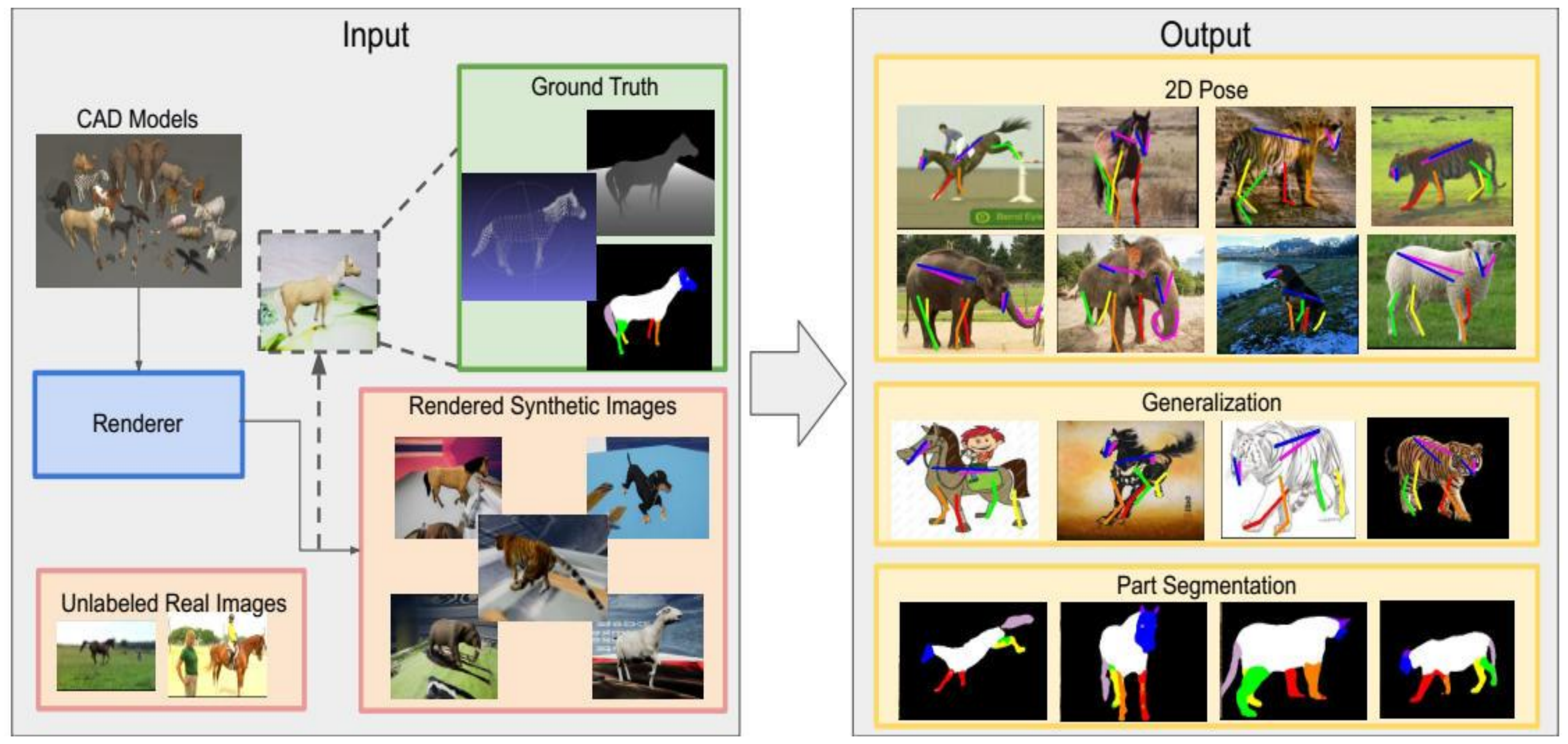
Learning from Synthetic Animals

Jiteng Mu*, Weichao Qiu*, Gregory D. Hager, Alan Yuille
Johns Hopkins University

HIGHLIGHTS

Motivation:
✦Parsing animals is important to our ecosystems and society.
✦Few large-scale animal annotated datasets.
✦Can we train animal parsing models using synthetic data which can generalize well on real world images?

Challenges:
✦Large domain gap between synthetic data and real data.
✦Synthetic data is limited by object diversity.



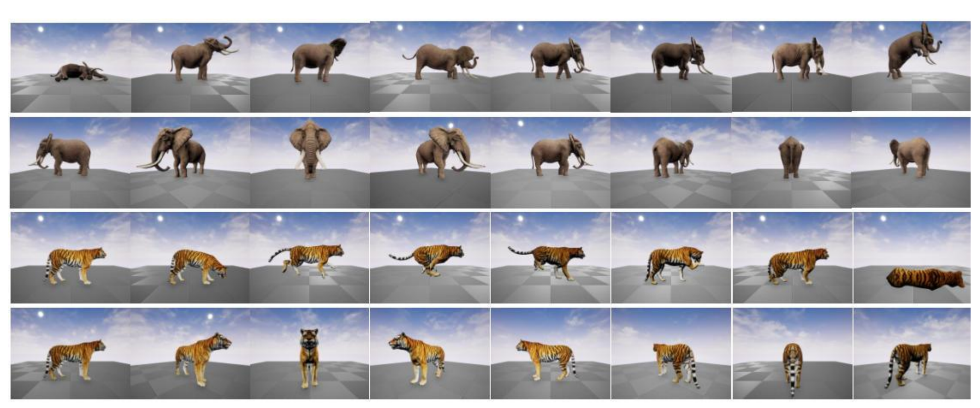
CONTRIBUTIONS

✦ We propose a consistency-constrained semi-supervised learning framework (CC-SSL) to learn a model with one single CAD object.
✦ We show that models trained jointly on synthetic and real images achieve better results compared to models trained only on real images when real image labels are available.
✦ We quantitatively demonstrate that models trained using synthetic data show better generalization performance than models trained on real-world images.
✦ We generate an animal dataset with 10+ different animal CAD models.

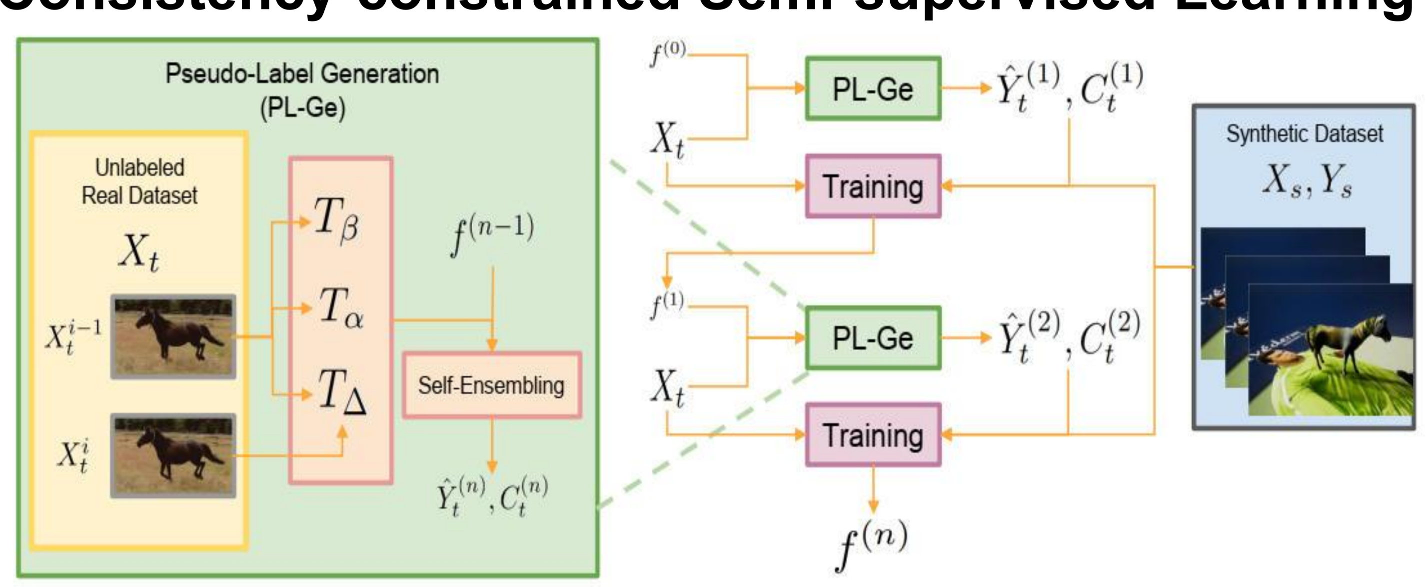
PROPOSED METHODS

Synthetic data generation

We randomize
✦ Textures
✦ Lighting
✦ Viewpoints
✦ Poses
✦ Background



Consistency-constrained Semi-supervised Learning



✦ Invariance Consistency: $f[T_\beta(X)] = f(X)$

✦ Equivariance Consistency: $f[T_\alpha(X)] = T_\alpha[f(X)]$

✦ Temporal Consistency: $f[T_\Delta(X)] = f(X) + \Delta$

✦ Loss for the *n*th iteration:

$$L^{(n)} = \sum_i^A L_{MSE}(f^{(n)}(X_s^i), Y_s^i) + \gamma \sum_j^B L_{MSE}(f^{(n)}(X_t^j), \hat{Y}_t^{(n-1),j})$$

2D POSE ESTIMATION

Quantitative results are evaluated on the TigDog dataset for both horses and tigers.

Quantitative Results

	Horse Accuracy								Tiger Accuracy							
	Eye	Chin	Shoulder	Hip	Elbow	Knee	Hoove	Mean	Eye	Chin	Shoulder	Hip	Elbow	Knee	Hoove	Mean
synthetic + real																
Real	79.04	89.71	**71.38**	91.78	82.85	80.80	72.76	78.98	**96.77**	93.68	65.90	94.99	67.64	80.25	81.72	81.99
CC-SSL-R	**89.39**	**92.01**	69.05	**92.28**	**86.39**	**83.72**	**76.89**	**82.43**	95.72	**96.32**	**74.41**	**91.64**	**71.25**	**82.37**	**82.73**	**84.00**
synthetic only																
Syn	46.08	53.86	20.46	32.53	20.20	24.20	17.45	25.33	23.45	27.88	14.26	52.99	17.32	16.27	19.29	21.17
CycleGAN [45]	70.73	84.46	56.97	69.30	52.94	49.91	35.95	51.86	71.80	62.49	29.77	61.22	36.16	37.48	40.59	46.47
BDL [26]	74.37	86.53	64.43	75.65	63.04	60.18	51.96	62.33	77.46	65.28	36.23	62.33	35.81	45.95	54.39	52.26
CyCADA [16]	67.57	84.77	56.92	76.75	55.47	48.72	43.08	55.57	75.17	69.64	35.04	65.41	38.40	42.89	48.90	51.48
CC-SSL	**84.60**	**90.26**	**69.69**	**85.89**	**68.58**	**68.73**	**61.33**	**70.77**	**96.75**	**90.46**	**44.84**	**77.61**	**55.82**	**42.85**	**64.55**	**64.14**

Qualitative Results

GENERALIZATION

Quantitative results are evaluated on different domains in the VisDA2019 dataset.

Quantitative Results

	Horse						Tiger					
	Visible Kpts Accuracy			Full Kpts Accuracy			Visible Kpts Accuracy			Full Kpts Accuracy		
	Sketch	Painting	Clipart	Sketch	Painting	Clipart	Sketch	Painting	Clipart	Sketch	Painting	Clipart
Real	65.37	64.45	64.43	61.28	58.19	60.49	48.10	61.48	53.36	46.23	53.14	50.92
CC-SSL	72.29	73.71	**73.47**	**70.31**	**71.56**	**72.24**	53.34	55.78	59.34	52.64	48.42	54.66
CC-SSL-R	**73.25**	**74.56**	71.78	67.82	65.15	65.87	**54.94**	**68.12**	**63.47**	**53.43**	**58.66**	**59.29**

Qualitative Results

REFERENCES

[16] Judy Hoffman et al. Cycada: Cycle-consistent adversarial domain adaptation. In ICML, 2018.
[26] Yunsheng Li et al. Bidirectional learning for domain adaptation of semantic segmentation. In CVPR, 2019.
[45] Jun-Yan Zhu et al. Unpaired image-to-image translation using cycle consistent adversarial networks. In ICCV, 2017.