Supplemental material for "Supervised Transformer Network for Efficient Face Detection"

Dong Chen, Gang Hua, Fang Wen, and Jian Sun

Microsoft Research
{doch,ganghua,fangwen,jiansun}@microsoft.com

1 The supervised transformer layer

In Section 2.3, we describe the detail of the supervised transformer layer. We need to obtain the derivatives for back-propagation by the chain rule.

$$\begin{aligned}\frac{\partial L}{\partial a} &= \sum_{\{\bar{x},\bar{y}\}} \frac{\partial L}{\partial \bar{I}(\bar{x},\bar{y})} \frac{\partial \bar{I}(\bar{x},\bar{y})}{\partial a} = \sum_{\{\bar{x},\bar{y}\}} \frac{\partial L}{\partial \bar{I}(\bar{x},\bar{y})} \frac{\partial I(x,y)}{\partial a} \\ &= \sum_{\{\bar{x},\bar{y}\}} \frac{\partial L}{\partial \bar{I}(\bar{x},\bar{y})} \left(\frac{\partial I(x,y)}{\partial x} \frac{\partial x}{\partial a} + \frac{\partial I(x,y)}{\partial y} \frac{\partial y}{\partial a} \right) \\ &= \sum_{\{\bar{x},\bar{y}\}} \frac{\partial L}{\partial \bar{I}(\bar{x},\bar{y})} \left(I_x \frac{\partial x}{\partial a} + I_y \frac{\partial y}{\partial a} \right)\end{aligned} \tag{1}$$

Similarly, we can obtain the derivative of other parameters.

$$\frac{\partial L}{\partial b} = \sum_{\{\bar{x},\bar{y}\}} \frac{\partial L}{\partial \bar{I}(\bar{x},\bar{y})} \left(I_x \frac{\partial x}{\partial b} + I_y \frac{\partial y}{\partial b} \right) \tag{2}$$

$$\frac{\partial L}{\partial m_x} = \sum_{\{\bar{x},\bar{y}\}} \frac{\partial L}{\partial \bar{I}(\bar{x},\bar{y})} I_x \tag{3}$$

$$\frac{\partial L}{\partial m_y} = \sum_{\{\bar{x},\bar{y}\}} \frac{\partial L}{\partial \bar{I}(\bar{x},\bar{y})} I_y \tag{4}$$

$$\frac{\partial L}{\partial m_{\bar{x}}} = \sum_{\{\bar{x},\bar{y}\}} \frac{\partial L}{\partial \bar{I}(\bar{x},\bar{y})} \left(-\frac{a}{a^2+b^2} I_x - \frac{b}{a^2+b^2} I_y \right) \tag{5}$$

$$\frac{\partial L}{\partial m_{\bar{y}}} = \sum_{\{\bar{x},\bar{y}\}} \frac{\partial L}{\partial \bar{I}(\bar{x},\bar{y})} \left(\frac{b}{a^2+b^2} I_x - \frac{a}{a^2+b^2} I_y \right) \tag{6}$$

As mentioned in the paper, $\frac{\partial L}{\partial \bar{I}(\bar{x},\bar{y})}$ is the gradient signals back propagated from the RCNN network. The I_x and I_y are horizontal and vertical gradient of the original image.

In Eqn. 1 and 2, the derivative $\frac{\partial x}{\partial a}$, $\frac{\partial x}{\partial b}$, $\frac{\partial y}{\partial a}$ and $\frac{\partial y}{\partial b}$ can be calculated based on Eqn. 4 in the paper.

$$\frac{\partial x}{\partial a} = -\frac{\partial y}{\partial b} = -\frac{a^2-b^2}{(a^2+b^2)^2}(\bar{x}-m_{\bar{x}}) + \frac{2ab}{(a^2+b^2)^2}(\bar{y}-m_{\bar{y}}) \tag{7}$$

$$\frac{\partial x}{\partial b} = \frac{\partial y}{\partial a} = -\frac{2ab}{(a^2+b^2)^2}(\bar{x}-m_{\bar{x}}) - \frac{a^2-b^2}{(a^2+b^2)^2}(\bar{y}-m_{\bar{y}}) \tag{8}$$

Finally we can obtain the gradient of positions of canonical facial landmarks and detected facial landmarks.

$$\left(\frac{\partial L}{\partial x_i}\frac{\partial L}{\partial y_i}\frac{\partial L}{\partial \bar{x}_i}\frac{\partial L}{\partial \bar{y}_i}\right) = G_1 * G_2 * G_3 \tag{9}$$

where

$$G_1 = \left(\frac{\partial L}{\partial a}\ \frac{\partial L}{\partial b}\ \frac{\partial L}{\partial m_x}\ \frac{\partial L}{\partial m_y}\ \frac{\partial L}{\partial m_{\bar{x}}}\ \frac{\partial L}{\partial m_{\bar{y}}}\right) \tag{10}$$

$$G_2 = \begin{pmatrix} \frac{\partial a}{\partial c_1} & \frac{\partial a}{\partial c_2} & \frac{\partial a}{\partial c_3} & \frac{\partial a}{\partial m_x} & \frac{\partial a}{\partial m_y} & \frac{\partial a}{\partial m_{\bar{x}}} & \frac{\partial a}{\partial m_{\bar{y}}} \\ \frac{\partial b}{\partial c_1} & \frac{\partial b}{\partial c_2} & \frac{\partial b}{\partial c_3} & \frac{\partial b}{\partial m_x} & \frac{\partial b}{\partial m_y} & \frac{\partial b}{\partial m_{\bar{x}}} & \frac{\partial b}{\partial m_{\bar{y}}} \\ \frac{\partial m_x}{\partial c_1} & \frac{\partial m_x}{\partial c_2} & \frac{\partial m_x}{\partial c_3} & \frac{\partial m_x}{\partial m_x} & \frac{\partial m_x}{\partial m_y} & \frac{\partial m_x}{\partial m_{\bar{x}}} & \frac{\partial m_x}{\partial m_{\bar{y}}} \\ \frac{\partial m_y}{\partial c_1} & \frac{\partial m_y}{\partial c_2} & \frac{\partial m_y}{\partial c_3} & \frac{\partial m_y}{\partial m_x} & \frac{\partial m_y}{\partial m_y} & \frac{\partial m_y}{\partial m_{\bar{x}}} & \frac{\partial m_y}{\partial m_{\bar{y}}} \\ \frac{\partial m_{\bar{x}}}{\partial c_1} & \frac{\partial m_{\bar{x}}}{\partial c_2} & \frac{\partial m_{\bar{x}}}{\partial c_3} & \frac{\partial m_{\bar{x}}}{\partial m_x} & \frac{\partial m_{\bar{x}}}{\partial m_y} & \frac{\partial m_{\bar{x}}}{\partial m_{\bar{x}}} & \frac{\partial m_{\bar{x}}}{\partial m_{\bar{y}}} \\ \frac{\partial m_{\bar{y}}}{\partial c_1} & \frac{\partial m_{\bar{y}}}{\partial c_2} & \frac{\partial m_{\bar{y}}}{\partial c_3} & \frac{\partial m_{\bar{y}}}{\partial m_x} & \frac{\partial m_{\bar{y}}}{\partial m_y} & \frac{\partial m_{\bar{y}}}{\partial m_{\bar{x}}} & \frac{\partial m_{\bar{y}}}{\partial m_{\bar{y}}} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{c_3} & 0 & -\frac{c_1}{c_3^2} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{c_3} & -\frac{c_2}{c_3^2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{11}$$

$$G_3 = \begin{pmatrix} \frac{\partial c_1}{\partial x_i} & \frac{\partial c_1}{\partial y_i} & \frac{\partial c_1}{\partial \bar{x}_i} & \frac{\partial c_1}{\partial \bar{y}_i} \\ \frac{\partial c_2}{\partial x_i} & \frac{\partial c_2}{\partial y_i} & \frac{\partial c_2}{\partial \bar{x}_i} & \frac{\partial c_2}{\partial \bar{y}_i} \\ \frac{\partial c_3}{\partial x_i} & \frac{\partial c_3}{\partial y_i} & \frac{\partial c_3}{\partial \bar{x}_i} & \frac{\partial c_3}{\partial \bar{y}_i} \\ \frac{\partial m_x}{\partial x_i} & \frac{\partial m_x}{\partial y_i} & \frac{\partial m_x}{\partial \bar{x}_i} & \frac{\partial m_x}{\partial \bar{y}_i} \\ \frac{\partial m_y}{\partial x_i} & \frac{\partial m_y}{\partial y_i} & \frac{\partial m_y}{\partial \bar{x}_i} & \frac{\partial m_y}{\partial \bar{y}_i} \\ \frac{\partial m_{\bar{x}}}{\partial x_i} & \frac{\partial m_{\bar{x}}}{\partial y_i} & \frac{\partial m_{\bar{x}}}{\partial \bar{x}_i} & \frac{\partial m_{\bar{x}}}{\partial \bar{y}_i} \\ \frac{\partial m_{\bar{y}}}{\partial x_i} & \frac{\partial m_{\bar{y}}}{\partial y_i} & \frac{\partial m_{\bar{y}}}{\partial \bar{x}_i} & \frac{\partial m_{\bar{y}}}{\partial \bar{y}_i} \end{pmatrix}$$

$$= \begin{pmatrix} \bar{x}_i - m_{\bar{x}} & \bar{y}_i - m_{\bar{y}} & x_i - m_x & y_i - m_y \\ -(\bar{y}_i - m_{\bar{y}}) & \bar{x} - m_{\bar{x}} & y_i - m_y & -(x_i - m_x) \\ 2(x_i - m_x) & 2(y_i - m_y) & 0 & 0 \\ \frac{1}{N} & 0 & 0 & 0 \\ 0 & \frac{1}{N} & 0 & 0 \\ 0 & 0 & \frac{1}{N} & 0 \\ 0 & 0 & 0 & \frac{1}{N} \end{pmatrix} \tag{12}$$

Fig. 1. Some unlabeled faces detected by our detector (shown in green rectangles). (a) FDDB, (b) AFW, (c) PASCAL faces datasets.

In Eqn. 10, each element of G_1 can be obtained from Eqn. $1 \sim 6$. In Eqn. 12, N is the facial points number. The proposed Supervised Transformer layer can be implemented very efficiently on GPU, and can be put between the RPN and RCNN network in the end-to-end training framework.

2 More face detection results

In this section, we provide more qualitative face detection result on FDDB, AFW and PASCAL datasets. In Figure 1, it shows some unlabeled faces detected by the proposed face detector. Figure 2 shows some faces missed by our detector. They mainly case by small or blur faces, large pose variations, and large occlusions.

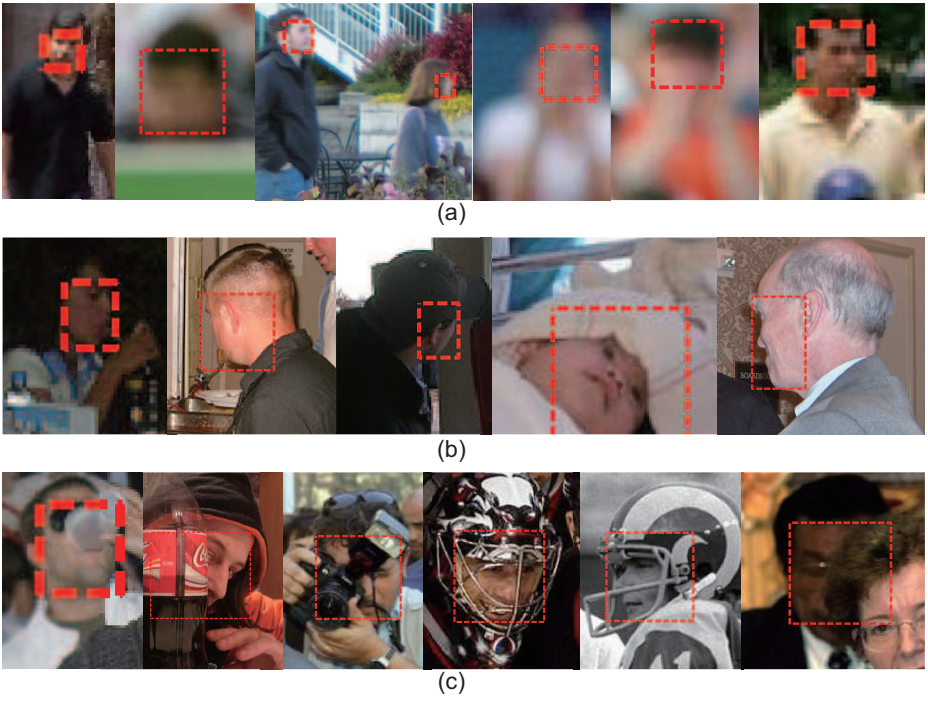

Fig. 2. Some faces missed by our detector. They mainly case by (a) small or blur faces, (b) large pose variations, (c) large occlusions.