Supplementary Materials for “Optimal Decisions from Probabilistic Models: the Intersection-over-Union Case”

Sebastian Nowozin
Microsoft Research
Sebastian.Nowozin@microsoft.com

Abstract

Supplementary materials to the main paper.

1. Proofs

Proof of Proposition 4

Proof. We have

$$\begin{aligned}\mathbb{E}[S_k] &= \sum_{i\in\mathcal{V}} \mathbb{E}_{z_i}[1_{\{z_i=k\wedge y_i=k\}}] \\ &= \sum_{i\in\mathcal{V}} p_i(k) 1_{\{y_i=k\}}.\end{aligned}$$

Likewise we have

$$\begin{aligned}\mathbb{E}[T_k] &= \sum_{i\in\mathcal{V}} \mathbb{E}_{z_i}[1_{\{z_i=k\vee y_i=k\}}] \\ &= \sum_{i\in\mathcal{V}} \mathbb{E}_{z_i}[1_{\{z_i=k\}} + 1_{\{y_i=k\}} - 1_{\{z_i=k\wedge y_i=k\}}] \\ &= \sum_{i\in\mathcal{V}} (p_i(k) + 1_{\{y_i=k\}} - p_i(k)\, 1_{\{y_i=k\}}) \\ &= \sum_{i\in\mathcal{V}} (1_{\{y_i=k\}} + p_i(k)\, 1_{\{y_i\neq k\}}).\end{aligned}$$

□

Proof of Proposition 5

Proof. In (5) the first central moment $\langle {}^1 T_k\rangle$ is always zero, hence the only term that could be non-zero is

$$\begin{aligned}\langle S_k, {}^1 T_k\rangle &= \mathbb{E}[(S_k - \mathbb{E}S_k)(T_k - \mathbb{E}T_k)] \\ &= \mathbb{E}[S_k T_k] - \mathbb{E}S_k \mathbb{E}T_k - \mathbb{E}T_k \mathbb{E}S_k + \mathbb{E}S_k \mathbb{E}T_k \\ &= \mathbb{E}[S_k T_k] - \mathbb{E}S_k \mathbb{E}T_k.\end{aligned}$$

We will show that these two terms are equal and hence their difference is zero. To this end we first expand the product as

$$\begin{aligned}\mathbb{E}[S_k T_k] &= \mathbb{E}\left[(\sum_{i\in\mathcal{V}} 1_{\{z_i=k\wedge y_i=k\}})(\sum_{i\in\mathcal{V}} 1_{\{z_i=k\vee y_i=k\}})\right] \\ &= \mathbb{E}\left[\sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{V}} 1_{\{z_i=k\wedge y_i=k\}} 1_{\{z_j=k\vee y_j=k\}}\right]. \qquad (1)\end{aligned}$$

In order to expand the expectation operator in (1) we would like to use the conditional independence assumption $p_{ij}(z_i, z_j) = p_i(z_i)\, p_j(z_j)$ for $i \neq j$. For this, we split the sum into the cases $i = j$ and $i \neq j$ as follows.

$$\begin{aligned}= &\sum_{i\in\mathcal{V}}\sum_{z_i\in\mathcal{Y}} p_i(z_i)\, 1_{\{z_i=k\wedge y_i=k\}} 1_{\{z_i=k\vee y_i=k\}} + \\ &\sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{V}\setminus\{i\}}\sum_{z_i\in\mathcal{Y}}\sum_{z_j\in\mathcal{Y}} p_i(z_i)\, p_j(z_j) \\ &\quad 1_{\{z_i=k\wedge y_i=k\}} 1_{\{z_j=k\vee y_j=k\}}\end{aligned}$$

The first sum can be simplified by observing that $1_{\{z_i=k\wedge y_i=k\}}\, 1_{\{z_j=k\vee y_j=k\}} = 1_{\{z_i=k\wedge y_i=k\}}$ and by observing that all terms in the sum are zero except for the case when $z_i = k$. For the second sum we reorder terms, obtaining

$$\begin{aligned}= &\sum_{i\in\mathcal{V}} p_i(k)\, 1_{\{y_i=k\}} + \\ &\sum_{i\in\mathcal{V}}\sum_{z_i\in\mathcal{Y}} p_i(z_i)\, 1_{\{z_i=k\wedge y_i=k\}} \\ &\quad \left(\sum_{j\in\mathcal{V}\setminus\{i\}}\sum_{z_j\in\mathcal{Y}} p_j(z_j)\, 1_{\{z_j=k\vee y_j=k\}}\right).\end{aligned}$$

We perform the same simplification as before on the second sum and then factor out the joint term, yielding

$$
\begin{aligned}
&= \sum_{i\in\mathcal{V}} p_i(k)\,1_{\{y_i=k\}} + \sum_{i\in\mathcal{V}} p_i(k)\,1_{\{y_i=k\}} \\
&\qquad \left(\sum_{j\in\mathcal{V}\setminus\{i\}} \left(p_j(k) + (1-p_j(k))\,1_{\{y_j=k\}}\right)\right) \\
&= \sum_{i\in\mathcal{V}} p_i(k)\,1_{\{y_i=k\}} \Big(1 + \\
&\qquad \sum_{j\in\mathcal{V}\setminus\{i\}} \left(p_j(k) + (1-p_j(k))\,1_{\{y_j=k\}}\right)\Big).
\end{aligned}
$$

Due to the factor $1_{\{y_i=k\}}$ we can simplify the case $j=i$ once we observe that $y_i=k$ implies $y_j=k$ and hence the inner summand is equal to one. This allows us to merge the 1 into the sum so that the sum runs over $\mathcal{V}$ instead of over just $\mathcal{V}\setminus\{i\}$.

$$
\begin{aligned}
&= \sum_{i\in\mathcal{V}} p_i(k)\,1_{\{y_i=k\}} \\
&\qquad \Big(\sum_{j\in\mathcal{V}} \left(p_j(k) + (1-p_j(k))\,1_{\{y_j=k\}}\right)\Big).
\end{aligned}
$$

Finally, by considering the cases $y_j=k$ and $y_j\neq k$ separately we observe that the inner summation can be equivalently replaced as follows.

$$
\begin{aligned}
&= \sum_{i\in\mathcal{V}} p_i(k)\,1_{\{y_i=k\}} \sum_{j\in\mathcal{V}} \left(1_{\{y_j=k\}} + p_j(k)\,1_{\{y_j\neq k\}}\right) \\
&= \left(\sum_{i\in\mathcal{V}} p_i(k)\,1_{\{y_i=k\}}\right) \\
&\qquad \left(\sum_{j\in\mathcal{V}} \left(1_{\{y_j=k\}} + p_j(k)\,1_{\{y_j\neq k\}}\right)\right) \\
&= \mathbb{E}[S_k]\,\mathbb{E}[T_k].
\end{aligned}
$$

The proof shows that S_k and T_k are *uncorrelated* random variables and we have $\langle S,^1 T\rangle = 0$. It follows that $\Psi_1 = 0$. $\square$

2. Algorithm Convergence

In Figure 1 we show the value of the optimization objective as a function of optimization iterations. By construction the algorithm produces monotonically increasing objective values.

3. Greedy Algorithm

The greedy algorithm we use is almost identical to the one of Tarlow and Adams [1] except that we use a simple sweep instead of prioritizing some variables over others.

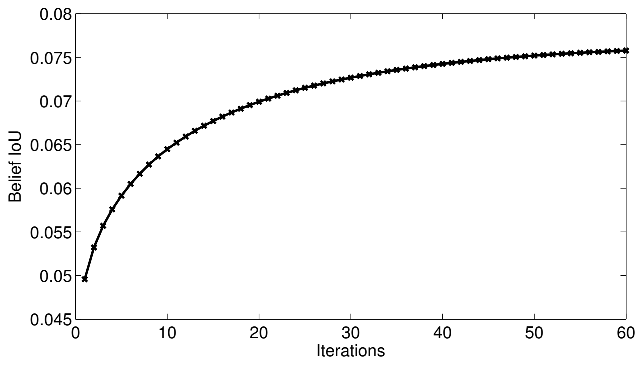

Figure 1. Optimization objective over 60 iterations. Convergence is monotonic by construction but levels off after about thirty iterations.

We start with a random initial solution λ and iteratively increase the objective function (9) as follows. We perform a sequence of *sweeps*. In each sweep we select a random ordering of variables $\mathcal{V}$ and process each variable $j\in\mathcal{V}$ in this order. For the variable j we attempt to relabel it with each of the K possible labels. Let us assume the variable has label k_1 at the moment and we test label $k_2\neq k_1$. To compute the difference $\Delta_{k_1\to k_2}$ in the objective (9) we first define the following statistics.

$$
\begin{aligned}
\alpha_k(\lambda) &= \sum_{i\in\mathcal{V}} p_i(k)\,\lambda_{i,k} \\
\beta_k(\lambda) &= \sum_{i\in\mathcal{V}} \left[p_i(k) + (1-p_i(k))\,\lambda_{i,k}\right]. \qquad (2)
\end{aligned}
$$

Given α and β we can compute the change as

$$
\begin{aligned}
\Delta_{k_1\to k_2} = \frac{1}{K}\Bigg(&\frac{\alpha_{k_2}(\lambda)+p_j(k_2)}{\beta_{k_2}(\lambda)+1-p_j(k_2)} \\
&+\frac{\alpha_{k_1}(\lambda)-p_j(k_1)}{\beta_{k_1}(\lambda)-1+p_j(k_1)} \\
&-\frac{\alpha_{k_2}(\lambda)}{\beta_{k_2}(\lambda)}-\frac{\alpha_{k_1}(\lambda)}{\beta_{k_1}(\lambda)}\Bigg).
\end{aligned}
$$

We test all possible labels $k_2\neq k_1$ and pick the label k^* with maximal $\Delta_{k_1\to k^*}$. If this quantity is strictly positive we relabel the variable and update α and β. If it is zero or negative, we do not change its label, so that it remains as k_1.

In the experiments we always run the greedy algorithm until no further change yields an increase in expected IoU score.

The greedy algorithm has the advantage of being simple and fast to implement, and that it also maintains an integral feasible solution. A potential disadvantage is that it is a local search method with small neighborhood which may get stuck in a suboptimal solution; however, our experiments confirm that this is not the case and therefore indicate that the IoU objective may have certain properties that make it well suited to the greedy local search procedure.

4. Visualization

We show some randomly selected test set predictions from the PASCAL VOC 2012 segmentation test set in Figure 2 and 3.

References

[1] D. Tarlow and R. P. Adams. Revisiting uncertainty in graph cut solutions. In *CVPR*, 2012. 2

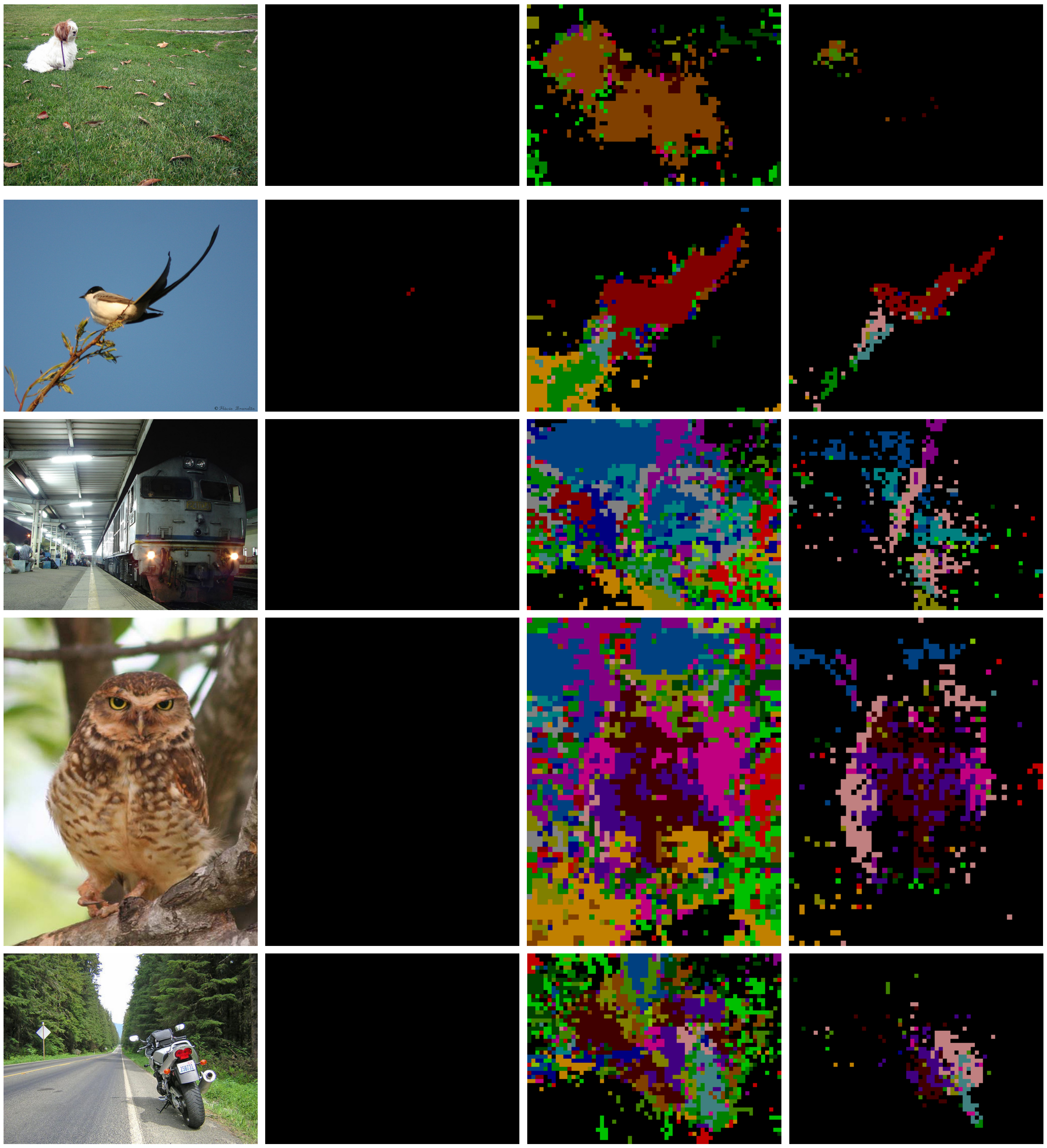

Figure 2. Visualizations of the predictions on the VOC 2012 segmentation test set. The five images are drawn at random from all test images. Columns, from left to right: input image, MAP prediction, inverse-weighted MAP (iwMAP), IoU-optimized predictions after 30 iterations (RF-IoU-opt30).

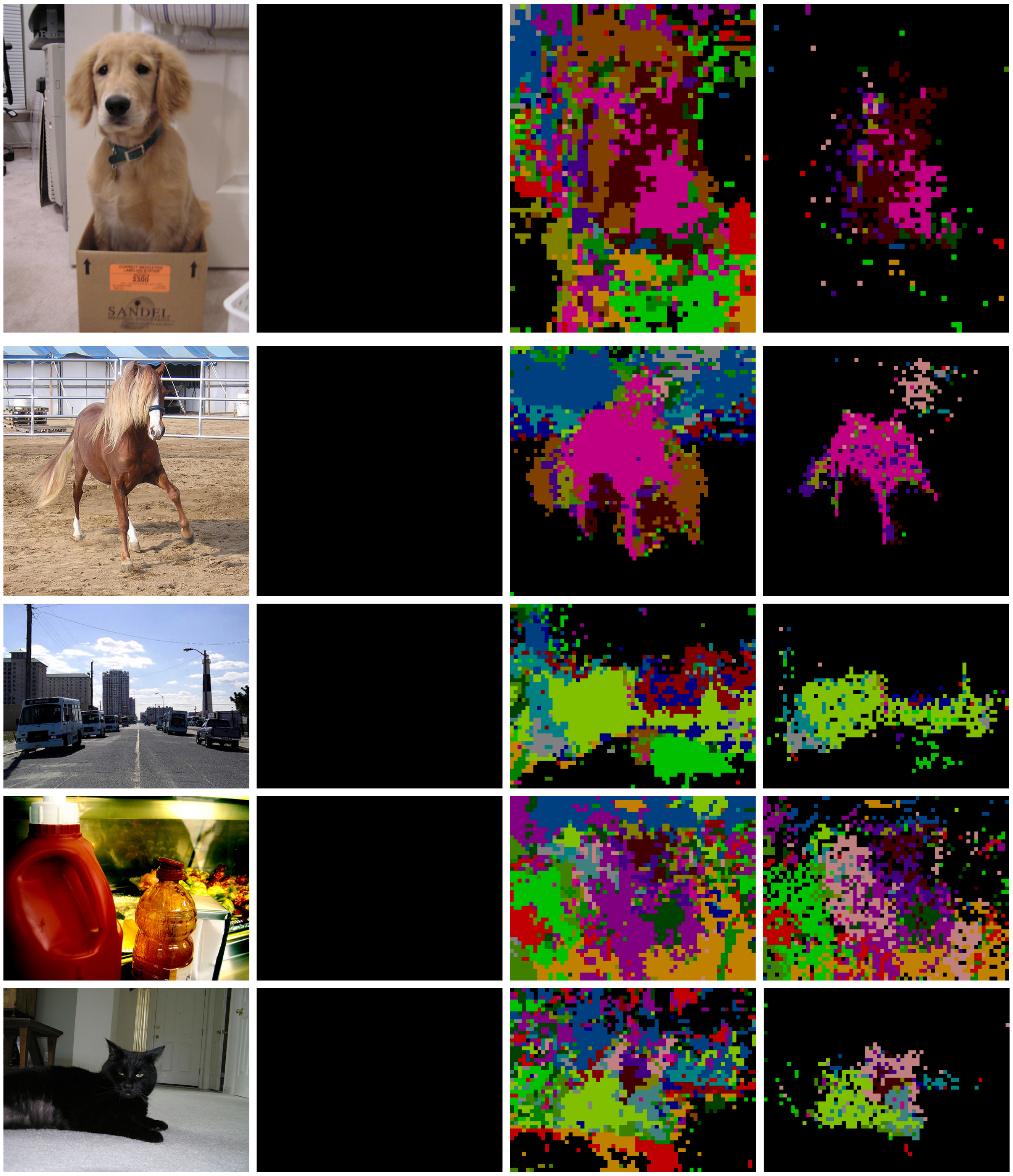
Figure 3. More visualizations, ordering as in Figure 2, from left to right: input image, MAP, iwMAP, IoU-opt30.