Supplementary of Exploiting Negative Evidence for Deep Latent Structured Models

Thibaut Durand, Nicolas Thome and Matthieu Cord

Appendix A
Proof of the upper bound

We show here that the loss function $\ell(\mathbf{x}_i, \mathbf{y}_i)$ (Eq. (6) of the submitted paper) is an upper bound of $\Delta(\mathbf{y}_i, \hat{\mathbf{y}})$, where $\mathbf{x}_i$ is the input, $\mathbf{y}_i$ is the ground truth, and $\hat{\mathbf{y}}$ the predicted output.
Proof:

$$\hat{\mathbf{y}} = \arg\max_{\mathbf{y}\in\mathcal{Y}} s_{\mathbf{w}}(\mathbf{x}_i, \mathbf{y}) \tag{1}$$

$$\Delta(\hat{\mathbf{y}}, \mathbf{y}_i) \leq \Delta(\hat{\mathbf{y}}, \mathbf{y}_i) + \underbrace{s_{\mathbf{w}}(x_i, \hat{\mathbf{y}}) - s_{\mathbf{w}}(\mathbf{x}_i, \mathbf{y}_i)}_{\geq 0} \tag{2}$$

$$\leq \max_{\mathbf{y}\in\mathcal{Y}} [\Delta(\mathbf{y}, \mathbf{y}_i) + s_{\mathbf{w}}(\mathbf{x}_i, \mathbf{y}) - s_{\mathbf{w}}(\mathbf{x}_i, \mathbf{y}_i)] \tag{3}$$

$$\leq \ell(\mathbf{x}_i, \mathbf{y}_i) \tag{4}$$

This proves that $\ell(\mathbf{x}_i, \mathbf{y}_i)$ is an upper bound of $\Delta(\mathbf{y}_i, \hat{\mathbf{y}})$.

Appendix B
Proof of Proposition 1

We prove that the inference is equivalent to a supervised inference, where each image x_i is represented by $s^{top}_{\mathbf{w},k}(F^t_{\mathbf{w}}(x_i, c)) + s^{low}_{\mathbf{w},k}(F^t_{\mathbf{w}}(x_i, c))$. We show that the selected regions can be predicted independently to ranking $\mathbf{y}$:

$$s(\mathbf{x}, \mathbf{y}) = s^{top}_{\mathbf{w},k}(F^t_{\mathbf{w}}(\mathbf{x}, \mathbf{y})) + s^{low}_{\mathbf{w},k}(F^t_{\mathbf{w}}(\mathbf{x}, \mathbf{y})) \tag{5}$$

$$\begin{aligned}=&\max_{\mathbf{h}} \sum_{p\in\mathcal{P}} \sum_{n\in\mathcal{N}} \mathbf{y}_{pn} \left[\sum_{z=1}^{r_p} h^{pn}_z F^t_{\mathbf{w}}(x_p, c, z) - \sum_{z'=1}^{r_n} h^{np}_{z'} F^t_{\mathbf{w}}(x_n, c, z')\right] \\ &+ \min_{\bar{\mathbf{h}}} \sum_{p\in\mathcal{P}} \sum_{n\in\mathcal{N}} \mathbf{y}_{pn} \left[\sum_{z=1}^{r_p} \bar{h}^{pn}_z F^t_{\mathbf{w}}(x_p, c, z) - \sum_{z'=1}^{r_n} \bar{h}^{np}_{z'} F^t_{\mathbf{w}}(x_n, c, z')\right]\end{aligned} \tag{6}$$

$$\begin{aligned}=&\sum_{p\in\mathcal{P}} \sum_{n\in\mathcal{N}} \Bigg(\max_{(\mathbf{h}^{pn}, \mathbf{h}^{np})} \mathbf{y}_{pn} \left[\sum_{z=1}^{r_p} h^{pn}_z F^t_{\mathbf{w}}(x_p, c, z) - \sum_{z'=1}^{r_n} h^{np}_z F^t_{\mathbf{w}}(x_n, c, z')\right] \\ &+ \min_{(\bar{\mathbf{h}}^{pn}, \bar{\mathbf{h}}^{np})} \mathbf{y}_{pn} \left[\sum_{z=1}^{r_p} \bar{h}^{pn}_z F^t_{\mathbf{w}}(x_p, c, z) - \sum_{z'=1}^{r_n} \bar{h}^{np}_{z'} F^t_{\mathbf{w}}(x_n, c, z')\right]\Bigg)\end{aligned} \tag{7}$$

The maximization (resp. minimization) can be decomposed for each term of the sum, so maximizing (resp. minimizing) the sum is equivalent to maximize (resp. minimize) each term of the sum. Now, we analyze the predicted regions with respect to $\mathbf{y}_{pn}$ value.
If $\mathbf{y}_{pn} = 1$

$$\begin{aligned}&\max_{(\mathbf{h}^{pn}, \mathbf{h}^{np})} \left[\sum_{z=1}^{r_p} h^{pn}_z F^t_{\mathbf{w}}(x_p, c, z) - \sum_{z'=1}^{r_n} h^{np}_z F^t_{\mathbf{w}}(x_n, c, z')\right] \\ &+ \min_{(\bar{\mathbf{h}}^{pn}, \bar{\mathbf{h}}^{np})} \left[\sum_{z=1}^{r_p} \bar{h}^{pn}_z F^t_{\mathbf{w}}(x_p, c, z) - \sum_{z'=1}^{r_n} \bar{h}^{np}_{z'} F^t_{\mathbf{w}}(x_n, c, z')\right] \\ =&\left(\max_{\mathbf{h}^{pn}} \sum_{z=1}^{r_p} h^{pn}_z F^t_{\mathbf{w}}(x_p, c, z) + \min_{\bar{\mathbf{h}}^{pn}} \sum_{z'=1}^{r_p} \bar{h}^{pn}_{z'} F^t_{\mathbf{w}}(x_p, c, z')\right) \\ &- \left(\max_{\bar{\mathbf{h}}^{np}} \sum_{z=1}^{n'_n} \bar{h}^{np}_z F^t_{\mathbf{w}}(x_n, c, z) + \min_{\bar{\mathbf{h}}^{np}} \sum_{z'=1}^{n'_n} h^{np}_{z'} F^t_{\mathbf{w}}(x_n, c, z')\right)\end{aligned} \tag{8}$$

$$\begin{aligned}=&s^{top}_{\mathbf{w},k}(F^t_{\mathbf{w}}(x_p, c)) + s^{low}_{\mathbf{w},k}(F^t_{\mathbf{w}}(x_p, c)) \\ &- (s^{top}_{\mathbf{w},k}(F^t_{\mathbf{w}}(x_n, c)) + s^{low}_{\mathbf{w},k}(F^t_{\mathbf{w}}(x_n, c))\end{aligned} \tag{9}$$

If $\mathbf{y}_{pn} = -1$

$$\begin{aligned}&\max_{(\mathbf{h}^{pn}, \mathbf{h}^{np})} - \left[\sum_{z=1}^{r_p} h^{pn}_z F^t_{\mathbf{w}}(x_p, c, z) - \sum_{z'=1}^{r_n} h^{np}_z F^t_{\mathbf{w}}(x_n, c, z')\right] \\ &+ \min_{(\bar{\mathbf{h}}^{pn}, \bar{\mathbf{h}}^{np})} - \left[\sum_{z=1}^{r_p} \bar{h}^{pn}_z F^t_{\mathbf{w}}(x_p, c, z) - \sum_{z'=1}^{r_n} \bar{h}^{np}_{z'} F^t_{\mathbf{w}}(x_n, c, z')\right] \\ =&\left(\max_{\mathbf{h}^{np}} \sum_{z=1}^{r_p} h^{np}_z F^t_{\mathbf{w}}(x_n, c, z) + \min_{\bar{\mathbf{h}}^{np}} \sum_{z'=1}^{r_p} \bar{h}^{np}_{z'} F^t_{\mathbf{w}}(x_n, c, z')\right) \\ &- \left(\max_{\bar{\mathbf{h}}^{pn}} \sum_{z=1}^{n'_n} \bar{h}^{pn}_z F^t_{\mathbf{w}}(x_p, c, z) + \min_{\bar{\mathbf{h}}^{pn}} \sum_{z'=1}^{n'_n} h^{pn}_{z'} F^t_{\mathbf{w}}(x_p, c, z')\right)\end{aligned} \tag{10}$$

$$\begin{aligned}=&-\Big(s^{top}_{\mathbf{w},k}(F^t_{\mathbf{w}}(x_p, c)) + s^{low}_{\mathbf{w},k}(F^t_{\mathbf{w}}(x_p, c)) \\ &- (s^{top}_{\mathbf{w},k}(F^t_{\mathbf{w}}(x_n, c)) + s^{low}_{\mathbf{w},k}(F^t_{\mathbf{w}}(x_n, c))\Big)\end{aligned} \tag{11}$$

We notice that the predicted regions are the same in the two cases: the predicted regions can be fixed independently to the value of $\mathbf{y}_{pn}$. The inference can be written as a supervised inference, where the region are fixed independently to the ranking matrix $\mathbf{y}$, and each image x_i is represented by $s^{top}_{\mathbf{w},k}(F^t_{\mathbf{w}}(x_i, c)) + s^{low}_{\mathbf{w},k}(F^t_{\mathbf{w}}(x_i, c))$.