Automatic Discovery of Meaningful Object Parts with Latent CRFs Supplemental Material

Paul Schnitzspan[1] Stefan Roth[1] Bernt Schiele[1,2]
[1]Department of Computer Science, TU Darmstadt [2]MPI Informatics, Saarbrücken

In the supplementary material we derive Eqs. (6), (9), (10) and (13) stated in the paper. Most of the derivations rely on factoring out terms and a permutation of the order of the summations.

1. Detecting object instances – Eq. (6)

$$\begin{aligned} p(y=1|\mathbf{x};\theta,E) &= \sum_{\mathbf{z}}\left(\left(\sum_{i\in V}\boldsymbol{\gamma}_i(z_i)\right)p(\mathbf{z}\,|\,\mathbf{x};\boldsymbol{\alpha},\mathbf{e},E)\right) &(1)\\ &= \sum_{i\in V}\sum_{z_i=0}^{P}\left(\boldsymbol{\gamma}_i(z_i)\left(\sum_{\mathbf{z}\setminus z_i}p(\mathbf{z}\,|\,\mathbf{x};\boldsymbol{\alpha},\mathbf{e},E)\right)\right) &(2)\\ &= \sum_{i\in V}\sum_{z_i=0}^{P}\gamma_i(z_i)p(z_i|\,\mathbf{x};\boldsymbol{\alpha},\mathbf{e},E). &(3)\end{aligned}$$

2. Expectation maximization – Eqs. (9) and (10)

Eq. (9). The gradient with respect to the parameters $\boldsymbol{\gamma}$ (Eq. (9) of the paper) can be approximated as follows. Here we make use of the approximation $\frac{p(y|\,\mathbf{z};\boldsymbol{\gamma}^{\text{old}})}{p(y|\,\mathbf{z};\boldsymbol{\gamma})}\approx 1$. Also recall that $p(\mathbf{z}\,|\,\mathbf{x};\boldsymbol{\alpha},\mathbf{e},E)$ does not depend on $\boldsymbol{\gamma}$.

$$\begin{aligned} \frac{\partial Q(\theta,\theta^{\text{old}})}{\partial\gamma_i(c)} &= \sum_{Z}\left(\prod_{j=1}^{M}p(\mathbf{z}^j|y^j,\mathbf{x}^j;\theta^{\text{old}},E^{\text{old}})\right)\left(\sum_{m=1}^{M}\frac{1}{p(y^m|\mathbf{z}^m;\boldsymbol{\gamma})}\frac{\partial}{\partial\gamma_i(c)}p(y^m|\mathbf{z}^m;\boldsymbol{\gamma})\right) &(4)\\ &= \sum_{m=1}^{M}\sum_{\mathbf{z}^m}\frac{p(\mathbf{z}^m|y^m,\mathbf{x}^m;\theta^{\text{old}},E^{\text{old}})}{p(y^m|\mathbf{z}^m;\boldsymbol{\gamma})}\frac{\partial p(y^m|\mathbf{z}^m;\boldsymbol{\gamma})}{\partial\gamma_i(c)}\underbrace{\left(\sum_{Z\setminus\mathbf{z}^m}\prod_{\substack{j=1\\ j\neq m}}^{M}p(\mathbf{z}^j|y^j,\mathbf{x}^j;\theta^{\text{old}},E^{\text{old}})\right)}_{=1} &(5)\\ &= \sum_{m=1}^{M}\sum_{\mathbf{z}^m}\frac{p(y^m|\,\mathbf{z}^m;\boldsymbol{\gamma}^{\text{old}})p(\mathbf{z}^m|\mathbf{x}^m;\boldsymbol{\alpha}^{\text{old}},\mathbf{e}^{\text{old}},E^{\text{old}})}{p(y^m|\mathbf{z}^m;\boldsymbol{\gamma})p(y^m|\mathbf{x}^m;\theta^{\text{old}},E^{\text{old}})}\frac{\partial}{\partial\gamma_i(c)}p(y^m|\mathbf{z}^m;\boldsymbol{\gamma}) &(6)\\ &\approx \sum_{m=1}^{M}\sum_{\mathbf{z}^m}\frac{p(\mathbf{z}^m|\mathbf{x}^m;\boldsymbol{\alpha}^{\text{old}},\mathbf{e}^{\text{old}},E^{\text{old}})}{p(y^m|\mathbf{x}^m;\theta^{\text{old}},E^{\text{old}})}\frac{\partial}{\partial\gamma_i(c)}p(y^m|\mathbf{z}^m;\boldsymbol{\gamma}). &(7)\end{aligned}$$

Here we observe that $\frac{\partial}{\partial\gamma_i(c)}p(y^m|\mathbf{z}^m;\boldsymbol{\gamma})$ does not depend on z_j^m for $j\neq i$ since

$$\frac{\partial}{\partial\gamma_i(c)}p(y^m|\mathbf{z}^m;\boldsymbol{\gamma}) = \begin{cases}1, & \text{if } y^m=1 \text{ and } z_i^m=c\\ -1, & \text{if } y^m=0 \text{ and } z_i^m=c\\ 0, & \text{otherwise.}\end{cases} \quad (8)$$

In the following we factor out $\frac{\partial}{\partial \gamma_i(c)} p(y^m|\mathbf{z}^m;\boldsymbol{\gamma})$ and approximate $p(z_i^m|\mathbf{x}^m;\boldsymbol{\alpha}^{\text{old}},\mathbf{e}^{\text{old}},E^{\text{old}})$ with the belief $b_i^{\text{old}}(z_i)$:

$$\frac{\partial Q(\theta,\theta^{\text{old}})}{\partial \gamma_i(c)} \approx \sum_{m=1}^{M}\sum_{z_i^m=0}^{P}\left(\left(\frac{\partial}{\partial \gamma_i(c)}p(y^m|\mathbf{z}^m;\boldsymbol{\gamma})\right)\sum_{\mathbf{z}^m\setminus z_i^m}\frac{p(\mathbf{z}^m|\mathbf{x}^m;\boldsymbol{\alpha}^{\text{old}},\mathbf{e}^{\text{old}},E^{\text{old}})}{p(y^m|\mathbf{x}^m;\theta^{\text{old}},E^{\text{old}})}\right) \tag{9}$$

$$= \sum_{m=1}^{M}\sum_{z_i^m=0}^{P}\left(\left(\frac{\partial}{\partial \gamma_i(c)}p(y^m|\mathbf{z}^m;\boldsymbol{\gamma})\right)\frac{p(z_i^m|\mathbf{x}^m;\boldsymbol{\alpha}^{\text{old}},\mathbf{e}^{\text{old}},E^{\text{old}})}{p(y^m|\mathbf{x}^m;\theta^{\text{old}},E^{\text{old}})}\right) \tag{10}$$

$$\approx \left(\sum_{m=1,y^m=1}^{M}\frac{b_i^{\text{old}}(c)}{p(y^m|\mathbf{x}^m;\theta^{\text{old}},E^{\text{old}})}\right) - \left(\sum_{m=1,y^m=0}^{M}\frac{b_i^{\text{old}}(c)}{p(y^m|\mathbf{x}^m;\theta^{\text{old}},E^{\text{old}})}\right). \tag{11}$$

Eq. (10). In order to derive Eq. (10) of the paper we simplify the notation of $Q(\cdot,\cdot)$.

$$Q(\theta,\theta^{\text{old}}) = \sum_{m=1}^{M}\sum_{Z}\left(\Big(\log p(y^m|\mathbf{z}^m;\boldsymbol{\gamma}) + \log p(\mathbf{z}^m|\mathbf{x}^m;\boldsymbol{\alpha},\mathbf{e},E)\Big)p(Z|Y,X;\theta^{\text{old}},E^{\text{old}})\right) \tag{12}$$

$$= \sum_{m=1}^{M}\sum_{\mathbf{z}^m}\left(\Big(\log p(y^m|\mathbf{z}^m;\boldsymbol{\gamma}) + \log p(\mathbf{z}^m|\mathbf{x}^m;\boldsymbol{\alpha},\mathbf{e},E)\Big)\sum_{Z\setminus\mathbf{z}^m}p(Z|Y,X;\theta^{\text{old}},E^{\text{old}})\right) \tag{13}$$

$$= \sum_{m=1}^{M}\sum_{\mathbf{z}^m}\left(\Big(\log p(y^m|\mathbf{z}^m;\boldsymbol{\gamma}) + \log p(\mathbf{z}^m|\mathbf{x}^m;\boldsymbol{\alpha},\mathbf{e},E)\Big)p(\mathbf{z}^m|y^m,\mathbf{x}^m;\theta^{\text{old}},E^{\text{old}})\right). \tag{14}$$

Following this we can easily derive the gradient with respect to the edge parameters $\mathbf{e}$. Recall that $p(y|\mathbf{z};\boldsymbol{\gamma})$ does not depend on the parameters $\mathbf{e}$.

$$\frac{\partial Q(\theta,\theta^{\text{old}})}{\partial \mathbf{e}_{ij}^{c_1c_2}} = \sum_{m=1}^{M}\sum_{\mathbf{z}^m}\left(\left(\frac{\partial}{\partial \mathbf{e}_{ij}^{c_1c_2}}\log(p(\mathbf{z}^m|\mathbf{x}^m;\boldsymbol{\alpha},\mathbf{e},E))\right)\frac{p(y^m|\mathbf{z}^m;\boldsymbol{\gamma}^{\text{old}})p(\mathbf{z}^m|\mathbf{x}^m;\boldsymbol{\alpha}^{\text{old}},\mathbf{e}^{\text{old}},E^{\text{old}})}{p(y^m|\mathbf{x}^m;\theta^{\text{old}},E^{\text{old}})}\right). \tag{15}$$

Thus, the gradient can be expressed in terms of the conditional log-likelihood gradient in CRFs.

3. Structure Learning – Eq. (13)

In the following we derive Eq. (13) of the paper. For estimating the structure of the domain of interest we want to maximize the ratio

$$\mathcal{R}(\mathbf{e}_{ij}) = \max_{(c_1,c_2)\in\{0..P\}^2}\left\|\sum_{m=1,y^m=1}^{M}\frac{\partial \log p(y^m=1|\mathbf{x}^m;\theta,E)}{\partial \mathbf{e}_{ij}^{c_1c_2}} - \sum_{m=1,y^m=0}^{M}\frac{\partial \log p(y^m=0|\mathbf{x}^m;\theta,E)}{\partial \mathbf{e}_{ij}^{c_1c_2}}\right\|$$

yielding the edge (i^*,j^*) that most likely improves the discriminative power of our model. By exploiting that

$$\frac{\partial}{\partial \mathbf{e}_{ij}^{c_1c_2}}p(\mathbf{z}|\mathbf{x};\boldsymbol{\alpha},\mathbf{e},E) = p(\mathbf{z}|\mathbf{x};\boldsymbol{\alpha},\mathbf{e},E)\cdot\frac{\partial}{\partial \mathbf{e}_{ij}^{c_1c_2}}\log p(\mathbf{z}|\mathbf{x};\boldsymbol{\alpha},\mathbf{e},E) \tag{16}$$

we derive this ratio for a fixed m and drop this index for notational simplicity:

$$\frac{\partial \log p(y|\mathbf{x};\theta,E)}{\partial \mathbf{e}_{ij}^{c_1c_2}} = \frac{1}{p(y|\mathbf{x};\theta,E)}\frac{\partial}{\partial \mathbf{e}_{ij}^{c_1c_2}}\left(\sum_{\mathbf{z}}\underbrace{p(y|\mathbf{z};\boldsymbol{\gamma})}_{\text{independent of }\mathbf{e}_{ij}^{c_1c_2}}p(\mathbf{z}|\mathbf{x};\boldsymbol{\alpha},\mathbf{e},E)\right) \tag{17}$$

$$= \sum_{\mathbf{z}}\left(\frac{p(y|\mathbf{z};\boldsymbol{\gamma})}{p(y|\mathbf{x};\theta,E)}\left(\frac{\partial}{\partial \mathbf{e}_{ij}^{c_1c_2}}p(\mathbf{z}|\mathbf{x};\boldsymbol{\alpha},\mathbf{e},E)\right)\right) \tag{18}$$

$$= \sum_{\mathbf{z}}\frac{p(y|\mathbf{z};\boldsymbol{\gamma})p(\mathbf{z}\mid\mathbf{x};\boldsymbol{\alpha},\mathbf{e},E)}{p(y|\mathbf{x};\theta,E)}\frac{\partial \log p(\mathbf{z}\mid\mathbf{x};\boldsymbol{\alpha},\mathbf{e},E)}{\partial \mathbf{e}_{ij}^{c_1c_2}}. \tag{19}$$

We can now compute $\mathcal{R}(\mathbf{e}_{ij})$ using the conditional log-likelihood gradient of a standard CRF.