Loss Functions for Image Restoration with Neural Networks

Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz

NVIDIA Research and MIT Media Lab

1 Computing the derivatives for the different loss functions we propose to use

In this section, we provide more details about how the derivatives of the different loss functions, specifically the derivatives of SSIM and MS-SSIM, as the other losses are either trivial or a direct consequence of these two.

1.1 The derivatives of SSIM

As indicated in the main paper, we can write SSIM for a pixel p as:

$$\text{SSIM}(p) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \cdot \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} = l(p) \cdot cs(p) \tag{1}$$

where we omitted the dependence of means and standard deviations on pixel p. Means and standard deviations are computed with a Gaussian filter with standard deviation σ_G:

$$\mu_x(p) = G_{\sigma_G} * P_x \tag{2}$$
$$\sigma_x^2(p) = G_{\sigma_G} * P_x^2 \quad - \mu_x^2(p) \tag{3}$$
$$\sigma_{xy}(p) = G_{\sigma_G} * (P_x \cdot P_y) \quad - \mu_x(p)\mu_y(p), \tag{4}$$

where P_x is a patch centered a pixel p, '$*$' denotes a convolution, and '$\cdot$' a point-wise multiplication. The values of $\mu_y(p)$ and $\sigma_y^2(p)$ can be computed similarly to Equations 2 and 3. For convenience we first compute the derivatives of the terms in Equations 2–4. For each term, we need to compute the derivatives at pixel p with respect to all the pixels q in patch P:

$$\frac{\partial \mu_x(p)}{\partial x(q)} = \frac{\partial}{\partial x(q)} \sum_{\hat{p} \in P} G_{\sigma_G}(p - \hat{p}) \cdot x(\hat{p}) = G_{\sigma_G}(p - q), \tag{5}$$

where we used the observation that $\partial \mu_x(p)/\partial x(q) = 0, \forall p \neq q$. Equation 5 shows that the derivatives with respect to all of the pixels in the patch are the coefficients of the

Gaussian filter. Similarly,

$$\begin{aligned}\frac{\partial \sigma_x^2(p)}{\partial x(q)} &= \frac{\partial}{\partial x(q)}\left[\sum_{\hat{p}\in P} G_{\sigma_G}(p-\hat{p})\cdot x(\hat{p})^2 \; - \mu_x^2(p)\right] \\ &= \frac{\partial}{\partial x(q)}\left[\sum_{\hat{p}\in P} G_{\sigma_G}(p-\hat{p})\cdot x(\hat{p})^2\right] - 2\mu_x(p)\frac{\partial \mu_x(p)}{\partial x(q)} \\ &= 2G_{\sigma_G}(p-q)\cdot x(q) \; - \; 2\mu_x(p)\cdot G_{\sigma_G}(p-q) \\ &= 2G_{\sigma_G}(p-q)\cdot (x(q)-\mu_x(p))\,, \end{aligned} \tag{6}$$

and

$$\begin{aligned}\frac{\partial \sigma_{xy}(p)}{\partial x(q)} &= \frac{\partial}{\partial x(q)}\left[\sum_{\hat{p}\in P} G_{\sigma_G}(p-\hat{p})\cdot x(\hat{p})y(\hat{p})\right] - \mu_y(p)\frac{\partial \mu_x(p)}{\partial x(q)} \\ &= G_{\sigma_G}(p-q)\cdot y(q) \; - \; \mu_y(p)\cdot G_{\sigma_G}(p-q) \\ &= G_{\sigma_G}(p-q)\cdot (y(q)-\mu_y(p))\,. \end{aligned} \tag{7}$$

With reference to Equation 1, we now need to compute $\partial l(p)/\partial x(q)$ and $\partial cs(p)/\partial x(q)$. Omitting the dependence of the means and variances on pixel p for clarity, we can write

$$\begin{aligned}\frac{\partial l(p)}{\partial x(q)} &= \frac{\partial l(p)}{\partial \mu_x}\cdot\frac{\partial \mu_x}{\partial x(q)} \\ &= \left(\frac{2\mu_y}{\mu_x^2+\mu_y^2+C_1} - \frac{2\mu_x\mu_y+C_1}{(\mu_x^2+\mu_y^2+C_1)^2}\cdot 2\mu_x\right)\cdot\frac{\partial \mu_x}{\partial x(q)} \\ &= 2\left(\frac{\mu_y-\mu_x\cdot l(p)}{\mu_x^2+\mu_y^2+C_1}\right)\cdot\frac{\partial \mu_x}{\partial x(q)} \\ &= 2G_{\sigma_G}(p-q)\cdot\left(\frac{\mu_y-\mu_x\cdot l(p)}{\mu_x^2+\mu_y^2+C_1}\right), \end{aligned} \tag{8}$$

and

$$\begin{aligned}\frac{\partial cs(p)}{\partial x(q)} &= \frac{2\frac{\partial \sigma_{xy}}{\partial x(q)}}{\sigma_x^2+\sigma_y^2+C_2} - \frac{(2\sigma_{xy}+C_2)\cdot\frac{\partial \sigma_x^2}{\partial x(q)}}{(\sigma_x^2+\sigma_y^2+C_2)^2} \\ &= \frac{1}{\sigma_x^2+\sigma_y^2+C_2}\left(2\frac{\partial \sigma_{xy}}{\partial x(q)} - cs(p)\cdot\frac{\partial \sigma_x^2}{\partial x(q)}\right) \\ &= \frac{1}{\sigma_x^2+\sigma_y^2+C_2}\big(2G_{\sigma_G}(p-q)\cdot(y(q)-\mu_y) \\ &\quad -cs(p)\cdot 2G_{\sigma_G}(p-q)\cdot(x(q)-\mu_x)\big) \\ &= \frac{2}{\sigma_x^2+\sigma_y^2+C_2}\cdot G_{\sigma_G}(p-q)\cdot \\ &\quad \cdot\big((y(q)-\mu_y) - cs(p)\cdot(x(q)-\mu_x)\big)\,. \end{aligned} \tag{9}$$

Now we can put everything together:

$$\frac{\partial}{\partial x(q)}\text{SSIM}(p) = \frac{\partial l(p)}{\partial x(q)} \cdot cs(p) + l(p) \cdot \frac{\partial cs(p)}{\partial x(q)}. \tag{10}$$

1.2 The derivatives of MS-SSIM

Multi-scale SSIM can be defined as:

$$\text{MS-SSIM}(p) = l_M(p)^{\alpha} \cdot \prod_{j=1}^{M} cs_j^{\beta_j}(p), \tag{11}$$

where we set $\alpha = \beta_j = 1, \forall j \in 1, .., M$. The derivative of Equation 11 can be written as

$$\begin{aligned}
\frac{\partial}{\partial x(q)}\text{MS-SSIM}(p) &= \frac{\partial l(p)}{\partial x(q)} \cdot \prod_{j=1}^{M} cs_j(p) + l(p) \cdot \frac{\partial}{\partial x(q)} \prod_{j=1}^{M} cs_j(p) \\
&= \frac{\partial l(p)}{\partial x(q)} \cdot \prod_{j=1}^{M} cs_j(p) + l(p) \cdot \sum_{i=1}^{M} \frac{\partial cs_i(p)}{\partial x(q)} \prod_{j \neq i} cs_j(p) \\
&= \frac{\partial l(p)}{\partial x(q)} \cdot \prod_{j=1}^{M} cs_j(p) + l(p) \cdot \sum_{i=1}^{M} \frac{1}{cs_i(p)} \frac{\partial cs_i(p)}{\partial x(q)} \prod_{j=1}^{M} cs_j(p) \\
&= \left(\frac{\partial l(p)}{\partial x(q)} + l(p) \cdot \sum_{i=1}^{M} \frac{1}{cs_i(p)} \frac{\partial cs_i(p)}{\partial x(q)} \right) \cdot \prod_{j=1}^{M} cs_j(p),
\end{aligned}$$

where the derivatives of l and cs are the same as in previous section.