Learning to Segment Under Various Forms of Weak Supervision
Supplementary Material

Jia Xu[1] Alexander G. Schwing[2] Raquel Urtasun[2]
[1]University of Wisconsin-Madison [2]University of Toronto
jiaxu@cs.wisc.edu {aschwing, urtasun}@cs.toronto.edu

1. Proof of Proposition 3.1

In this section we prove Proposition 3.1 in the main paper. We start by presenting some preliminaries that are necessary for the proof. We refer the reader to [3] for more details.

Definition 1.1. [3] A matrix A is totally unimodular (TU), iff the determinants of all square submatrices of A are either -1, 0, or 1.

Theorem 1.2. *[3] A $(0, +1, -1)$ matrix A is totally unimodular if both of the following conditions are satisfied:*

- *Each column contains at most two nonzero elements*
- *The rows of A can be partitioned into two sets A_1 and A_2 such that two nonzero entries in a column are in the same set of rows if they have different signs and in different sets of rows if they have the same sign.*

Corollary 1.3. *[3] A $(0, +1, -1)$ matrix A is totally unimodular if it contains no more than one $+1$ and no more than one -1 in each column.*

Theorem 1.4. *[4, 1, 2] If A is totally unimodular and b is integral, then solving linear programs of form $\{\min \mathbf{c}^T\mathbf{x} \mid A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq 0\}$ have integral optima, for any $\mathbf{c}$.*

The main idea of our proof is to show that the matrix describing our linear constraints is totally unimodular. Employing Theorem 1.4 we then know that the LP relaxation gives integral optima since the right hand side is integral in our optimization problem (Eq. (8) in the main paper). Given that our inference problem is fully decomposable with respect to images, we first decompose it into small LPs, one for each image. More formally, for each image i, we have,

$$\begin{aligned} \max_{H^i} \quad & \text{tr}\left((X^iW)^T H^i\right) \\ \text{s.t.} \quad & H^i \mathbf{1}_C = \mathbf{1}_{n^i} \\ & B'^T H^i \geq \mathbf{z}'^i \\ & 0 \leq H^i \leq B^i \mathbf{z}^i, \end{aligned} \tag{1}$$

where n^i is the number of super-pixels in image i and H^i is a binary label matrix for n^i super-pixels in image i. $B'^T H^i \geq \mathbf{z}'^i$ are the constraints from both the bounding boxes as well as tags. Note that for the semi-supervised case we remove the labeled super-pixels in the above LP. Additionally, the corresponding row c in $B'^T H^i \geq \mathbf{z}'^i$ is already satisfied (one instance is labeled) for class c by default, hence it is no longer required to be considered.

Given $0 \leq H^i \leq B^i\mathbf{z}^i$, those not-tagged classes will be filtered out in the final solution, i.e., $H^i_{pc} = 0$ if $z^i_c = 0$. We remove these classes and obtain the following LP:

$$\begin{aligned} \max_{H^i} \quad & \text{tr}\left((X^iW_{C'})^T H^i_{C'}\right) \\ \text{s.t.} \quad & H^i_{C'} \mathbf{1}_{C'} = \mathbf{1}_{n^i} \\ & \mathbf{1}^T_{n^i} H^i_{C'} \geq \mathbf{1}_{C'} \\ & H^i_{C'} \geq 0, \end{aligned} \tag{2}$$

where C' is the number of potential classes for the super-pixels in image i.

Next, let us rephrase our LP into the canonical form. We vectorize $H^i_{C'}$ by stacking each row into $\mathbf{x} \in \{0,1\}^{n^i C' \times 1}$: $\mathbf{x} = [\mathbf{h}_1; \mathbf{h}_2; \cdots; \mathbf{h}_{n^i}]$ to obtain

$$\begin{aligned} \max_{\mathbf{x}} \quad & \mathbf{c}^T \mathbf{x} \\ \text{s.t.} \quad & A_1 \mathbf{x} = \mathbf{1}_{n^i} \\ & A_2 \mathbf{x} \geq \mathbf{1}_{C'} \\ & \mathbf{x} \geq 0, \end{aligned} \tag{3}$$

with

$$A_1 = I_{C'} \otimes \mathbf{1}^T_{n^i} = \begin{bmatrix} 1 & \cdots & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 & 0 & 0 & \cdots & 0 \\ & & & & & & \cdots & & & \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 1 & \cdots & 1 \end{bmatrix},$$

$$A_2 = \mathbf{1}_{C'} \otimes I_{n^i} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & \\ 0 & 1 & \cdots & 0 & 0 & 1 & \cdots & 0 & \cdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 1 & \end{bmatrix}.$$

Following Corollary 1.3, A_1, A_2 are both total unimodular. Next we introduce slack variables $\mathbf{y}$ to rephrase our LP (3) into the following form,

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{c}^T \mathbf{x} \\ \text{s.t.} \quad & A_1 \mathbf{x} = \mathbf{1}_{n^i} \\ & A_2 \mathbf{x} - \mathbf{y} = \mathbf{1}_{C'} \\ & \mathbf{x}, \mathbf{y} \geq 0. \end{aligned} \tag{4}$$

Consequently the coefficient matrix reads as

$$A = \begin{bmatrix} A_1 & 0 \\ A_2 & -I \end{bmatrix}. \tag{5}$$

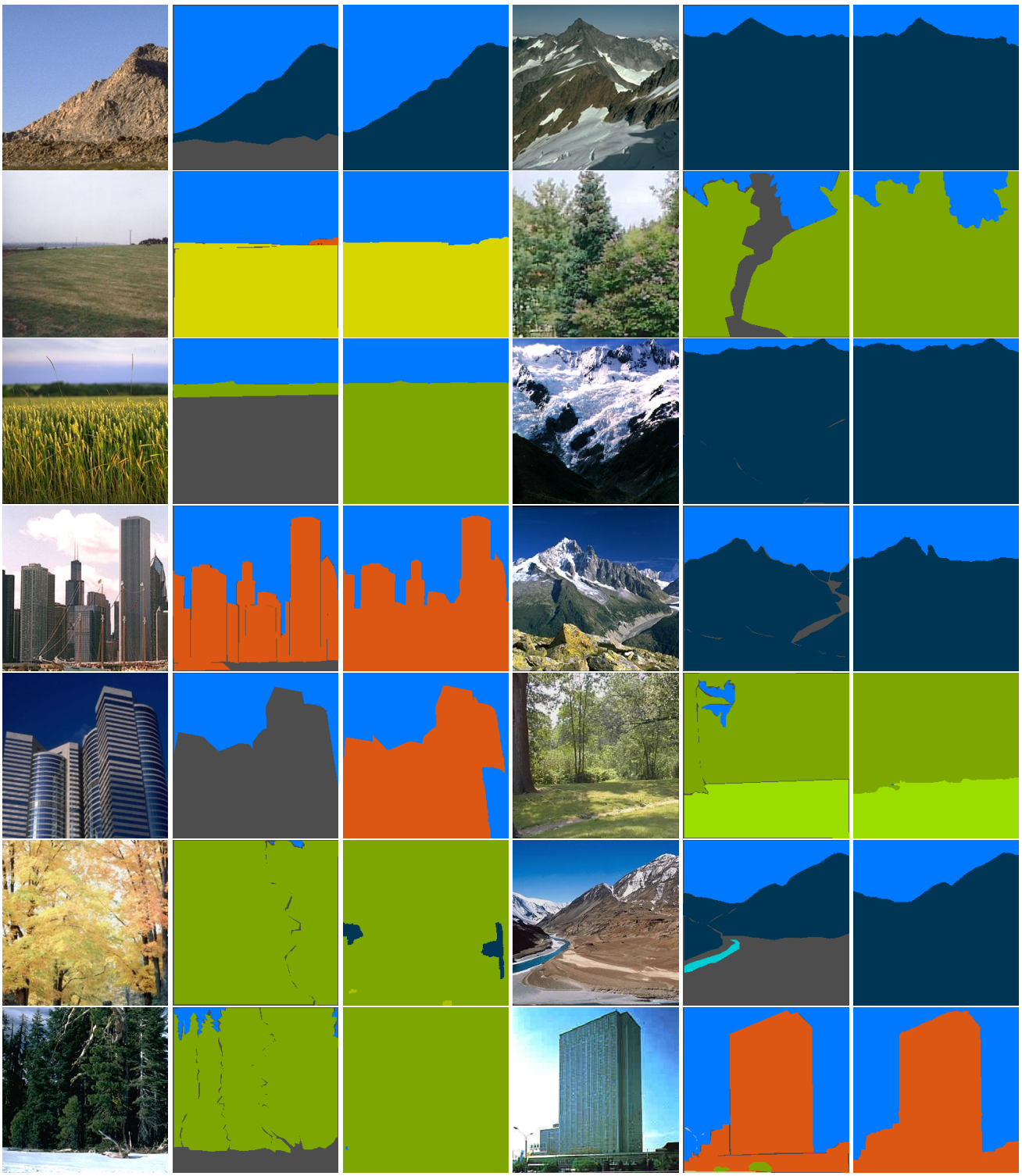
It is obvious that A has at most two non-zero entries in every column. Further, if we partition matrix A into $\begin{bmatrix} A_1 & 0 \end{bmatrix}$ and $\begin{bmatrix} A_2 & -I \end{bmatrix}$, two nonzero entries are in different sets of rows if they have the same sign. Hence, A is total unimodular by Theorem 1.2. Finally, employing Theorem 1.4, the LP relaxation gives integral optima. This concludes our proof.

2. More Results

In the following, we present more results from Ours (ILT+transductive) setting. Figs 1,2,3,4 and 5 (page 3 to page 7) present good cases, which cover scene types of forest, coast, opencountry, mountain, tallbuilding and highway. Note that the third and sixth columns show our segmentation results. Figure 6 (page 7) shows failure cases for scene types of street and inside-city. These cases are mainly due to under segmentation and cluttered textures.

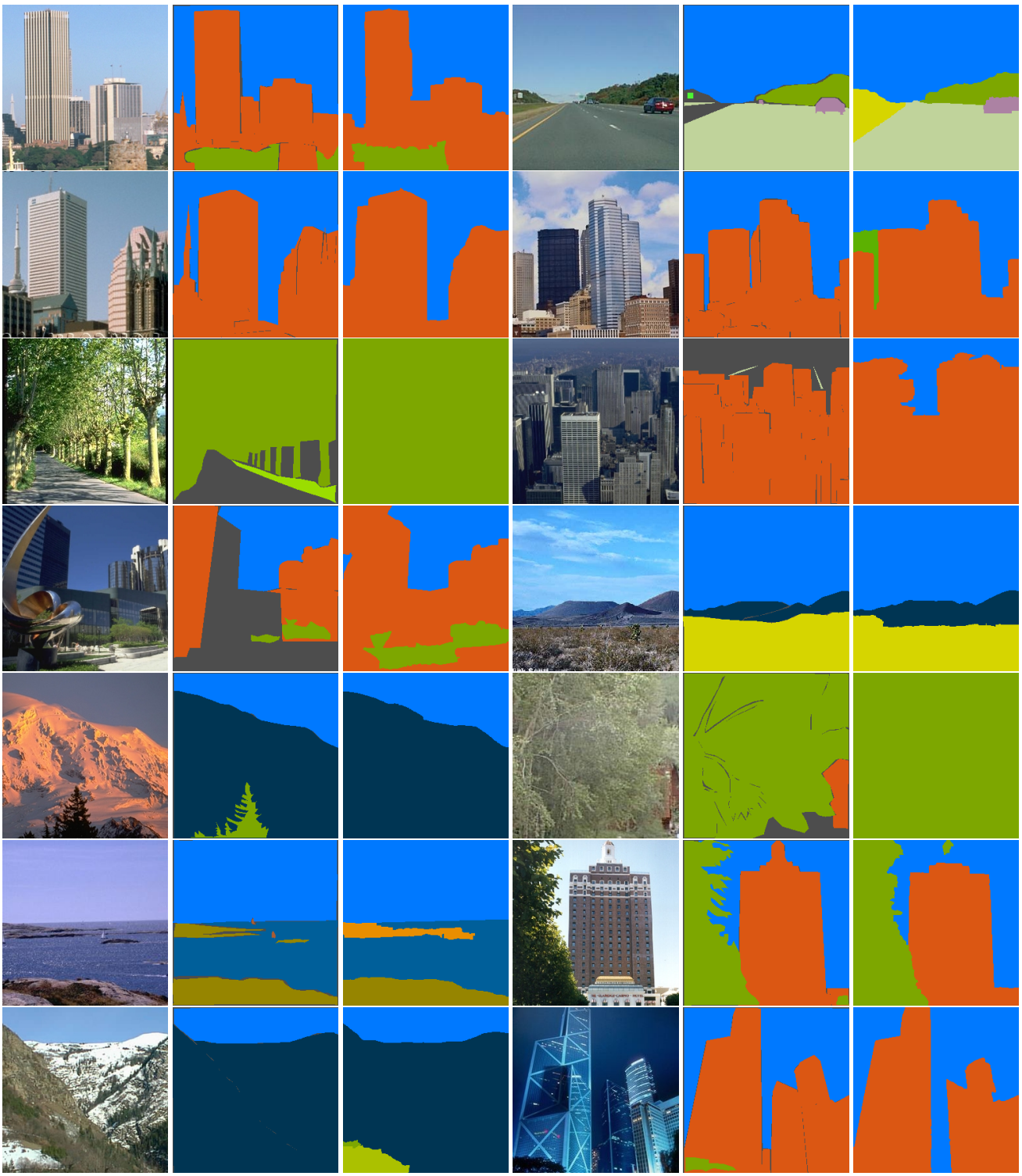
Original Image Ground Truth Ours Original Image Ground Truth Ours

Figure 1. Sample results from "Ours(ILT+transductive)". Note gray regions in the second and fifth column are not labeled in ground truth. **Best viewed in color.**



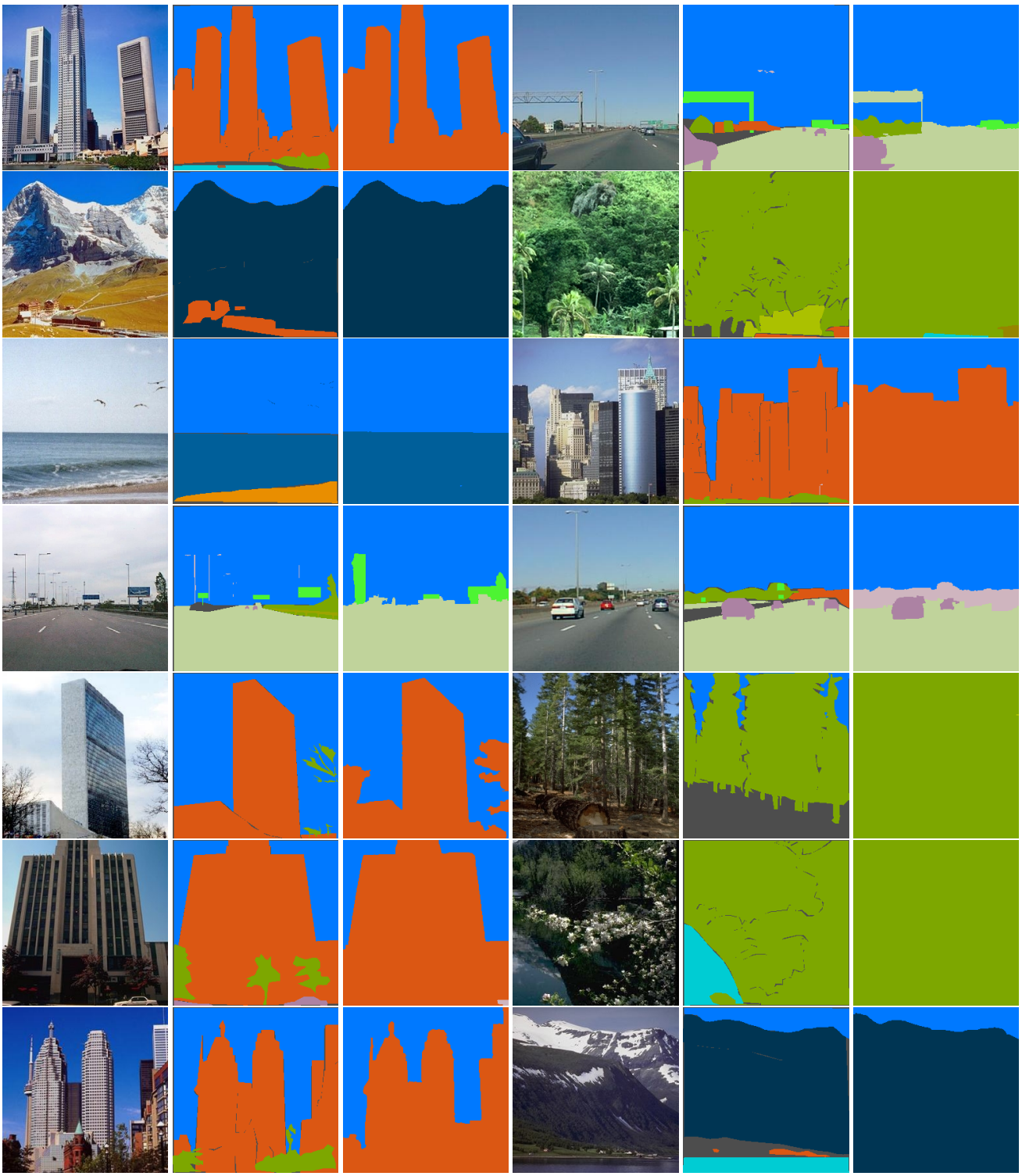
Original Image Ground Truth Ours Original Image Ground Truth Ours

Figure 2. Sample results from "Ours(ILT+transductive)". Note gray regions in the second and fifth column are not labeled in ground truth. **Best viewed in color.**



Original Image Ground Truth Ours Original Image Ground Truth Ours

Figure 3. Sample results from "Ours(ILT+transductive)". Note gray regions in the second and fifth column are not labeled in ground truth. **Best viewed in color.**

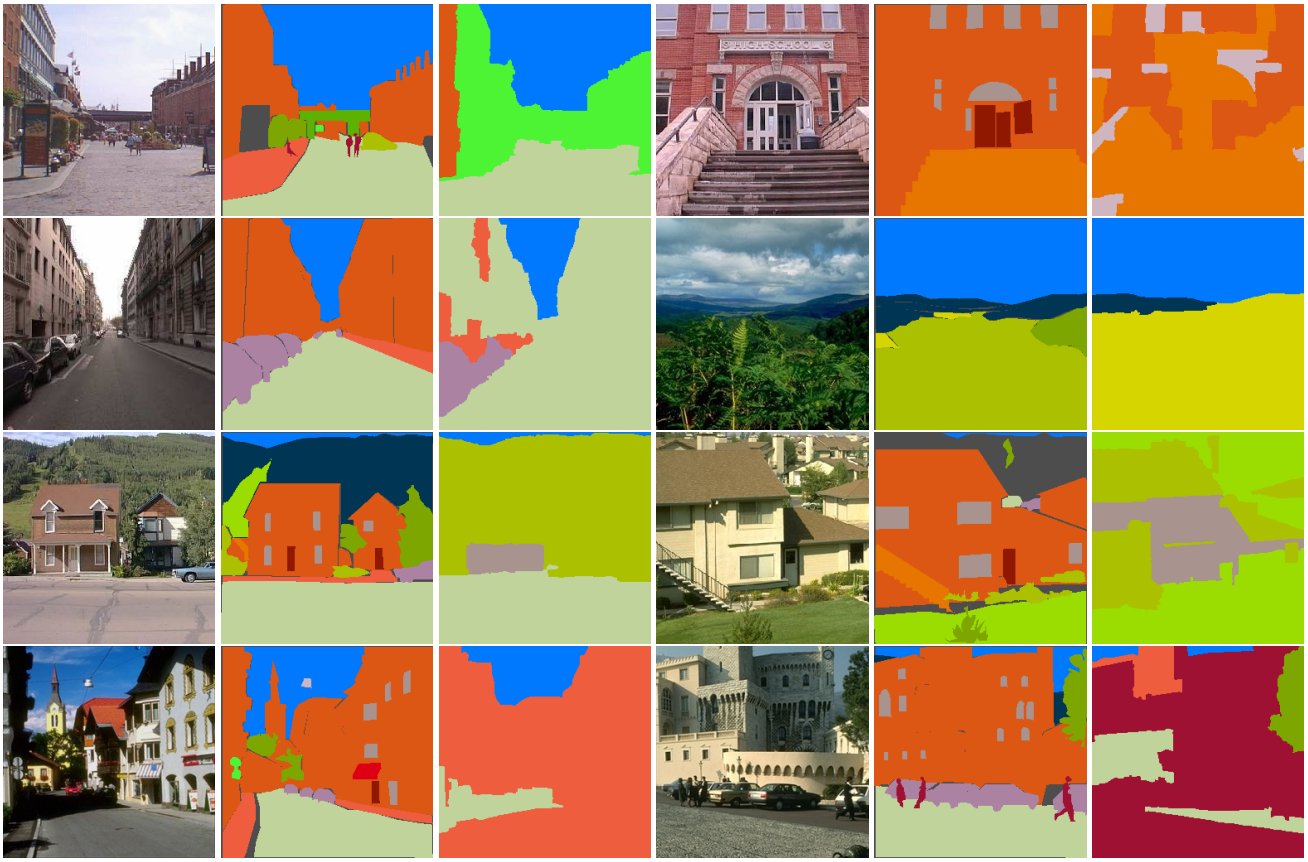


Original Image Ground Truth Ours Original Image Ground Truth Ours

Figure 4. Sample results from "Ours(ILT+transductive)". Note gray regions in the second and fifth column are not labeled in ground truth. **Best viewed in color.**

Original Image Ground Truth Ours Original Image Ground Truth Ours

Figure 5. Sample results from "Ours(ILT+transductive)". Note gray regions in the second and fifth column are not labeled in ground truth. **Best viewed in color.**



Original Image Ground Truth Ours Original Image Ground Truth Ours

Figure 6. Failure cases. **Best viewed in color.**

References

[1] L. Grady. Minimal Surfaces Extend Shortest Path Segmentation Methods to 3D. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(2):321–334, 2010. 1

[2] L. Grady and J. R. Polimeni. *Discrete Calculus: Applied Analysis on Graphs for Computational Science*. Springer, 2010. 1

[3] G. L. Nemhauser and L. A. Wolsey. *Integer and combinatorial optimization*, volume 18. Wiley New York, 1988. 1

[4] K. Truemper. Algebraic characterizations of unimodular matrices. *SIAM Journal on Applied Mathematics*, 35(2):328–332, 1978. 1