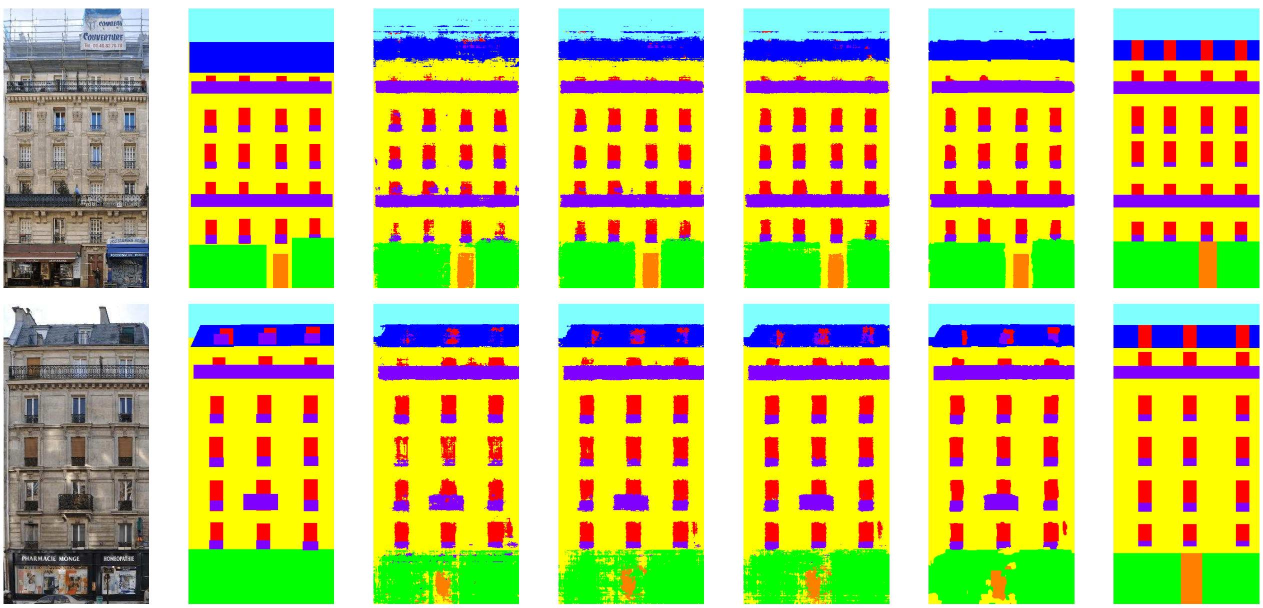
Efficient Facade Segmentation using Auto-Context
Supplementary Material

Varun Jampani*
MPI for Intelligent Systems
Tübingen, Germany
varun.jampani@tuebingen.mpg.de

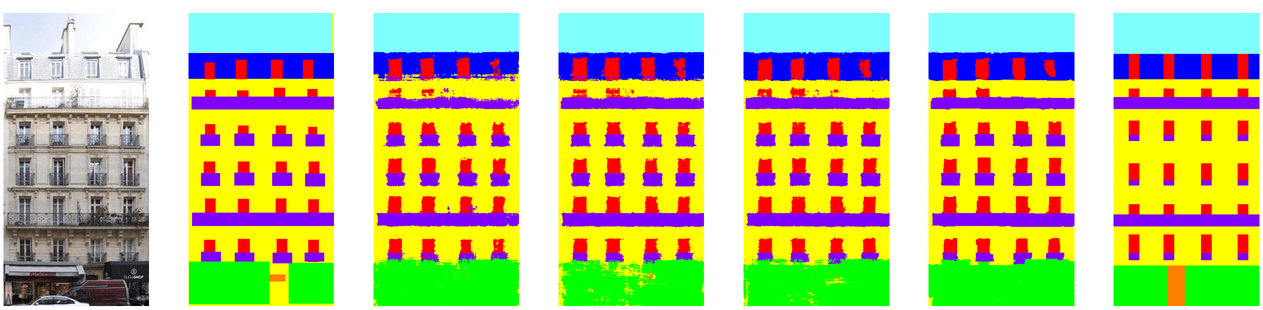
Raghudeep Gadde*
Université Paris-Est, LIGM
ENPC, ECP, Paris, France
gadder@imagine.enpc.fr

Peter V. Gehler
MPI for Intelligent Systems
Tübingen, Germany
pgehler@tuebingen.mpg.de

We present more qualitative (Figures 1 to 6) and quantitative (Table 1 and Table 2) facade segmentation results. The images have been selected based on the absolute overall pixel-accuracy of ST3 and include images with the (i) highest, (ii) average, and (iii) lowest performance.



(a) Facade (b) GT (c) ST1 (d) ST2 (e) ST3 (f) PW3 (g) Parse



(a) Facade (b) GT (c) ST1 (d) ST2 (e) ST3 (f) PW3 (g) Parse

Figure 1. (a) Sample facade images from ECP dataset, (b) ground truth segmentation, (c,d,e) result of various classification stages (ST1, ST2, ST3) of the auto-context method, (f) Potts model using ST3 as unaries, and (g) the result obtained by applying reinforcement learning [8] using the output of ST3.

*The first two authors contribute equally to this work.

(a) CMP Dataset

Class	[6]	Auto Context ST1	Auto Context ST2	Auto Context ST3	AC + Potts Model PW1	AC + Potts Model PW2	AC + Potts Model PW3
Background	58	67.1	71.8	**72.6**	68.0	**72.6**	73.1
Facade	73	74.6	75.3	75.2	**80.5**	79.9	79.3
Window	61	71.6	76.1	77.0	74.1	77.4	**78.1**
Door	**54**	37.9	45.5	47.0	39.6	46.4	48.7
Cornice	41	39.1	47.5	**49.6**	40.0	48.3	**50.1**
Sill	27	21.1	32.8	**36.2**	16.9	30.3	34.6
Balcony	46	31.6	44.1	46.7	31.6	45.2	**48.1**
Blind	**48**	22.7	35.8	40.1	19.5	34.7	39.9
Deco	**24**	10.4	13	13.8	6.1	10.0	11.4
Molding	54	63.2	65.4	66.5	64.2	66.0	**67.2**
Pillar	**25**	5.71	11.2	13.6	1.33	7.72	9.78
Shop	**59**	40.9	45.6	45.6	42.8	46.7	46.8
Average	47.5	40.50	47.00	**48.65**	40.38	47.1	**48.92**
Overall	60.3	61.83	65.47	66.24	64.46	67.48	**68.08**
IoU	-	29.26	34.46	35.86	30.67	36.02	**37.47**

(b) Graz Dataset

Class	Auto Context (AC) ST1	Auto Context (AC) ST2	Auto Context (AC) ST3	AC + Potts Model PW1	AC + Potts Model PW2	AC + Potts Model PW3	[7]
Door	57.3	62.4	**62.7**	57.3	**62.8**	**63**	41
Window	78.2	81.2	**81.5**	77.8	80.6	80.9	60
Wall	94.9	94.7	94.9	**95.8**	**95.6**	**95.8**	84
Sky	87.4	**91.2**	**90.5**	87.7	**91.4**	**90.6**	**91**
Average	79.47	82.40	82.42	79.65	**82.61**	**82.56**	69
Overall	90.18	91.02	91.16	90.78	**91.53**	**91.68**	78
IoU	71.25	73.31	73.25	72.49	**74.45**	**74.39**	58

(c) labelMeFacades Dataset

Class	[2]	[5]	Auto Context(AC) ST1	Auto Context(AC) ST2	Auto Context(AC) ST3	AC + Potts Model PW1	AC + Potts Model PW2	AC + Potts Model PW3
Building	-	-	87.7	88.1	88.2	**92.7**	91.8	92.1
Car	-	-	47.1	53.6	54.8	51.1	57.0	**58.2**
Door	-	-	**6.52**	6.03	5.12	2.61	3.22	1.71
Pavement	-	-	24	**25.3**	**24.6**	22.0	24.2	23.3
Road	-	-	80.3	82.1	84.5	85.3	85.1	**87.6**
Sky	-	-	86.2	87.2	87.4	88.3	**88.6**	**88.9**
Vegetation	-	-	53.3	**57.5**	**57.6**	53.4	**58.1**	**57.9**
Window	-	-	20.3	22.6	**25.4**	13.0	16.9	19.5
Various	-	-	19.9	**20.6**	**21.0**	11.6	12.2	12.1
Average	**56.61**	-	47.26	49.22	49.84	46.68	48.56	49.04
Overall	67.33	71.28	71.52	72.9	73.46	74.1	74.62	**75.23**
IoU	-	35.96	37.01	38.69	39.36	37.74	38.96	**39.57**

Table 1. Segmentation results of various methods on CMP, Graz and labelmeFacades datasets. ST1, ST2, and ST3 correspond to the classification stages in the auto-context method. PW1, PW2, and PW3 refer to a Potts model using ST1, ST2, and ST3, respectively, as unaries. Published results are shown for comparisons. The method of [7] parses the image into a lattice representation and is not trying to maximize pixel accuracy results.

(a) eTRIMS Dataset

Class				Auto Context (AC)			AC + Potts Model		
	[4]	**[1]**	**[3]**	**ST1**	**ST2**	**ST3**	**PW1**	**PW2**	**PW3**
Building	91	91	84	90.3	90.5	90.9	**92.7**	**92.5**	**92.5**
Car	69	70	51	63.3	74.8	72.4	69.4	**79.1**	76.6
Door	18	18	**73**	62.7	62.3	63.6	66.0	63.6	65.3
Pavement	33	33	55	43.0	46.5	47.1	43.1	**48.6**	**48.8**
Road	55	57	81	78.2	82.3	80.3	80.9	**84.7**	82.1
Sky	93	97	**99**	97.6	**98.5**	**98.6**	98.2	**98.8**	**98.9**
Vegetation	89	90	92	91.1	92.1	92.3	92.4	**92.8**	**92.9**
Window	**74**	71	78	65.9	67.1	68.4	65.6	66.5	68.2
Average	65.3	65.9	66.4	74.01	76.78	76.7	76.04	**78.32**	78.14
Overall	83.16	83.84	83.40	84.68	85.95	86.12	86.39	**87.29**	**87.29**
IoU	-	-	-	58.7	61.26	61.48	61.49	63.39	**63.54**

Table 2. Segmentation results of various methods on eTRIMS dataset. ST1, ST2, and ST3 correspond to the classification stages in the auto-context method. PW1, PW2, and PW3 refer to a Potts model using ST1, ST2, and ST3, respectively, as unaries. Published results are shown for comparisons.

References

[1] A. Cohen, A. G. Schwing, and M. Pollefeys. Efficient structured parsing of facades using dynamic programming. In *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, pages 3206–3213. IEEE, 2014.

[2] B. Fröhlich, E. Rodner, and J. Denzler. Semantic segmentation with millions of features: Integrating multiple cues in a combined random forest approach. In *Asian Conference on Computer Vision (ACCV)*, pages 218–231. Springer, 2012.

[3] C. Gatta and F. Ciompi. Stacked sequential scale-space taylor context. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014.

[4] A. Martinovic, M. Mathias, J. Weissenberg, and L. Van Gool. A three-layered approach to facade parsing. In *European Conference on Computer Vision (ECCV)*, pages 416–429. Springer, 2012.

[5] S. Nowozin. Optimal decisions from probabilistic models: the intersection-over-union case. In *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, 2014.

[6] R. Š. Radim Tyleček. Spatial pattern templates for recognition of objects with regular structure. In *Proc. German Conference on Pattern Recognition*, Saarbrücken, Germany, 2013.

[7] H. Riemenschneider, U. Krispel, W. Thaller, M. Donoser, S. Havemann, D. Fellner, and H. Bischof. Irregular lattices for complex shape grammar facade parsing. In *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, pages 1640–1647, 2012.

[8] O. Teboul, I. Kokkinos, L. Simon, P. Koutsourakis, and N. Paragios. Shape grammar parsing via reinforcement learning. In *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, pages 2273–2280. IEEE, 2011.

Figure 2. Qualitative results on ECP dataset images along with overall pixel accuracy (Stage-3 results).

Figure 3. Qualitative results on Graz dataset images along with overall pixel accuracy (Stage-3 Results).

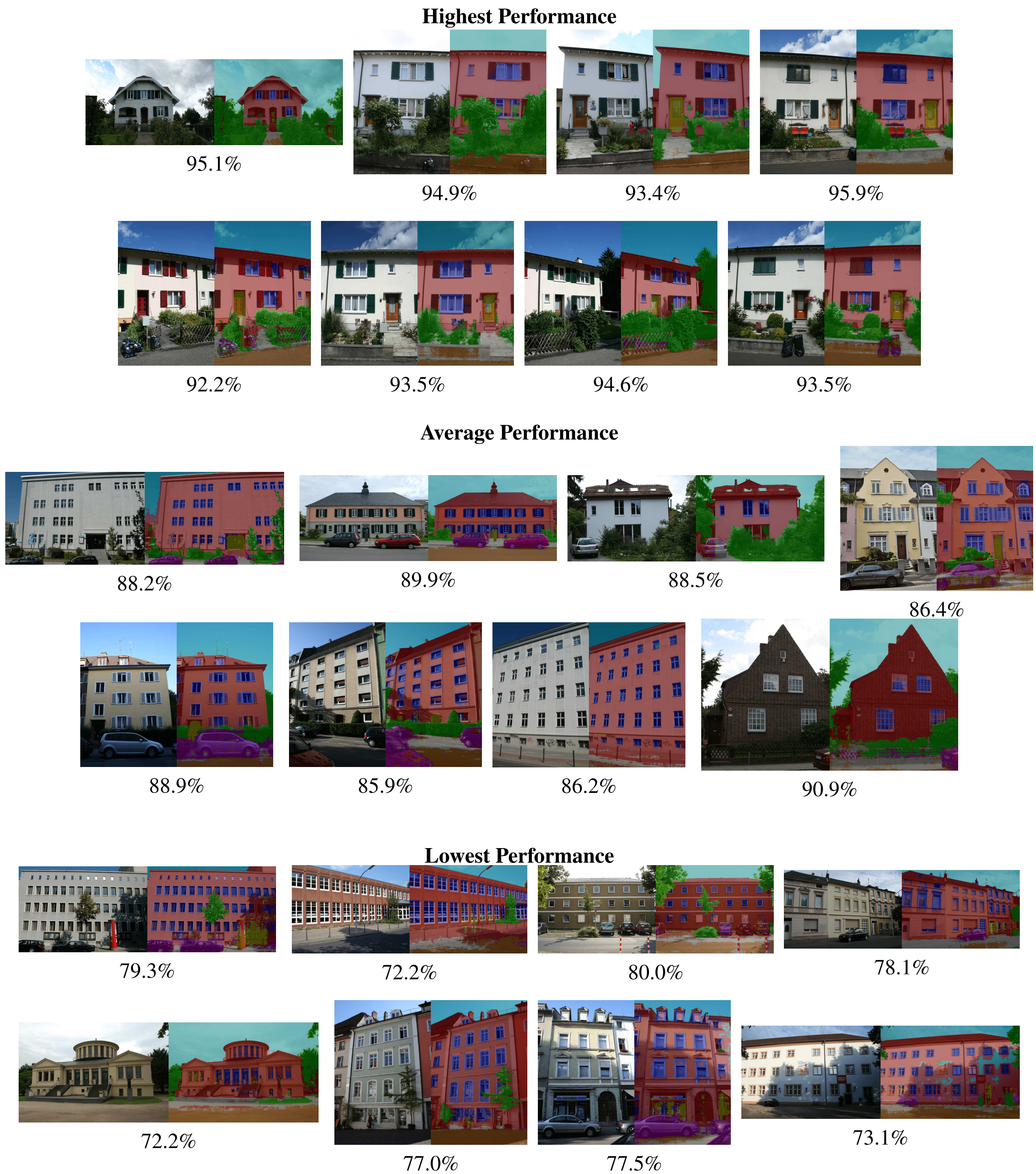


Figure 4. Qualitative results on eTRIMS dataset images along with overall pixel accuracy (Stage-3 results).

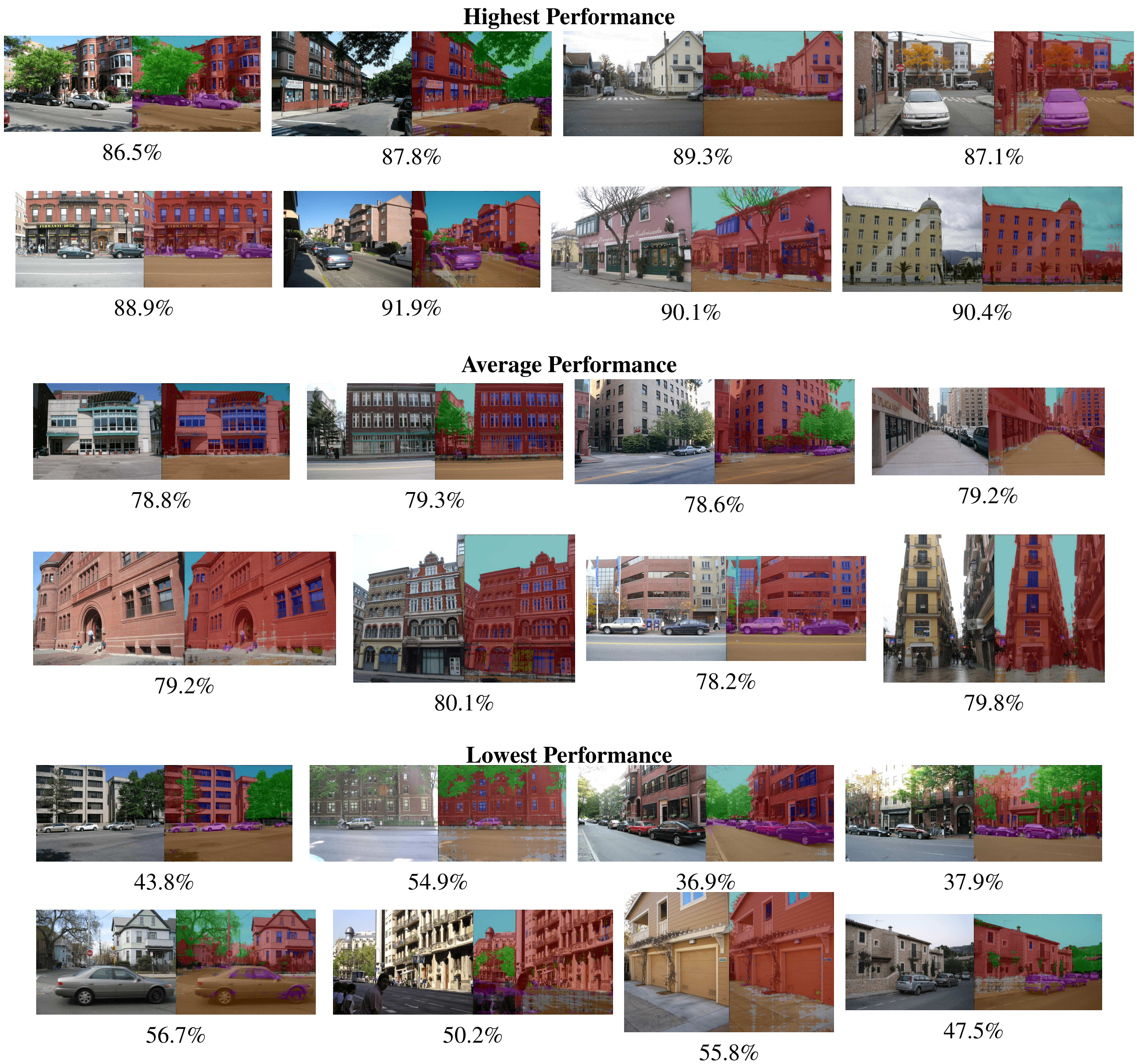


Figure 5. Qualitative results on labelmeFacades dataset images along with overall pixel accuracy (Stage-3 results).

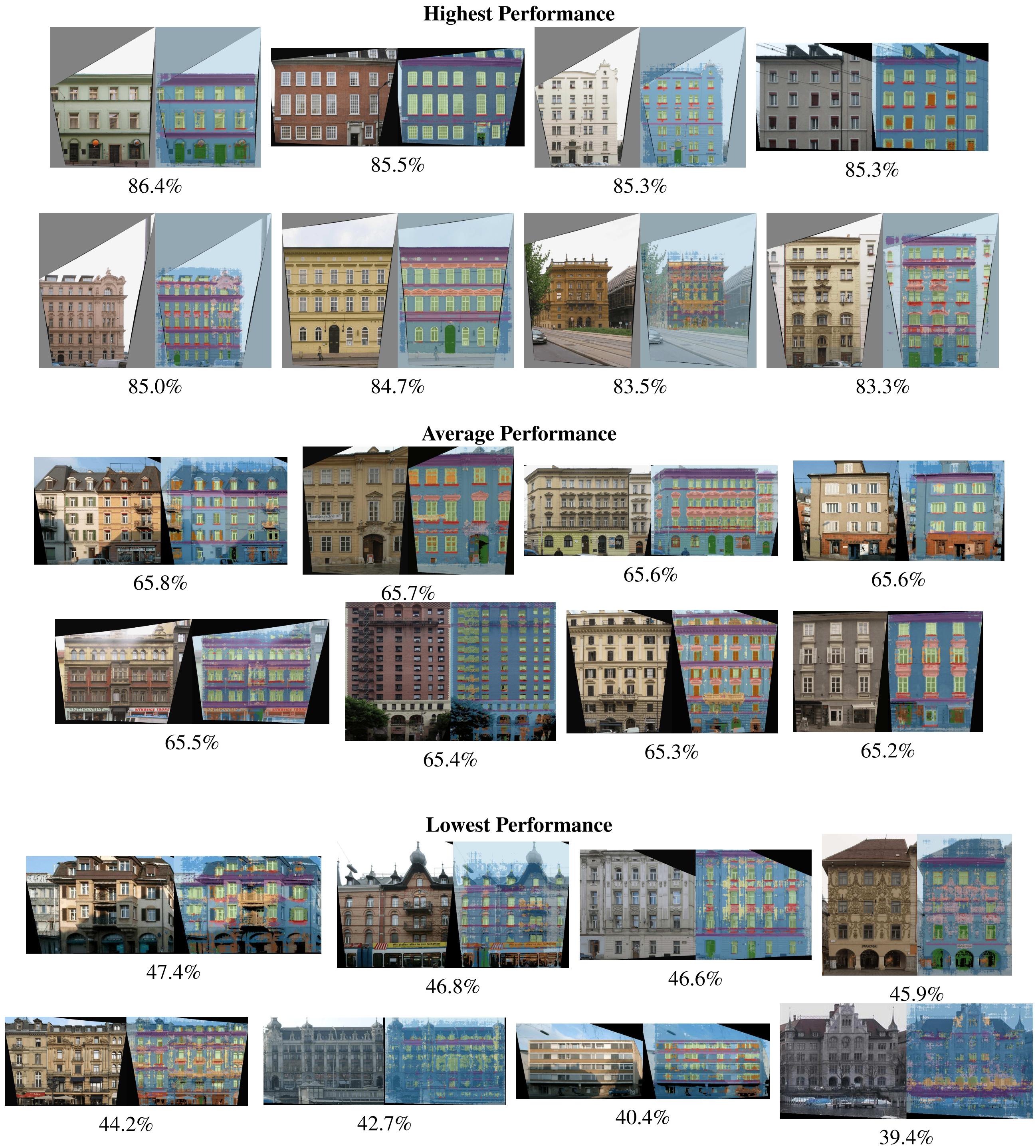


Figure 6. Qualitative results on CMP dataset images along with overall pixel accuracy (Stage-3 results).