Deep Gaussian Conditional Random Field Network: A Model-based Deep Network for Discriminative Denoising - Supplementary Material

Raviteja Vemulapalli
Center for Automation Research, UMIACS
University of Maryland, College Park

Oncel Tuzel, Ming-Yu Liu
Mitsubishi Electric Research Laboratories
Cambridge, MA

Section 1 of this supplementary material provides detailed derivations for the derivative formulas presented in the appendix of the main submission, and section 2 of this supplementary material provides an algorithmic description of the proposed deep Gaussian CRF network.

1. Backpropagation

1.1. Backpropagation through the combination layer

Forward step:

$$\mathbf{\Sigma}_{ij} = \sum_k \gamma_{ij}^k \mathbf{\Psi}_k. \tag{1}$$

Backward step:

$$\frac{dL}{d\gamma_{ij}^k} = \sum_{pq} \frac{dL}{d\mathbf{\Sigma}_{ij}(p,q)} \frac{d\mathbf{\Sigma}_{ij}(p,q)}{d\gamma_{ij}^k} = \sum_{pq} \frac{dL}{d\mathbf{\Sigma}_{ij}(p,q)} \mathbf{\Psi}_k(p,q) = \text{trace}\left(\mathbf{\Psi}_k^\top \frac{dL}{d\mathbf{\Sigma}_{ij}}\right). \tag{2}$$

$$\frac{dL}{d\mathbf{\Psi}_k(p,q)} = \sum_{ij} \frac{dL}{d\mathbf{\Sigma}_{ij}(p,q)} \frac{d\mathbf{\Sigma}_{ij}(p,q)}{d\mathbf{\Psi}_k(p,q)} = \sum_{ij} \frac{dL}{d\mathbf{\Sigma}_{ij}(p,q)} \gamma_{ij}^k \implies \frac{dL}{d\mathbf{\Psi}_k} = \sum_{ij} \gamma_{ij}^k \frac{dL}{d\mathbf{\Sigma}_{ij}}. \tag{3}$$

1.2. Backpropagation through the quadratic layer

Forward step:

$$s_{ij}^k = -\frac{1}{2} \bar{\mathbf{x}}_{ij}^\top \left(\mathbf{W}_k + \sigma^2 \mathbf{I}\right)^{-1} \bar{\mathbf{x}}_{ij} + b_k. \tag{4}$$

Backward step:

$$\frac{dL}{db_k} = \sum_{ij} \frac{dL}{ds_{ij}^k} \frac{ds_{ij}^k}{db_k} = \sum_{ij} \frac{dL}{ds_{ij}^k}. \tag{5}$$

$$\frac{dL}{d\bar{\mathbf{x}}_{ij}} = \sum_k \frac{dL}{ds_{ij}^k} \frac{ds_{ij}^k}{d\bar{\mathbf{x}}_{ij}} = -\sum_k \frac{dL}{ds_{ij}^k} \left(\mathbf{W}_k + \sigma^2 \mathbf{I}\right)^{-1} \bar{\mathbf{x}}_{ij}. \tag{6}$$

$$\begin{aligned}
\frac{dL}{d\mathbf{W}_k(p,q)} = \sum_{ij} \frac{dL}{ds_{ij}^k} \frac{ds_{ij}^k}{d\mathbf{W}_k(p,q)} &= -\frac{1}{2} \sum_{ij} \frac{dL}{ds_{ij}^k} \bar{\mathbf{x}}_{ij}^\top \frac{d(\mathbf{W}_k + \sigma^2 I)^{-1}}{d\mathbf{W}_k(p,q)} \bar{\mathbf{x}}_{ij} \\
&= \frac{1}{2} \sum_{ij} \frac{dL}{ds_{ij}^k} \bar{\mathbf{x}}_{ij}^\top (\mathbf{W}_k + \sigma^2 I)^{-1} \frac{d(\mathbf{W}_k + \sigma^2 I)}{d\mathbf{W}_k(p,q)} (\mathbf{W}_k + \sigma^2 I)^{-1} \bar{\mathbf{x}}_{ij} \\
&= \frac{1}{2} \sum_{ij} \frac{dL}{ds_{ij}^k} \left[(\mathbf{W}_k + \sigma^2 I)^{-1} \bar{\mathbf{x}}_{ij} \bar{\mathbf{x}}_{ij}^\top (\mathbf{W}_k + \sigma^2 I)^{-1}\right](p,q)
\end{aligned} \tag{7}$$

From (7), we have

$$\frac{dL}{d\mathbf{W}_k} = \frac{1}{2}\sum_{ij} \frac{dL}{ds_{ij}^k} \left[(\mathbf{W}_k + \sigma^2 I)^{-1}\, \bar{\mathbf{x}}_{ij}\, \bar{\mathbf{x}}_{ij}^\top\, (\mathbf{W}_k + \sigma^2 I)^{-1} \right]$$
$$= (\mathbf{W}_k + \sigma^2 I)^{-1} \left(\sum_{ij} \frac{dL}{ds_{ij}^k} \frac{\bar{\mathbf{x}}_{ij}\, \bar{\mathbf{x}}_{ij}^\top}{2} \right) (\mathbf{W}_k + \sigma^2 I)^{-1} \tag{8}$$

1.3. Backpropagation through the patch inference layer

Forward step:

$$\mathbf{z}_{ij} = \left(\mathbf{I} - \mathbf{G}^\top \left(\beta \mathbf{\Sigma}_{ij} + \mathbf{G}\mathbf{G}^\top \right)^{-1} \mathbf{G} \right) \mathbf{y}_{ij}. \tag{9}$$

Backward step:

Let $\mathbf{A}_{ij} = \mathbf{I} - \mathbf{G}^\top \left(\beta \mathbf{\Sigma}_{ij} + \mathbf{G}\mathbf{G}^\top \right)^{-1} \mathbf{G}$. Let $\mathbb{I}_{pq}$ be a matrix with (p,q) element as one and all other elements as zero. Note that the matrix $\mathbf{G} = \mathbf{I} - \frac{1}{d^2}\mathbf{1}\mathbf{1}^\top$ satisfies $\mathbf{G}\mathbf{G}^\top = \mathbf{G}$.

$$\mathbf{z}_{ij} = \mathbf{A}_{ij}\mathbf{y}_{ij} \implies \frac{dL}{d\mathbf{y}_{ij}} = \mathbf{A}_{ij}^\top \frac{dL}{d\mathbf{z}_{ij}} = \left(\mathbf{I} - \mathbf{G}^\top \left(\beta \mathbf{\Sigma}_{ij} + \mathbf{G}\mathbf{G}^\top \right)^{-1} \mathbf{G} \right) \frac{dL}{d\mathbf{z}_{ij}} = \left(\mathbf{I} - \mathbf{G}^\top \left(\beta \mathbf{\Sigma}_{ij} + \mathbf{G} \right)^{-1} \mathbf{G} \right) \frac{dL}{d\mathbf{z}_{ij}}. \tag{10}$$

$$\mathbf{z}_{ij} = \mathbf{A}_{ij}\mathbf{y}_{ij} \implies \frac{dL}{d\mathbf{A}_{ij}} = \frac{dL}{d\mathbf{z}_{ij}} \mathbf{y}_{ij}^\top \tag{11}$$

$$\frac{dL}{d\mathbf{\Sigma}_{ij}(p,q)} = \sum_{r,s} \frac{dL}{d\mathbf{A}_{ij}(r,s)} \frac{d\mathbf{A}_{ij}(r,s)}{d\mathbf{\Sigma}_{ij}(p,q)} = \text{trace}\left(\left(\frac{dL}{d\mathbf{A}_{ij}} \right)^\top \frac{d\mathbf{A}_{ij}}{d\mathbf{\Sigma}_{ij}(p,q)} \right) \tag{12}$$

$$\frac{d\mathbf{A}_{ij}}{d\mathbf{\Sigma}_{ij}(p,q)} = -\mathbf{G}^\top \frac{d\left(\beta \mathbf{\Sigma}_{ij} + \mathbf{G}\mathbf{G}^\top \right)^{-1}}{d\mathbf{\Sigma}_{ij}(p,q)} \mathbf{G}$$
$$= \mathbf{G}^\top \left(\beta \mathbf{\Sigma}_{ij} + \mathbf{G}\mathbf{G}^\top \right)^{-1} \frac{d\left(\beta \mathbf{\Sigma}_{ij} + \mathbf{G}\mathbf{G}^\top \right)}{d\mathbf{\Sigma}_{ij}(p,q)} \left(\beta \mathbf{\Sigma}_{ij} + \mathbf{G}\mathbf{G}^\top \right)^{-1} \mathbf{G} \tag{13}$$
$$= \mathbf{G}^\top \left(\beta \mathbf{\Sigma}_{ij} + \mathbf{G}\mathbf{G}^\top \right)^{-1} \left(\beta \mathbb{I}_{pq} \right) \left(\beta \mathbf{\Sigma}_{ij} + \mathbf{G}\mathbf{G}^\top \right)^{-1} \mathbf{G}$$

From (12) and (13), we have

$$\frac{dL}{d\mathbf{\Sigma}_{ij}(p,q)} = \text{trace}\left(\left(\frac{dL}{d\mathbf{A}_{ij}} \right)^\top \left(\mathbf{G}^\top \left(\beta \mathbf{\Sigma}_{ij} + \mathbf{G}\mathbf{G}^\top \right)^{-1} \left(\beta \mathbb{I}_{pq} \right) \left(\beta \mathbf{\Sigma}_{ij} + \mathbf{G}\mathbf{G}^\top \right)^{-1} \mathbf{G} \right) \right)$$
$$= \text{trace}\left(\left(\beta \mathbf{\Sigma}_{ij} + \mathbf{G}\mathbf{G}^\top \right)^{-1} \mathbf{G} \left(\frac{dL}{d\mathbf{A}_{ij}} \right)^\top \mathbf{G}^\top \left(\beta \mathbf{\Sigma}_{ij} + \mathbf{G}\mathbf{G}^\top \right)^{-1} \left(\beta \mathbb{I}_{pq} \right) \right) \tag{14}$$

From (11) and (14), we have

$$\frac{dL}{d\mathbf{\Sigma}_{ij}} = \beta \left(\left(\beta \mathbf{\Sigma}_{ij} + \mathbf{G}\mathbf{G}^\top \right)^{-1} \mathbf{G} \left(\frac{dL}{d\mathbf{A}_{ij}} \right)^\top \mathbf{G}^\top \left(\beta \mathbf{\Sigma}_{ij} + \mathbf{G}\mathbf{G}^\top \right)^{-1} \right)^\top$$
$$= \beta \left(\beta \mathbf{\Sigma}_{ij} + \mathbf{G}\mathbf{G}^\top \right)^{-1} \mathbf{G} \left(\frac{dL}{d\mathbf{A}_{ij}} \right) \mathbf{G}^\top \left(\beta \mathbf{\Sigma}_{ij} + \mathbf{G}\mathbf{G}^\top \right)^{-1} \tag{15}$$
$$= \beta \left(\beta \mathbf{\Sigma}_{ij} + \mathbf{G} \right)^{-1} \mathbf{G} \frac{dL}{d\mathbf{z}_{ij}} \mathbf{y}_{ij}^\top \mathbf{G}^\top \left(\beta \mathbf{\Sigma}_{ij} + \mathbf{G} \right)^{-1}$$

2. Algorithmic description of the proposed deep Gaussian CRF network

Algorithm 1 Deep Gaussian CRF Network

Input: Noisy image $\mathbf{X}$, noise variance σ^2

Initialize the network input $\mathbf{Y}^0$ with the noisy input image $\mathbf{X}$.

for $t = 1$ to T **do**

Parameter generation network (PgNet):

Patch extraction layer: Extract the $d \times d$ mean subtracted patches $\{\bar{\mathbf{y}}_{ij}^{t-1}\}$ from the image $\mathbf{Y}^{t-1}$ at all pixels (i, j).

Selection network (parameters $\{(\mathbf{W}_k, b_k)\}$): Compute the combination weights $\{\gamma_{ij}^k\}$ for each patch $\bar{\mathbf{y}}_{ij}^{t-1}$ using

$$s_{ij}^k = -\frac{1}{2}\left(\bar{\mathbf{y}}_{ij}^{t-1}\right)^\top \left(\mathbf{W}_k + \sigma^2 \mathbf{I}\right)^{-1} \bar{\mathbf{y}}_{ij}^{t-1} + b_k, \quad \gamma_{ij}^k = e^{s_{ij}^k} / \sum_{p=1}^{K} e^{s_{ij}^p}. \tag{16}$$

Combination layer (parameters $\{\mathbf{\Psi}_k\}$): Compute the potential parameters $\mathbf{\Sigma}_{ij}$ for each patch $\bar{\mathbf{y}}_{ij}^{t-1}$ using

$$\mathbf{\Sigma}_{ij} = \sum_{k=1}^{K} \gamma_{ij}^k \mathbf{\Psi}_k. \tag{17}$$

HQS layer of the inference network (InfNet):

Patch inference layer (PI): Compute the auxiliary patches $\{\mathbf{z}_{ij}\}$ using the patches $\{\mathbf{y}_{ij}^{t-1}\}$ extracted from the image $\mathbf{Y}^{t-1}$, and the pairwise potential parameters $\{\mathbf{\Sigma}_{ij}\}$ given by the PgNet:

$$\mathbf{z}_{ij} = \left(\mathbf{I} - \mathbf{G}^\top \left(\beta_t \mathbf{\Sigma}_{ij} + \mathbf{G}\mathbf{G}^\top\right)^{-1} \mathbf{G}\right) \mathbf{y}_{ij}^{t-1}, \text{ where } \mathbf{G} = \mathbf{I} - \frac{1}{d^2}\mathbf{1}\mathbf{1}^\top \text{ is the mean subtraction matrix.} \tag{18}$$

Image formation layer (IF): Compute the new clean image estimate $\mathbf{Y}^t$ using the auxiliary patches $\{\mathbf{z}_{ij}\}$ and the original noisy image $\mathbf{X}$:

$$\mathbf{Y}^t(i, j) = \frac{\mathbf{X}(i, j)}{1 + \beta_t \sigma^2} + \frac{\beta_t \sigma^2}{(1 + \beta_t \sigma^2) d^2} \sum_{p,q=-\lfloor \frac{d-1}{2} \rfloor}^{\lceil \frac{d-1}{2} \rceil} \mathbf{z}_{pq}(i, j). \tag{19}$$

Output: Clean image $\mathbf{Y} = \mathbf{Y}^T$.