Supplementary file: Learning a Discriminative Model for the Perception of Realism using Composite Images

In this supplementary document, we derive the gradient of our objective function $\frac{\partial E}{\partial g}$ used in Section 4 in the main paper.

Recall that the object composition process is formulated as $I_g = \alpha \cdot g(F) + (1 - \alpha) \cdot B$ where F is the source object, B is the background scene, and $\alpha \in [0, 1]$ is the alpha mask for the foreground object. We denote F^p as the RGB color values for each pixel p in the foreground object, α^p as its alpha channel and B^p as the color values in the background image. We denote I_g^p as the RGB values for pixel p in the composite image.

The color adjustment model $g(\cdot)$ adjusts the visual properties of the foreground to be compatible with background image. Our objective function for color adjustment is written as follows:

$$E(g, F) = -f(I_g; \theta) + w \cdot E_{reg}(g), \tag{1}$$

where f measures the visual realism of the composite and E_{reg} imposes a regularizer on the space of possible adjustments. The weight w controls the relative importance between the two terms. We apply a very simple brightness and contrast model to the source object F for each channel independently. For each pixel we map the foreground color values $F^p = (c_1^p, c_2^p, c_3^p)$ to $g(F^p) = (\lambda_1 c_1^p + \beta_1, \lambda_2 c_2^p + \beta_2, \lambda_3 c_3^p + \beta_3)$. The regularization term for this model is formulated as:

$$E_{reg}(g) = \frac{1}{N} \sum_p \Big(\|I_g^p - I_0^p\|_2 + \sum_{i,j} \|(\lambda_i - 1) \cdot c_i^p + \beta_i - (\lambda_j - 1) \cdot c_j^p - \beta_j\|_2 \Big) \tag{2}$$

where N is the number of foreground pixels in the image, and $I_0 = \alpha \cdot F + (1 - \alpha) \cdot B$ is the composite image without recoloring, I_g^p and I_0^p denotes the color values for pixel p in the composite image. The first term penalizes large change between the original object and recolored object, and the second term discourages independent color channel variations (roughly hue change). In practice, we normalize these two terms so that they have a similar magnitude.

We would like to optimize the color adjustment function $g^* = \arg\min_g E(g, F)$. Our objective (Equation 1) is differentiable if the color adjustment function g is also differentiable. This allows us to optimize for the color adjustment using gradient-descent.

To optimize the function, we decompose the gradient into the following two terms:

$$\begin{aligned} \frac{\partial E}{\partial g} &= -\frac{\partial f(I_g, \theta)}{\partial g} + \frac{\partial E_{reg}}{\partial g} \\ &= -\sum_p \frac{\partial f(I_g, \theta)}{\partial I_g^p} \cdot \frac{\partial I_g^p}{\partial g} + \frac{\partial E_{reg}}{\partial g} \end{aligned}$$

Notice that $-\frac{\partial f(I_g, \theta)}{\partial I_g^p} = (\Delta_1^p, \Delta_2^p, \Delta_3^p)$ can be computed through backpropagation of CNN model from the loss layer to the image layer. Given this, we can now give the closed form of gradient for each parameter separately $(\lambda_1, \lambda_2, \lambda_3, \beta_1, \beta_2, \beta_3)$. For the first part:

$$\begin{aligned} \frac{\partial f(I_g, \theta)}{\partial \lambda_i} &= \sum_p \alpha^p \cdot c_i^p \cdot \Delta_i^p \\ \frac{\partial f(I_g, \theta)}{\partial \beta_i} &= \sum_p \alpha^p \cdot \Delta_i^p \end{aligned}$$

For the second part, we have:

$$\begin{aligned} \frac{1}{2} \frac{\partial E_{reg}}{\partial \lambda_i} &= \frac{1}{N} \sum_p \alpha^p \cdot c_i^p \cdot \Big(\lambda_i c_i^p + \beta_i - c_i^p + \sum_{j \neq i} [(\lambda_i - 1) \cdot c_i^p + \beta_i - (\lambda_j - 1) \cdot c_j^p - \beta_j] \Big) \\ \frac{1}{2} \frac{\partial E_{reg}}{\partial \beta_i} &= \frac{1}{N} \sum_p \alpha^p \cdot \Big(\lambda_i c_i^p + \beta_i - c_i^p + \sum_{j \neq i} [(\lambda_i - 1) \cdot c_i^p + \beta_i - (\lambda_j - 1) \cdot c_j^p - \beta_j] \Big) \end{aligned}$$

We optimize the cost function using L-BFGS-B [1]. We set the search range of color mapping parameters α_i to $[0.4, 2.0]$, and β_i to $[-127, 127]$. Since the objective is non-convex, we start from multiple random initializations. We initialize the color bias terms β_i to 0 and sample the color gain values λ_i from $[0.6, 0.8, 1.0, 1.2, 1.4]$. We output the solution with the minimal cost.

References

[1] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208, 1995.