

Deep Semantic Matching with Foreground Detection and Cycle-Consistency

Yun-Chun Chen^{1,2} Po-Hsiang Huang² Li-Yu Yu² Jia-Bin Huang³ Ming-Hsuan Yang^{4,5} Yen-Yu Lin¹

¹Academia Sinica ²National Taiwan University ³Virginia Tech

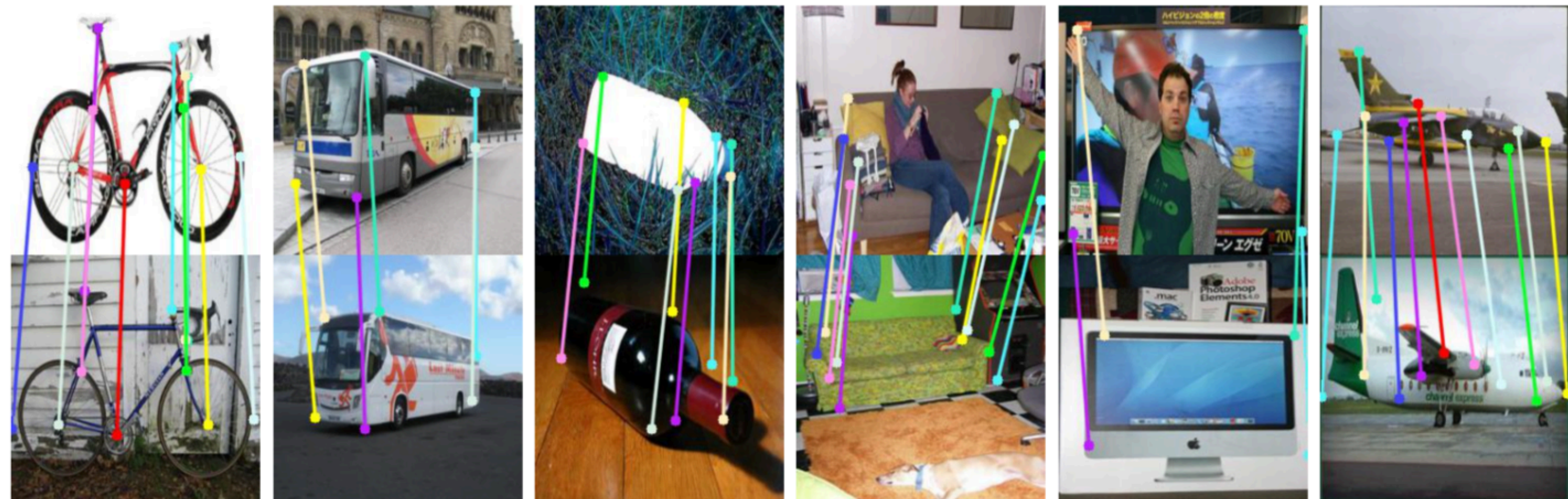
⁴University of California, Merced ⁵Google Cloud

Code available!



Semantic matching

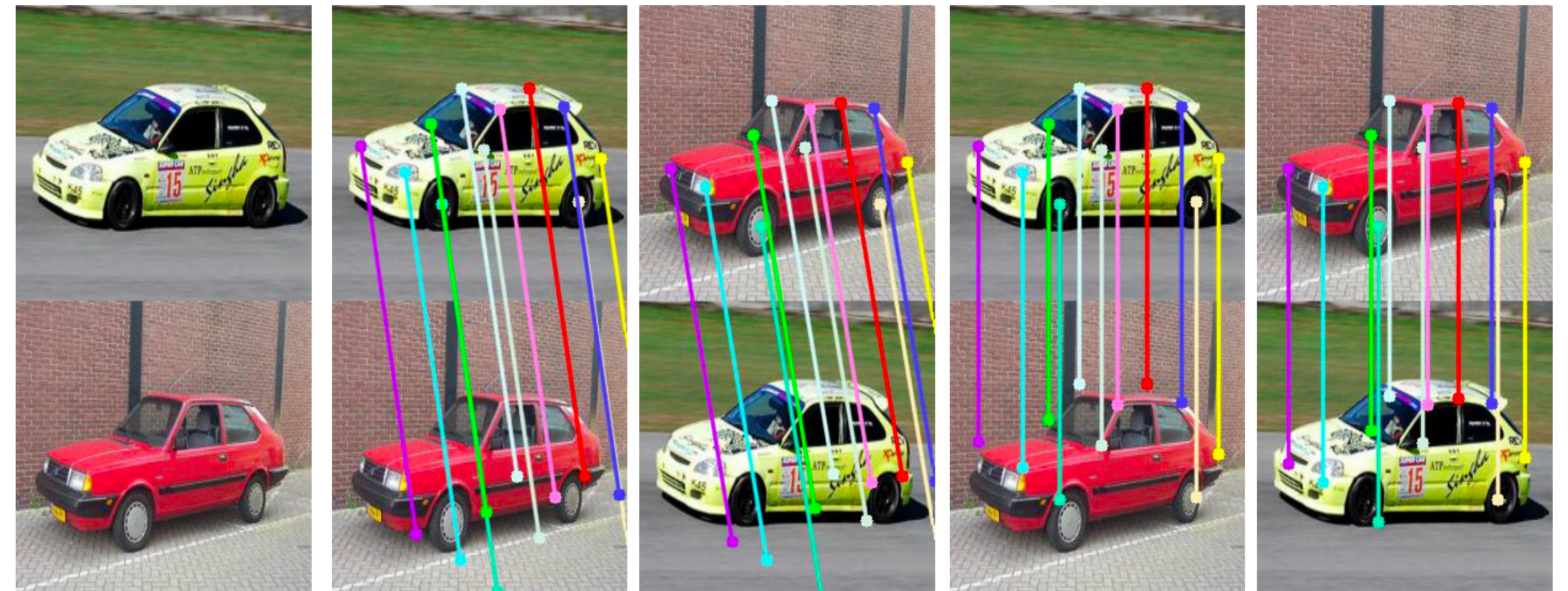
- **Input:** A set of images containing instances of an object category
- **Goal:** Estimate correspondences between image pairs



Our matching results.

Challenges

- Background clutters
- Inconsistent bidirectional matching
- Require keypoint annotation for supervision

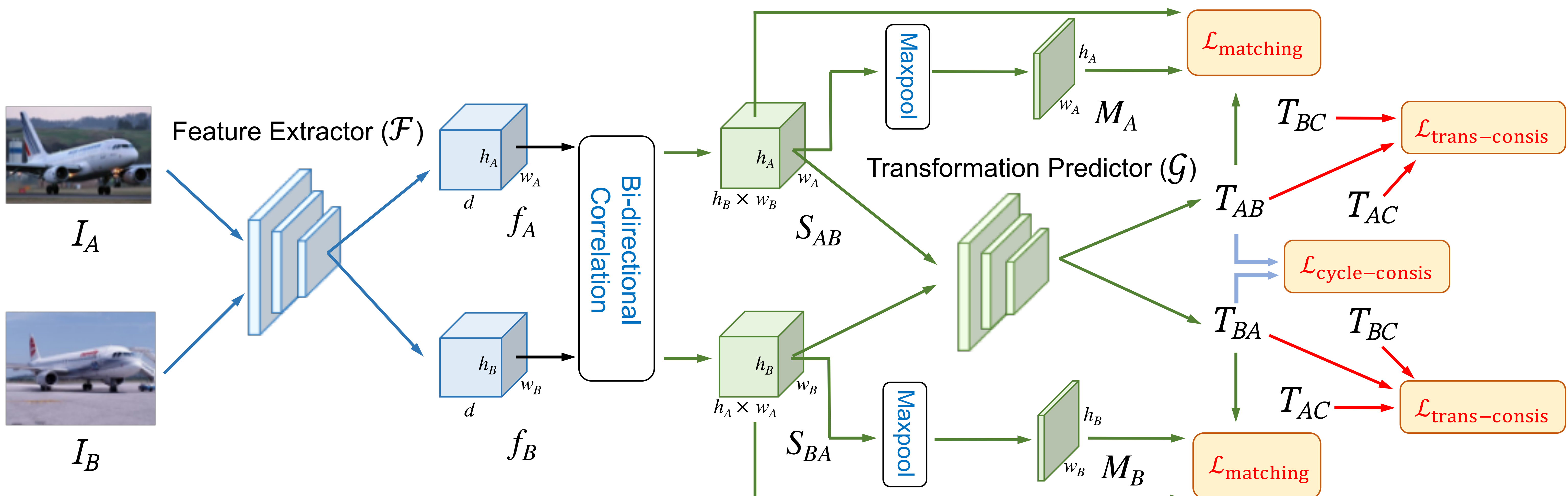


Input

Rocco *et al.*

Ours

Weakly-supervised semantic matching network



Experimental results

(a) Results on PF-PASCAL.

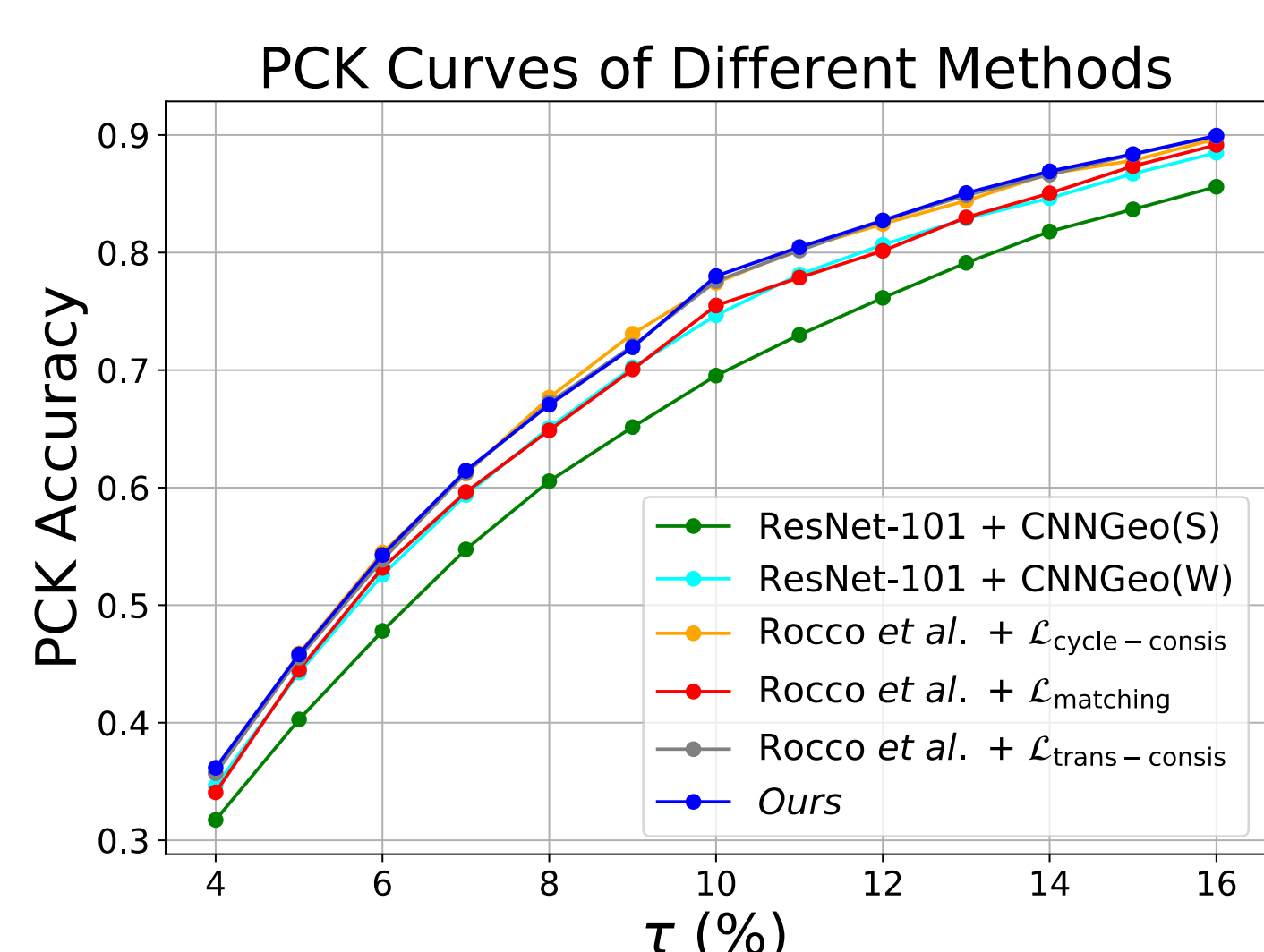
Method	mean PCK
HOG+PF-LOM	62.5
UCN	55.6
VGG-16+SCNet-A	66.3
VGG-16+SCNet-AG	69.7
VGG-16+SCNet-AG+	72.2
VGG-16+CNNGeo	62.6
ResNet-101+CNNGeo(S)	69.5
ResNet-101+CNNGeo(W)	74.8
Ours	78.0

(b) Results on PF-WILLOW.

Method	$\tau = 0.05$	$\tau = 0.1$	$\tau = 0.15$
SIFT Flow	0.247	0.380	0.504
VGG w/SF	0.324	0.456	0.555
FCSS w/SF	0.354	0.532	0.681
UCN	0.291	0.417	0.513
SCNet-A	0.390	0.725	0.873
SCNet-AG	0.394	0.721	0.871
SCNet-AG+	0.386	0.704	0.853
ResNet-101+CNNGeo(S)	0.448	0.777	0.899
ResNet-101+CNNGeo(W)	0.477	0.812	0.917
Ours	0.491	0.819	0.922

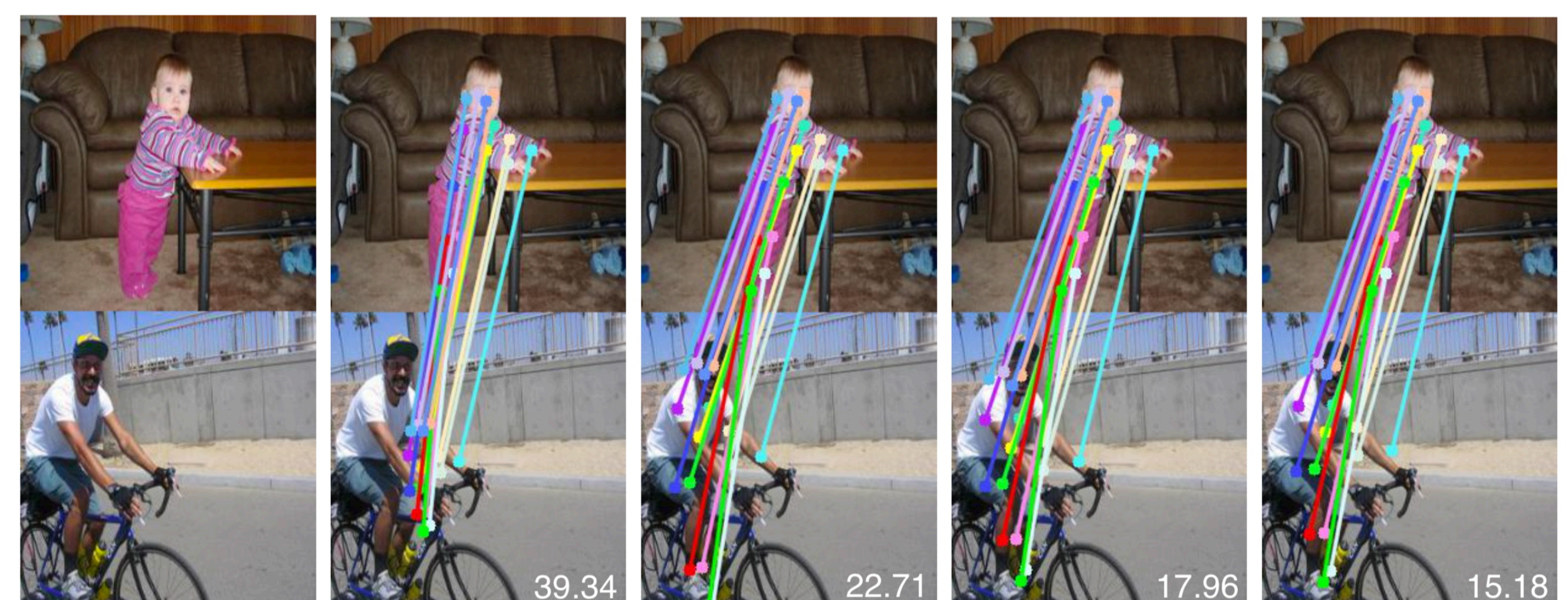
(c) Results on TSS.

Method	FG3DCar	JODS	PASCAL	Avg.
HOG+PF-LOM	0.786	0.653	0.531	0.657
HOG+TSS	0.830	0.595	0.483	0.636
FCSS+SIFT Flow	0.830	0.656	0.494	0.660
FCSS+PF-LOM	0.839	0.635	0.582	0.685
HOG+OADSC*	0.875	0.708	0.729	0.771
FCSS+DCTM	0.891	0.721	0.610	0.740
VGG-16+CNNGeo	0.835	0.656	0.527	0.673
ResNet-101+CNNGeo(S)	0.886	0.758	0.560	0.735
ResNet-101+CNNGeo(W)	0.892	0.758	0.562	0.737
Ours	0.898	0.768	0.560	0.742



(d) Ablation study on PF-PASCAL.

Method	mean
Rocco <i>et al.</i>	74.8
Rocco <i>et al.</i> + $\mathcal{L}_{\text{matching}}$	75.5
Rocco <i>et al.</i> + $\mathcal{L}_{\text{cycle-consis}}$	77.4
Rocco <i>et al.</i> + $\mathcal{L}_{\text{trans-consis}}$	77.6
Ours	78.0



An image pair

Rocco *et al.*

Rocco *et al.*
+ foreground
detection

Rocco *et al.*
+ cycle
consistency

Ours

Ground Truth Label		Ground Truth Label		Accuracy		Variance	
Warped Label	Foreground	82.3 %	29.1 %	79.1 %	81.2 %	10.3 %	6.73 %
Warped Label	Background	17.7 %	70.9 %				
		Foreground	Background	Baseline	Ours		
Rocco <i>et al.</i>		Ours		Comparison			

(e) Effect of using the foreground-guided matching loss $\mathcal{L}_{\text{matching}}$.

References:

- Choy et al. Universal Correspondence Network. In NIPS, 2016.
- Han et al. SCNet: Learning Semantic Correspondence. In ICCV, 2017.
- Ham et al. Proposal Flow: Semantic Correspondences from Object Proposals. TPAMI, 2017.
- Rocco et al. Convolutional Neural Network Architecture for Geometric Matching. In CVPR, 2017.
- Rocco et al. End-to-End Weakly-Supervised Semantic Alignment. In CVPR, 2018.