Learning of 3D Morphable Models for Common Objects in Neural Feature Space

Supplementary Material

Name	Value
Optimizer	Adam
Learning Rate	1.00E-04
Batch Size	12
Batch Accumulation	2
Losses	
λ_{app}	1.00E-01
λ_{CD}	1.00E-01
λ_{m}	1.00E+00
λ_{mdt}	1.00E+02
λ_{sdf}	1.00E-02
λ_{def}	1.00E-01
λ_{def-sm}	1.00E-02
κ	14.3
Vertex Sampling	150
Mesh Model	
Tetrahedron Grid Size	16
Instance Shape Encoder	
ResNet Blocks	4
ResNet Block Types	bottleneck
Out Dimensions	[256, 256, 256, 256]
Strides	[2, 2, 2, 2]
Pre-Upsampling	[1, 1, 1]
Affine Transformation Field	
Layers	5
Hidden Dimension	256
Out Dimension	6
Feature Adapter	
ResNet Blocks	4
ResNet Block Types	bottleneck
Out Dimensions	[512, 512, 128]
Strides	[1, 1, 1]
Pre-Upsampling	[1, 1, 2]
SDF Field	
Layers	5
Hidden Dimension	256
Out Dimension	1
Feature Field	
Layers	5
Hidden Dimension	256
Out Dimension	128

Table 7. The following parameters are set for training and testing our model.

6. Parameters

In Tab. 7 we list our model's parameters and their values.

We set σ as the average nearest neighbor distance for all vertices V to the sampled vertices V′, as follows

$$\sigma = \sqrt{\frac{1}{|\mathrm{V}|}\sum_{\mathrm{v}\in\mathrm{V}} \min_{\mathrm{v}'\in\mathrm{V}'} ||\mathrm{v}-\mathrm{v}'||^2}. \tag{14}$$

7. Comprehensive Results

We report all missing categorical quantitative results in Tabs. 8 to 10. Additional qualitative results are shown for ObjectNet3D in Fig. 4 and for SPair-71k in Fig. 6.

	Method	bicycle	bus	car	chair	couch	motorcycle	tv	AVG
Sup.	StarMap [83]	83.2	94.4	90	75.4	79.8	68.8	85.8	82.5
	VoGE [63]	82.6	98.1	99	90.5	94.9	87.5	83.9	90.9
Unsup.	ZSP [14]	61.7	21.4	61.6	42.6	52.9	43.1	39	46
	UOP3D [52]	58.4	79.3	98.2	51.9	76.6	67	**53.1**	69.2
	Ours	**67.2**	**82.9**	**99.3**	**67.8**	**80.9**	**78.1**	51.1	**75.3**

Table 8. **Comprehensive Evaluation** object 3D pose on PASCAl3D+. We report the 30° accuracy for 7 categories.

	b'pack	bench	b'cycle	bus	car	phone	chair	couch	cup	h'dryer	
ZSP [14]	23.1	50.8	58.6	30.5	60.3	46.4	36.8	55.5	33	**21.7**	
UOP3D [52]	18	62.1	57.8	78.3	98.1	**54.6**	52.2	76.6	38.2	14.1	
Ours	**25.5**	**77**	**65.7**	**81.1**	**99**	51.8	**65.9**	**79**	**44.7**	19.4	
	k'board	**laptop**	**m'wave**	**m'cycle**	**mouse**	**remote**	**suitcase**	**toaster**	**toilet**	**tv**	**AVG**
ZSP [14]	**46.8**	**60.5**	50.5	50.3	28.8	41.6	**25.8**	28.8	56.3	37.9	42.2
UOP3D [52]	26.9	53.3	**80.3**	69	**44.7**	**54.4**	15.5	**60.6**	39.6	**53.2**	52.4
Ours	34.5	54.2	76.1	**81**	38.5	52.4	15.7	57.3	**65**	52.4	**56.8**

Table 9. **Comprehensive evaluation** of object 3D pose on ObjectNet3D. We report the 30° accuracy for 20 categories.

	b'pack	bench	b'cycle	bus	car	phone	chair	couch	cup	h'dryer	
UOP3D [52]	68.9	49.9	27.9	**66.8**	79.2	67.7	42.9	73	**69.6**	50.9	
Ours	**69.4**	**54.1**	**31.5**	64.9	**82.8**	**68.1**	**53.3**	**77.8**	66.8	**53.6**	
	k'board	**laptop**	**m'wave**	**m'cycle**	**mouse**	**remote**	**suitcase**	**toaster**	**toilet**	**tv**	**AVG**
UOP3D [52]	76.9	62.4	**75.3**	64.3	**64.1**	61.4	62.5	77.9	64.7	73.8	64
Ours	**83.1**	**70.6**	75.2	**66.7**	62.7	**71.7**	**65.3**	**78.2**	**66.7**	**78.3**	**67**

Table 10. **Comprehensive evaluation** of instance segmentation on ObjectNet3D. We report the IoU for 20 categories.

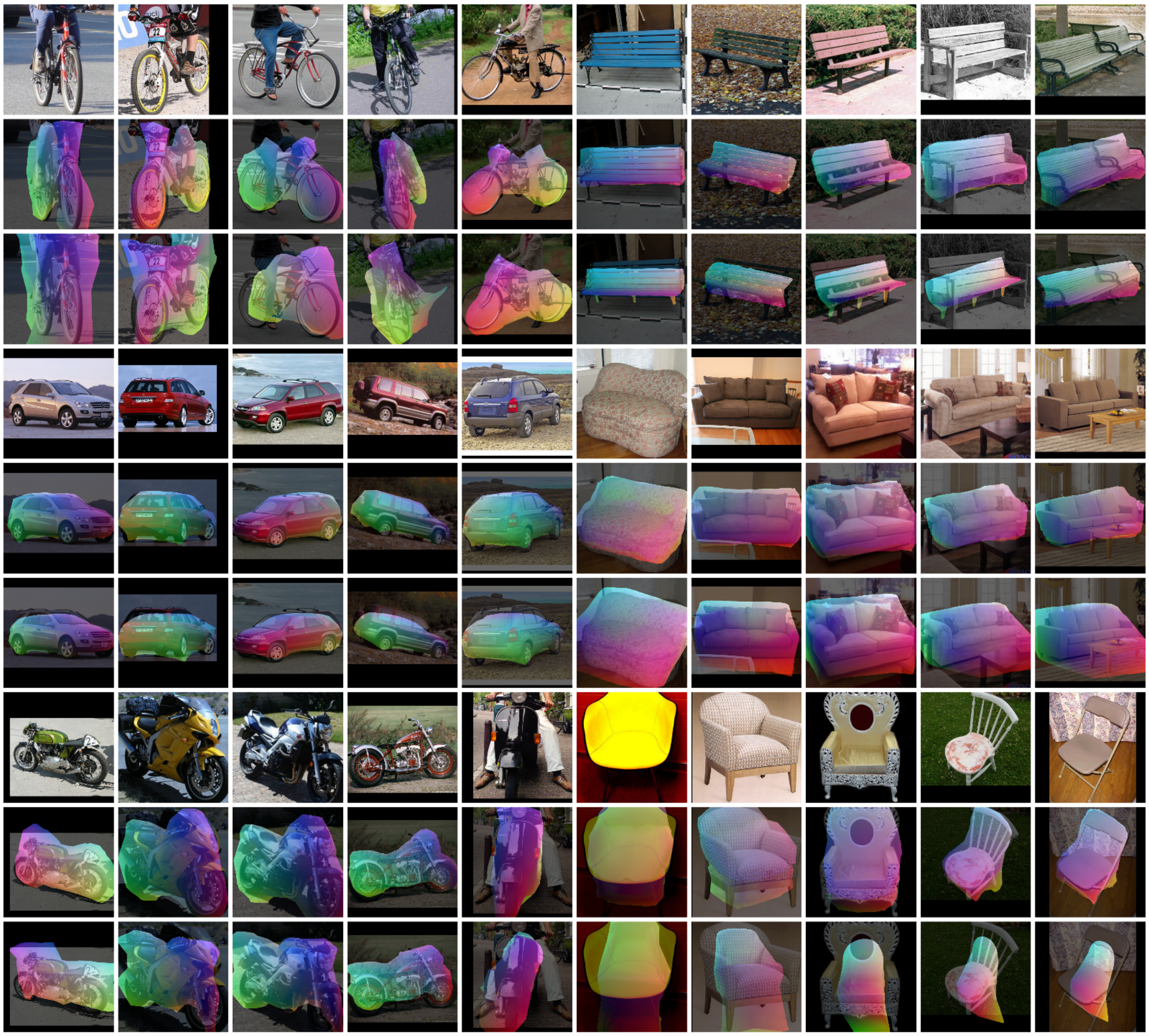

Figure 4. **Comprehensive qualitative results** on the ObjectNet3D dataset. In the second row the results of our method are illustrated, in the third the results of UOP3D [52]. Notably, our method fits the object more accurately, resulting in improved 3D pose and segmentation accuracy.

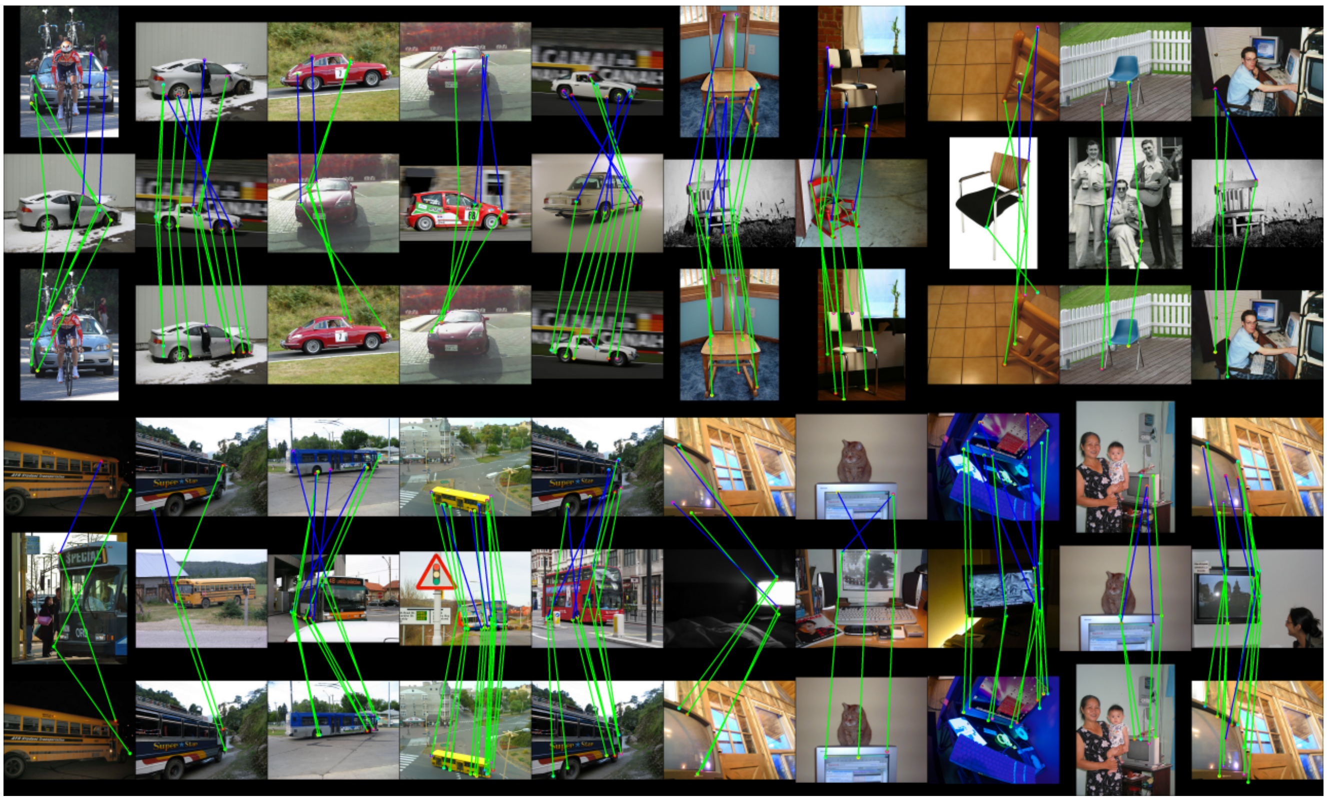

Figure 5. **Comprehensive qualitative results** on the SPair-71k dataset. In the first row the results of DINOv2 are illustrated, in the third the results of our method. Our method can improve DINOv2 correspondences by resolving ambiguities in parts and symmetries.

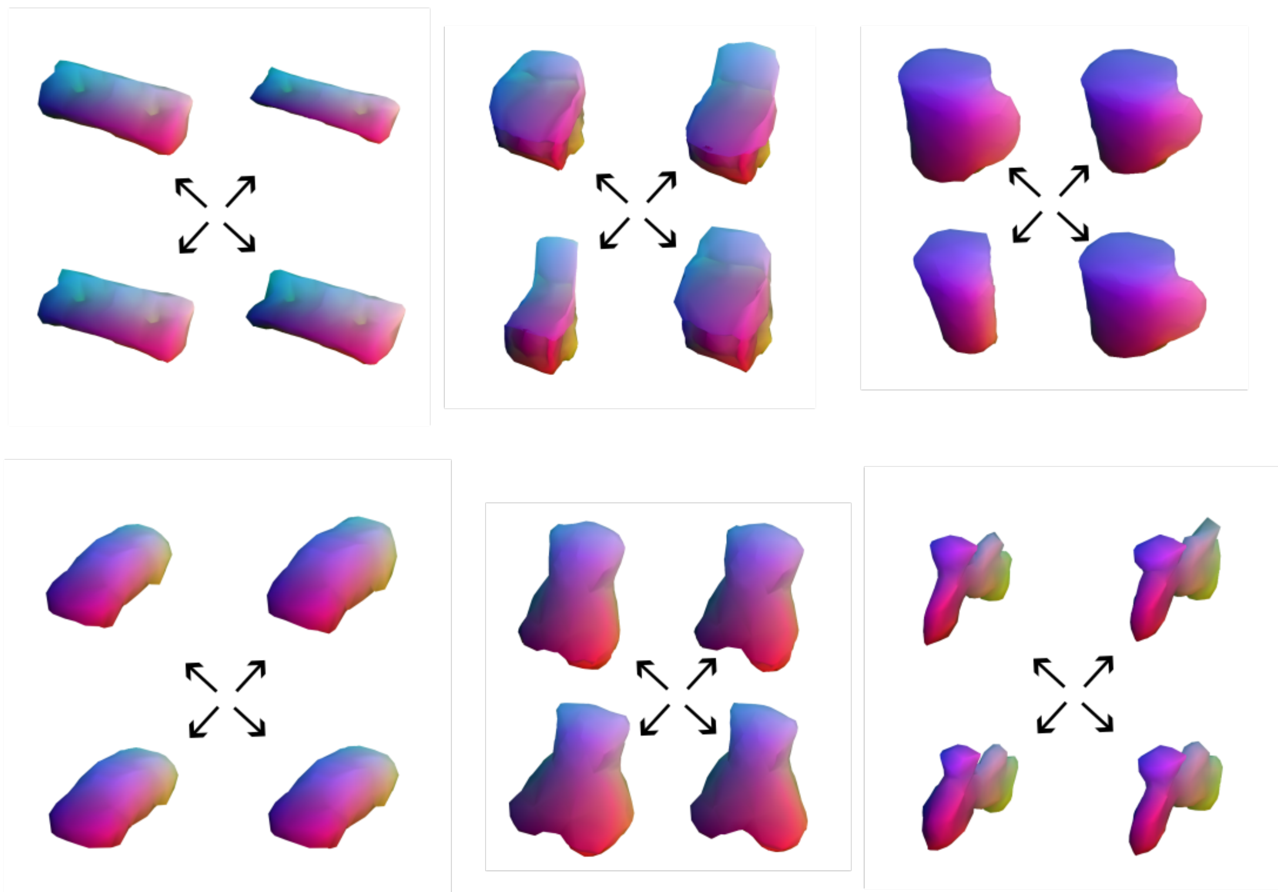

Figure 6. **3D morphable models with their latent deformations.**