Leveraging Semi-Supervised Learning in Video Sequences for Urban Scene Segmentation – Supplementary Material –

Anonymous ECCV submission

Paper ID 942

Pseudo-Labeled Video Sequences: We visualize the generated pseudo-labeled sequences at iteration 1 and iteration 2 for a short video sequence from train-sequence set. For each second, we visualize the input image (left), pseudo-labels at iteration 1 (Teacher with X-71), and pseudo-labels at iteration 2 (Teacher with WR-41). We also overlay the pseudo-labels with images.

Architecture of proposed WR-41: In Tab. 1, we provide the architecture details of WR-38 [1] and our proposed WR-41.

layer name	input size	output size	WR-38 [1]	WR-41
conv1	224×224	224×224	3×3, 64	
B2	224×224	112×112	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times3$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times3$
B3	112×112	56×56	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times3$
B4	56×56	28×28	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times6$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times6$
B5	28×28	14×14	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 1024 \end{bmatrix}\times3$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 1024 \end{bmatrix}\times3$
B6	14×14	7×7	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 1024 \\ 1\times1, 2048 \end{bmatrix}\times1$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 1024 \\ 1\times1, 2048 \end{bmatrix}\times3$
B7	7×7	7×7	$\begin{bmatrix} 1\times1, 1024 \\ 3\times3, 2048 \\ 1\times1, 4096 \end{bmatrix}\times1$	
	7×7	1×1	average pool, 1000-d fc, softmax	
Params (M)			109	111
M-Adds (B)			46	49.3

Table 1. Architectures for WR-38 [1] and our proposed WR-41 on ImageNet. The M-Adds are computed w.r.t. an 224 × 224 input.

References

1. Wu, Z., Shen, C., Van Den Hengel, A.: Wider or deeper: Revisiting the resnet model for visual recognition. Pattern Recognition (2019) 1