NeRFRenderCore Loss Function Derivation

By James Perlman, January 3, 2023

Problem: Given the volumetric rendering function and the network loss function, evaluate the gradient of the loss with respect to the network's output, dL/doutput.

Volumetric Rendering Function

A ray's accumulated color R can be expressed as the sum of the products of the transmittance T_i, the alpha α_i, and the predicted sample color c_i (adapted from [Mildenhall, et al. NeRF, page 6])

(1) $R_r = \sum_{i=1}^{N} T_i \alpha_i c_{ri}$

(2) $\alpha_i = 1 - exp\left(-\sigma_i \delta_i\right)$

(3) $T_i = exp\left(-\sum_{j=1}^{i} \sigma_j \delta_j\right)$

Where:

R_r is the accumulated ray color (per each color component, r, g, and b)

N is the number of samples along the ray

i is the sample index along the ray

c_{ri} is the color network's output at sample i (per each color component, r, g, and b)

σ_i is the density network's output at sample i

δ_i is the difference between the ray's start and end points $t_{i+1} - t_i$

- A ray's position given t is defined by $r_x(t) = o_x + td_x$, per component x,y,z

A slightly special case comes for calculating the ray's accumulated opacity, since the color network does not output an alpha channel we assume $c_{\alpha i} = 1$ for each sample i:

(4) $R_\alpha = \sum_{i=1}^{N} T_i \alpha_i c_{\alpha i} = \sum_{i=1}^{N} T_i \alpha_i$

Loss Function

The network's loss functions can be calculated per color channel:

(5) $L_r = \frac{1}{M}\sum_{k=1}^{M}\left(R_{rk} - gt_{rk}\right)^2$ ← for channels r, g, and b

(6) $L_\alpha = \frac{1}{M}\sum_{k=1}^{M}\left(R_{\alpha k} - 1\right)^2$ ← for alpha channel*

** Assuming ground-truth alpha is 1 for each training pixel. We could potentially use ground truth alpha here if we have it, or weighted masking for painting out features.*

Where:
M is the number of rays in the training batch
R_{rk} is the accumulated color for ray k (per channel, r, g, b)
gt_{rk} is the ground-truth pixel color corresponding to ray k (per channel, r, g, b)

The total network loss is the average of each loss per channel:

(7) $L = \frac{1}{4}\left(L_r + L_g + L_b + L_\alpha\right)$

Deriving dL/doutput

Now, we need to determine the gradient, which is just the partial derivative of each network output sample with respect to the total network loss. We will now see how these results were obtained:

$$\frac{dL}{dc_{rki}} = \frac{\left(R_{rk} - gt_{rk}\right)}{2M} \cdot T_{ki}\alpha_{ki}$$
$$\frac{dL}{dc_{gki}} = \frac{\left(R_{gk} - gt_{gk}\right)}{2M} \cdot T_{ki}\alpha_{ki}$$
$$\frac{dL}{dc_{bki}} = \frac{\left(R_{bk} - gt_{bk}\right)}{2M} \cdot T_{ki}\alpha_{ki}$$
$$\frac{dL}{d\sigma_{ki}} = \frac{1}{2M} \cdot \left(1 - 2\alpha_{ki}\right) \cdot \left(\delta_{ki}T_{ki}\right) \cdot \left[\left(R_{rk} - gt_{rk}\right)c_{rki} + \left(R_{gk} - gt_{gk}\right)c_{gki} + \left(R_{bk} - gt_{bk}\right)c_{bki} + \left(R_{\alpha k} - 1\right)\right]$$

Here, the notation c_{rki} signifies the color sample of the red channel at index i of ray k, where:

$i \in [1, N_k]$ – N_k is the number of samples at ray k

$k \in [1, M]$ – M is the number of rays in the training batch.

The color samples will all be similar per each channel, so let's just focus on red for now.

$$\frac{dL}{dc_{rki}} = \frac{d}{dc_{rki}}L = \frac{d}{dc_{rki}}\left[\frac{1}{4}\left(L_r + L_g + L_b + L_\alpha\right)\right]$$

Since c_{rki} only contributes to L_r, that means $\frac{d}{dc_{rki}}L_g = \frac{d}{dc_{rki}}L_b = \frac{d}{dc_{rki}}L_\alpha = 0$

$$\frac{dL}{dc_{rki}} = \frac{1}{4} \cdot \frac{d}{dc_{rki}}\left[L_r\right]$$

Expanding L_r we can write:

$$\frac{dL}{dc_{rki}} = \frac{1}{4} \cdot \frac{d}{dc_{rki}}\left[\frac{1}{M}\sum_{m=1}^{M}\left(R_{rm} - gt_{rm}\right)^2\right]$$
$$\frac{dL}{dc_{rki}} = \frac{1}{4M} \cdot \frac{d}{dc_{rki}}\left[\sum_{m=1}^{M}\left(R_{rm} - gt_{rm}\right)^2\right]$$

If we expand the summation, we can find more terms whose partial derivatives evaluate to zero:

$$\frac{dL}{dc_{rki}} = \frac{1}{4M} \cdot \frac{d}{dc_{rki}}\left[\left(R_{r1} - gt_{r1}\right)^2 + \ \dots\ + \left(R_{rk} - gt_{rk}\right)^2 + \ \dots\ + \left(R_{rM} - gt_{rM}\right)^2\right]$$

Any term that does not contain R_k or gt_k will not contain c_{rki} and thus will evaluate to zero.

$$\frac{dL}{dc_{rki}} = \frac{1}{4M} \cdot \frac{d}{dc_{rki}}\left[\left(R_{rk} - gt_{rk}\right)^2\right]$$

Using the chain rule, we can rewrite this as:

$$\frac{dL}{dc_{rki}} = \frac{1}{4M} \cdot \left[2\left(R_{rk} - gt_{rk}\right) \cdot \frac{d}{dc_{rki}}R_{rk}\right]$$

And simplify:

$$(8)\ \frac{dL}{dc_{rki}} = \frac{\left(R_{rk} - gt_{rk}\right)}{2M} \cdot \frac{d}{dc_{rki}}R_{rk}$$

Let's focus on evaluating $\frac{d}{dc_{rki}}R_{rk}$, and make a substitution from section 1:

$$R_{rk} = \sum_{i=1}^{N_k} T_{ki}\alpha_{ki}c_{rki}$$

This is a modified form of the equation from section 1, where N_k is the number of samples along ray k, T_{ki} is the transmittance of sample i on ray k, etc…

Until further noted, let's drop the k subscript while evaluating $\frac{d}{dc_{rki}}R_k$ – all samples here exist along ray k so it will reduce visual clutter to leave it out.

$$\frac{d}{dc_{rki}}R_{rk} = \frac{d}{dc_{ri}}R_r = \frac{d}{dc_{ri}}\left[\sum_{n=1}^{N} T_n\alpha_n c_{rn}\right]$$

If we expand this summation, we can write:

$$\frac{d}{dc_{ri}}R_r = \frac{d}{dc_{ri}}\left[T_1\alpha_1 c_{r1} + \ \dots\ + T_i\alpha_i c_{ri} + \ \dots\ + T_N\alpha_N c_{rN}\right]$$

And we can drop terms that do not contain c_{ri} since their partial derivatives will be zero:

$$\frac{d}{dc_{ri}}R_r = \frac{d}{dc_{ri}}\left[T_i\alpha_i c_{ri}\right]$$

Using the chain rule, we can write this as:

$$(9)\ \frac{d}{dc_{ri}}R_r = \left(\frac{d}{dc_{ri}}T_i \cdot \alpha_i \cdot c_{ri}\right) + \left(T_i \cdot \frac{d}{dc_{ri}}\alpha_i \cdot c_{ri}\right) + \left(T_i \cdot \alpha_i \cdot \frac{d}{dc_{ri}}c_{ri}\right)$$

Let's work on each of these $\frac{d}{dc_{ri}}$ terms individually.

(a) $\frac{d}{dc_{ri}} T_i$

This is the derivative of the transmittance of sample i with respect to the intensity of the red channel. Expanding T_i to its full form as in section 1, we can write:

$$\frac{d}{dc_{ri}} T_i = \frac{d}{dc_{ri}} \left[exp\left(- \sum_{j=1}^{i} \sigma_j \delta_j \right) \right] = 0$$

T_i does not depend on c_{ri}, so its derivative is zero.

(b) $\frac{d}{dc_{ri}} \alpha_i$

This is the derivative of the alpha of sample i with respect to the intensity of the red channel. Substituting again from section 1, we can write:

$$\frac{d}{dc_{ri}} \alpha_i = \frac{d}{dc_{ri}} \left[1 - exp\left(- \sigma_i \delta_i \right) \right] = 0$$

Again, this expression has no correlation with c_{ri}, so its derivative evaluates to zero.

(c) $\frac{d}{dc_{ri}} c_{ri} = 1$

Finally, we have the derivative of c_{ri} with respect to itself, which is just 1. Now, substituting these terms back into equation (9) we have:

$$\frac{d}{dc_{ri}} R_r = \left(0 \cdot \alpha_i \cdot c_{ri} \right) + \left(T_i \cdot 0 \cdot c_{ri} \right) + \left(T_i \cdot \alpha_i \cdot 1 \right) = T_i \alpha_i$$

This is identical to expressing the weight for sample i ([pg. 8 of Mildenhall, et al. "NeRF"]). It is constant with regard to c_{ri}. Expanding terms to their full form:

$$\frac{d}{dc_{ri}} R_r = exp\left(- \sum_{j=1}^{i} \sigma_j \delta_j \right) \cdot \left(1 - exp\left(- \sigma_i \delta_i \right) \right)$$

Now we may reintroduce the k subscript:

$$\frac{d}{dc_{rki}} R_{rk} = exp\left(- \sum_{j=1}^{i} \sigma_{kj} \delta_{kj}\right) \cdot \left(1 - exp\left(- \sigma_{ki} \delta_{ki}\right)\right) = T_{ki} \alpha_{ki}$$

And place this back where it belongs, in equation (8):

$$\frac{dL}{dc_{rki}} = \frac{\left(R_{rk} - gt_{rk}\right)}{2M} \cdot exp\left(- \sum_{j=1}^{i} \sigma_{j} \delta_{j}\right) \cdot \left(1 - exp\left(- \sigma_{i} \delta_{i}\right)\right)$$

Which can also be written:

$$\frac{dL}{dc_{rki}} = \frac{\left(R_{rk} - gt_{rk}\right)}{2M} \cdot T_{ki} \alpha_{ki}$$

This process can be repeated to obtain the equations for the green and blue channels as well.

$$\frac{dL}{dc_{gki}} = \frac{\left(R_{gk} - gt_{gk}\right)}{2M} \cdot T_{ki} \alpha_{ki}$$

$$\frac{dL}{dc_{bki}} = \frac{\left(R_{bk} - gt_{bk}\right)}{2M} \cdot T_{ki} \alpha_{ki}$$

Now, let's move on to differentiating with respect to sigma (σ_{ki})...

Deriving $\frac{dL}{d\sigma_{ki}}$

Let's start with the total loss function.

$$\frac{dL}{d\sigma_{ki}} = \frac{d}{d\sigma_{ki}} L = \frac{d}{d\sigma_{ki}}\left[\frac{1}{4}\left(L_r + L_g + L_b + L_\alpha\right)\right]$$

We can move some symbols around to obtain the following:

$$\text{(10)} \qquad \frac{dL}{d\sigma_{ki}} = \frac{1}{4}\left[\frac{d}{d\sigma_{ki}}L_r + \frac{d}{d\sigma_{ki}}L_g + \frac{d}{d\sigma_{ki}}L_b + \frac{d}{d\sigma_{ki}}L_\alpha\right]$$

The partial derivatives for each color channel will have the same form, so again we can focus on just the red channel first and extrapolate later. Let's focus on $\frac{d}{d\sigma_{ki}}L_r$ first:

$$\frac{d}{d\sigma_{ki}}L_r = \frac{d}{d\sigma_{ki}}\left[\frac{1}{M}\sum_{m=1}^{M}\left(R_{rm} - gt_{rm}\right)^2\right]$$

Expanding summation terms and using the distributive property, we can rewrite this as:

$$\frac{d}{d\sigma_{ki}}L_r = \frac{1}{M} \cdot \frac{d}{d\sigma_{ki}}\left[\left(R_{r1} - gt_{r1}\right)^2 + \ \dots \ + \left(R_{rk} - gt_{rk}\right)^2 + \ \dots \ + \left(R_{rM} - gt_{rM}\right)^2\right]$$

All derivatives without the rk subscript will evaluate to zero:

$$\frac{d}{d\sigma_{ki}}L_r = \frac{1}{M} \cdot \frac{d}{d\sigma_{ki}}\left(R_{rk} - gt_{rk}\right)^2$$

And using the chain rule, we have:

$$\frac{d}{d\sigma_{ki}}L_r = \frac{2}{M} \cdot \left(R_{rk} - gt_{rk}\right) \cdot \frac{d}{d\sigma_{ki}}\left(R_{rk} - gt_{rk}\right)$$

Which simplifies to:

$$\text{(11)} \qquad \frac{d}{d\sigma_{ki}}L_r = \frac{2}{M} \cdot \left(R_{rk} - gt_{rk}\right) \cdot \frac{d}{d\sigma_{ki}}R_{rk}$$

Now let's just focus on $\frac{d}{d\sigma_{ki}} R_{rk}$, bringing in equation (1) and dropping the k subscript for now:

$$\frac{d}{d\sigma_i} R_r = \frac{d}{d\sigma_i}\left[\sum_{n=1}^{N} T_n \alpha_n c_{rn}\right]$$

Expanding summation terms, we can clean this up a bit:

$$\frac{d}{d\sigma_i} R_r = \frac{d}{d\sigma_i}\left[T_1 \alpha_1 c_{r1} + \ \dots \ + T_i \alpha_i c_{ri} + \ \dots \ + T_N \alpha_N c_{rN}\right]$$

$$\frac{d}{d\sigma_i} R_r = \frac{d}{d\sigma_i}\left[T_i \alpha_i c_{ri}\right]$$

And applying the chain rule, we have:

$$\frac{d}{d\sigma_i} R_r = \left(\frac{d}{d\sigma_i} T_i \cdot \alpha_i \cdot \ c_{ri}\right) + \left(T_i \cdot \frac{d}{d\sigma_i} \alpha_i \cdot \ c_{ri}\right) + \left(T_i \cdot \alpha_i \cdot \frac{d}{d\sigma_i} c_{ri}\right) \tag{12}$$

Now let's break this down further, and solve for each of our 3 partial derivative terms.
Substituting from equation (3):

$$\frac{d}{d\sigma_i} T_i = \frac{d}{d\sigma_i} exp\left(- \sum_{j=1}^{i} \sigma_j \delta_j\right)$$

Expand!

$$\frac{d}{d\sigma_i} T_i = \frac{d}{d\sigma_i} exp\left(- \left[\sigma_1 \delta_1 + \ \dots \ + \sigma_i \delta_i\right]\right)$$

Chain rule!

$$\frac{d}{d\sigma_i} T_i = \left(- \ \delta_i\right) \cdot \ exp\left(- \left[\sigma_1 \delta_1 + \ \dots \ + \sigma_i \delta_i\right]\right)$$

Contract!

$$\frac{d}{d\sigma_i} T_i = \left(- \ \delta_i\right) \cdot \ exp\left(- \sum_{j=1}^{i} \sigma_j \delta_j\right)$$

Next, using equation (2) to substitute:

$$\frac{d}{d\sigma_i}\alpha_i = \frac{d}{d\sigma_i}\left[1 - exp\left(-\sigma_i\delta_i\right)\right]$$

This is a pretty easy one:

$$\frac{d}{d\sigma_i}\alpha_i = \delta_i exp\left(-\sigma_i\delta_i\right)$$

And finally, the last of the 3 partial derivatives from equation (12):

$$\frac{d}{d\sigma_{ki}}c_{ri} = 0$$

Now let's put everything back into equation (12):

$$\frac{d}{d\sigma_i}R_r = \left(\left[\left(-\delta_i\right)\cdot exp\left(-\sum_{j=1}^{i}\sigma_j\delta_j\right)\right]\cdot \alpha_i \cdot c_{ri}\right) + \left(T_i \cdot \left[\delta_i exp\left(-\sigma_i\delta_i\right)\right]\cdot c_{ri}\right) + \left(T_i \cdot \alpha_i \cdot 0\right)$$

Let's make this prettier to look at by eliminating the unneeded term:

$$\frac{d}{d\sigma_i}R_r = \left(\left[\left(-\delta_i\right)\cdot exp\left(-\sum_{j=1}^{i}\sigma_j\delta_j\right)\right]\cdot \alpha_i \cdot c_{ri}\right) + \left(T_i \cdot \left[\delta_i exp\left(-\sigma_i\delta_i\right)\right]\cdot c_{ri}\right)$$

Factoring out δ_i and c_{ri}:

$$\frac{d}{d\sigma_i}R_r = \delta_i c_{ri} \cdot \left[\left(-exp\left(-\sum_{j=1}^{i}\sigma_j\delta_j\right)\cdot \alpha_i\right) + \left(T_i \cdot exp\left(-\sigma_i\delta_i\right)\right)\right]$$

Replacing α_i with equation (2) and T_i with equation (3):

$$\frac{d}{d\sigma_i}R_r = \delta_i c_{ri} \cdot \left[\left(-exp\left(-\sum_{j=1}^{i}\sigma_j\delta_j\right)\cdot \left(1 - exp\left(-\sigma_i\delta_i\right)\right)\right) + \left(exp\left(-\sum_{j=1}^{i}\sigma_j\delta_j\right)\cdot exp\left(-\sigma_i\delta_i\right)\right)\right]$$

Factoring out $exp\left(-\sum_{j=1}^{i}\sigma_j\delta_j\right)$:

$$\frac{d}{d\sigma_i}R_r = \delta_i c_{ri} \cdot exp\left(-\sum_{j=1}^{i}\sigma_j\delta_j\right)\cdot \left[\left(-\left(1 - exp\left(-\sigma_i\delta_i\right)\right)\right) + exp\left(-\sigma_i\delta_i\right)\right]$$

Removing parentheses:

$$\frac{d}{d\sigma_i}R_r = \delta_i c_{ri} \cdot exp\left(-\sum_{j=1}^{i}\sigma_j\delta_j\right)\cdot\left[-1 + exp\left(-\sigma_i\delta_i\right) + exp\left(-\sigma_i\delta_i\right)\right]$$

And we are left with this:

$$\frac{d}{d\sigma_i}R_r = \delta_i c_{ri} \cdot exp\left(-\sum_{j=1}^{i}\sigma_j\delta_j\right)\cdot\left[2 \cdot exp\left(-\sigma_i\delta_i\right) - 1\right]$$

Which can also be rewritten as:

$$\frac{d}{d\sigma_i}R_r = \delta_i c_{ri} \cdot exp\left(-\sum_{j=1}^{i}\sigma_j\delta_j\right)\cdot\left[-1 \cdot \left[1 - 2exp\left(-\sigma_i\delta_i\right) + 1 - 1\right]\right]$$

$$\frac{d}{d\sigma_i}R_r = \delta_i c_{ri} \cdot exp\left(-\sum_{j=1}^{i}\sigma_j\delta_j\right)\cdot\left[-1 \cdot \left[2 - 2exp\left(-\sigma_i\delta_i\right) - 1\right]\right]$$

$$\frac{d}{d\sigma_i}R_r = \delta_i c_{ri} \cdot exp\left(-\sum_{j=1}^{i}\sigma_j\delta_j\right)\cdot\left[-1 \cdot \left[2\left(1 - exp\left(-\sigma_i\delta_i\right)\right) - 1\right]\right]$$

Or:

$$\frac{d}{d\sigma_i}R_r = \delta_i c_{ri} \cdot T_i \cdot \left[-1 \cdot \left[2\alpha_i - 1\right]\right]$$

And finally:

$$\frac{d}{d\sigma_i}R_r = \delta_i c_{ri} \cdot T_i \cdot \left(1 - 2\alpha_i\right)$$

Now we can plug this back into equation 11, and bring back the k subscript:

$$\frac{d}{d\sigma_{ki}}R_{rk} = \delta_{ki} c_{rki} \cdot T_{ki} \cdot \left(1 - 2\alpha_{ki}\right)$$

$$\frac{d}{d\sigma_{ki}}L_r = \frac{2}{M} \cdot \left(R_{rk} - gt_{rk}\right) \cdot \frac{d}{d\sigma_{ki}}R_{rk}$$

$$\frac{d}{d\sigma_{ki}}L_r = \frac{2}{M} \cdot \left(R_{rk} - gt_{rk}\right) \cdot \left(\delta_{ki} c_{rki} \cdot T_{ki} \cdot \left(1 - 2\alpha_{ki}\right)\right)$$

$$\frac{d}{d\sigma_{ki}}L_r = \frac{2}{M} \cdot \left(R_{rk} - gt_{rk}\right) \cdot \left(\delta_{ki} c_{rki} T_{ki}\right) \cdot \left(1 - 2\alpha_{ki}\right)$$

This process is similar for $\frac{d}{d\sigma_{ki}} L_g$, and $\frac{d}{d\sigma_{ki}} L_b$, and we should obtain:

$$\frac{d}{d\sigma_{ki}} L_g = \frac{2}{M} \cdot \left(R_{gk} - gt_{gk}\right) \cdot \left(\delta_{ki} c_{gki} T_{ki}\right) \cdot \left(1 - 2\alpha_{ki}\right)$$

$$\frac{d}{d\sigma_{ki}} L_b = \frac{2}{M} \cdot \left(R_{bk} - gt_{bk}\right) \cdot \left(\delta_{ki} c_{bki} T_{ki}\right) \cdot \left(1 - 2\alpha_{ki}\right)$$

Now we move on to evaluating $\frac{d}{d\sigma_{ki}} L_{\alpha}$ from equation (10). From equation (6) we can make the substitution:

$$\frac{d}{d\sigma_{ki}} L_{\alpha} = \frac{d}{d\sigma_{ki}} \left[\frac{1}{M} \sum_{m=1}^{M} \left(R_{\alpha m} - 1\right)^2 \right]$$

Reorganizing and expanding, we see that this is equivalent to:

$$\frac{d}{d\sigma_{ki}} L_{\alpha} = \frac{1}{M} \cdot \frac{d}{d\sigma_{ki}} \left[\left(R_{\alpha 1} - 1\right)^2 + \ ... \ + \ \left(R_{\alpha k} - 1\right)^2 + \ ... \ + \left(R_{\alpha M} - 1\right)^2 \right]$$

Again, terms without any reference to ray k can be eliminated:

$$\frac{d}{d\sigma_{ki}} L_{\alpha} = \frac{1}{M} \cdot \frac{d}{d\sigma_{ki}} \left[\left(R_{\alpha k} - 1\right)^2 \right]$$

Using the chain rule:

$$\frac{d}{d\sigma_{ki}} L_{\alpha} = \frac{1}{M} \cdot \left[2 \cdot \left(R_{\alpha k} - 1\right) \cdot \frac{d}{d\sigma_{ki}} R_{\alpha k} \right]$$

And we have:

$$(13) \qquad \frac{d}{d\sigma_{ki}} L_{\alpha} = \frac{2}{M} \cdot \left(R_{\alpha k} - 1\right) \cdot \frac{d}{d\sigma_{ki}} R_{\alpha k}$$

Focusing on $\frac{d}{d\sigma_{ki}} R_{\alpha k}$, dropping the k subscript, and substituting from (4):

$$\frac{d}{d\sigma_{i}} R_{\alpha} = \frac{d}{d\sigma_{i}} \sum_{n=1}^{N} T_{n} \alpha_{n}$$

Expanding terms:

$$\frac{d}{d\sigma_{i}} R_{\alpha} = \frac{d}{d\sigma_{i}} \left[T_{1} \alpha_{1} + \ ... \ + T_{i} \alpha_{i} + \ ... \ + T_{N} \alpha_{N} \right]$$

Expanded elements representing samples that have no correlation to sample n can be dropped:

$$\frac{d}{d\sigma_{i}} R_{\alpha} = \frac{d}{d\sigma_{i}} \left[T_{i} \alpha_{i} \right]$$

Using the chain rule:

$$(14)\qquad \frac{d}{d\sigma_i}R_\alpha = \left(\frac{d}{d\sigma_i}T_i \cdot \alpha_i\right) + \left(T_i \cdot \frac{d}{d\sigma_i}\alpha_i\right)$$

Let's focus on $\frac{d}{d\sigma_i}T_n$, using equation (3)

$$\frac{d}{d\sigma_i}T_i = \frac{d}{d\sigma_i}exp\left(-\sum_{j=1}^{i}\sigma_j\delta_j\right)$$

We've seen this before… it's equivalent to:

$$\frac{d}{d\sigma_i}T_i = -\ \delta_i exp\left(-\sum_{j=1}^{i}\sigma_j\delta_j\right)$$

Or:

$$\frac{d}{d\sigma_i}T_i = -\ \delta_i T_i$$

Now $\frac{d}{d\sigma_i}\alpha_n$, substituting equation (2)

$$\frac{d}{d\sigma_i}\alpha_i = \frac{d}{d\sigma_i}\left[1 - exp\left(-\sigma_i\delta_i\right)\right]$$

This one is pretty simple, and reduces to:

$$\frac{d}{d\sigma_i}\alpha_i = \delta_i exp\left(-\sigma_i\delta_i\right)$$

Or just:

$$\frac{d}{d\sigma_i}\alpha_i = -\ \delta_i\left[1 - exp\left(-\sigma_i\delta_i\right) - 1\right]$$

$$\frac{d}{d\sigma_i}\alpha_i = -\ \delta_i\left(\alpha_i - 1\right)$$

Now we can put these terms back into equation (14):

$$\frac{d}{d\sigma_i}R_\alpha = \left(\left[-\ \delta_i T_i\right] \cdot \alpha_i\right) + \left(T_i \cdot \left[-\ \delta_i\left(\alpha_i - 1\right)\right]\right)$$

$$\frac{d}{d\sigma_i}R_\alpha = \left(-\ \delta_i \cdot T_i \cdot \alpha_i\right) - \left(\delta_i \cdot T_i \cdot \left(\alpha_i - 1\right)\right)$$

Simplifying, we obtain:

$$\frac{d}{d\sigma_i} R_\alpha = \delta_i \cdot T_i \left(1 - 2\alpha_i\right)$$

Bringing back the k subscript:

$$\frac{d}{d\sigma_{ki}} R_{\alpha k} = \delta_{ki} \cdot T_{ki} \left(1 - 2\alpha_{ki}\right)$$

And plugging into equation (13):

$$\frac{d}{d\sigma_{ki}} L_\alpha = \frac{2}{M} \cdot \left(R_{\alpha k} - 1\right) \cdot \left(\delta_{ki} \cdot T_{ki} \left(1 - 2\alpha_{ki}\right)\right)$$

We can simplify to:

$$\frac{d}{d\sigma_{ki}} L_\alpha = \frac{2}{M} \cdot \left(R_{\alpha k} - 1\right) \cdot \left(\delta_{ki} T_{ki}\right) \cdot \left(1 - 2\alpha_{ki}\right)$$

And remembering our other partial derivatives for the color channels:

$$\frac{d}{d\sigma_{ki}} L_r = \frac{2}{M} \cdot \left(R_{rk} - gt_{rk}\right) \cdot \left(\delta_{ki} c_{rki} T_{ki}\right) \cdot \left(1 - 2\alpha_{ki}\right)$$
$$\frac{d}{d\sigma_{ki}} L_g = \frac{2}{M} \cdot \left(R_{gk} - gt_{gk}\right) \cdot \left(\delta_{ki} c_{gki} T_{ki}\right) \cdot \left(1 - 2\alpha_{ki}\right)$$
$$\frac{d}{d\sigma_{ki}} L_b = \frac{2}{M} \cdot \left(R_{bk} - gt_{bk}\right) \cdot \left(\delta_{ki} c_{bki} T_{ki}\right) \cdot \left(1 - 2\alpha_{ki}\right)$$

We now have everything we need to plug back into the total loss function (10):

$$\frac{dL}{d\sigma_{ki}} = \frac{1}{2M} \cdot \left(1 - 2\alpha_{ki}\right) \cdot \left(\delta_{ki} T_{ki}\right) \cdot \left[\left(R_{rk} - gt_{rk}\right) c_{rki} + \left(R_{gk} - gt_{gk}\right) c_{gki} + \left(R_{bk} - gt_{bk}\right) c_{bki} + \left(R_{\alpha k} - 1\right)\right]$$