VI^{3}NR: Variance Informed Initialization for Implicit Neural Representations

Supplementary Material

6. Derivations

6.1. Derivation for the backward pass

Given

$$\frac{\partial \mathcal{L}}{\partial [z_{i-1}]_j} = \sum_{k=1}^{N_i} \frac{\partial \mathcal{L}}{\partial [z_i]_k} [W_i]_{kj} f'([z_{i-1}]_j), \tag{17}$$

we have that

$$\mu\left(\frac{\partial \mathcal{L}}{\partial z_{i-1}}\right) = \sum_{k=1}^{N_n} \mathbf{E}\left[[W_i]_{kj}\right] \mathbf{E}\left[\frac{\partial \mathcal{L}}{\partial [z_i]_k} f'([z_{i-1}]_j)\right] \tag{18}$$

$$= 0 \tag{19}$$

and

$$\sigma^2\left(\frac{\partial \mathcal{L}}{\partial z_{i-1}}\right) = \sum_{k=1}^{N_i} \mathbf{Var}\left[\frac{\partial \mathcal{L}}{\partial [z_i]_k}\right] \mathbf{Var}\left[[W_i]_{kj}\right] \left(\mathbf{E}\left[f'([z_{i-1}]_j)\right]^2 + \mathbf{Var}\left[f'([z_{i-1}]_j)\right]\right) \tag{20}$$

$$= N_i \sigma^2\left(\frac{\partial \mathcal{L}}{\partial z_i}\right) \sigma^2(W_i) (\mu^2(f'(z_i)) + \sigma^2(f'(z_i))). \tag{21}$$

Finally note that $N_i = M_{i+1}$.

6.2. Initialization for the first layer

For our INR experiments, we first normalize the input coordinates x_0 to be within $[-1, 1]^D$ (for some input dimension D). We then model our element input distribution as a uniform distribution over $[-1, 1]$, so $\mathcal{D}_{\textbf{in}} = \mathcal{U}([-1, 1])$. Then we have that

$$\mathbf{E}[x_0] = 0 \tag{22}$$

$$\mathbf{Var}(x_0) = \frac{1}{12} 2^2 = \frac{1}{3} \tag{23}$$

so $\mu^2(x_0) + \sigma^2(x_0) = \frac{1}{3}$. Thus by Eq. (4) we initialize our first layer weights with variance

$$\sigma^2(W_0) = \frac{\sigma_p^2}{M_0\left(\mu^2(x_0) + \sigma^2(x_0)\right)} \tag{24}$$

$$= \frac{3\sigma_p^2}{M_0} \tag{25}$$

which is equivalent to by $\mathcal{U}([-c, c])$ where

$$c = \sigma_p \sqrt{\frac{3}{M_0\left(\mu^2(x_0) + \sigma^2(x_0)\right)}} \tag{26}$$

$$= \sigma_p \sqrt{\frac{9}{M_0}}. \tag{27}$$

6.3. Analytical Expectations for Gaussians

Let us assume that the preactivations at layer $i-1$, z_{i-1}, have variance σ_p^2. Then to ensure that z_i has the same variance, we set the variance of W_i according to Eq. (4), which requires us to compute the mean and variance of x_i. We do the analytical derivation of this now.

Given $X \sim \mathcal{N}(0, \sigma_p^2)$ and $Y = \exp\left(\frac{-X^2}{2\sigma_a^2}\right)$ then

$$F_X(x) = \frac{1}{2} + \frac{1}{2}\operatorname{erf}\left(\frac{x}{\sigma_p\sqrt{2}}\right) \tag{28}$$

and $0 \leq Y \leq 1$ and

$$F_Y(y) = \mathrm{P}\left(\exp\left(\frac{-X^2}{2\sigma_a^2}\right) \leq y\right) \tag{29}$$

$$= \mathrm{P}\left(\frac{-X^2}{2\sigma_a^2} \leq \log(y)\right) \tag{30}$$

$$= \mathrm{P}\left(X^2 \geq -2\sigma_a^2 \log(y)\right) \tag{31}$$

$$= \mathrm{P}\left(|X| \geq \sigma_a\sqrt{-2\log(y)}\right) \tag{32}$$

$$= 2\mathrm{P}\left(X \leq -\sigma_a\sqrt{-2\log(y)}\right)$$

(as $\mathcal{N}(0, \sigma_p^2)$ is symmetric around 0)

$$= 2\left(\frac{1}{2} + \frac{1}{2}\operatorname{erf}\left(\frac{-\sigma_a\sqrt{-2\log(y)}}{\sigma_p\sqrt{2}}\right)\right) \tag{33}$$

$$= 1 + \operatorname{erf}\left(-\frac{\sigma_a}{\sigma_p}\sqrt{-\log(y)}\right) \tag{34}$$

$$= 1 + \operatorname{erf}\left(-\sigma_r\sqrt{-\log(y)}\right) \tag{35}$$

where $\sigma_r = \frac{\sigma_a}{\sigma_p}$.

Thus

$$
\begin{aligned}
f_Y(y) &= \frac{\mathrm{d}}{\mathrm{d}y} F_y(y) && (36)\\
&= \frac{\mathrm{d}}{\mathrm{d}y}\left(1+\operatorname{erf}\left(-\sigma_r\sqrt{-\log(y)}\right)\right) && (37)\\
&= \frac{2}{\sqrt{\pi}}\exp(\sigma_r^2\log(y))\frac{\mathrm{d}}{\mathrm{d}y}\left(-\sigma_r\sqrt{-\log(y)}\right) && (38)\\
&= \frac{2}{\sqrt{\pi}}y^{\sigma_r^2}\frac{\sigma_r}{2y\sqrt{-\log(y)}} && (39)\\
&= \frac{\sigma_r}{\sqrt{-\pi\log(y)}}y^{\sigma_r^2-1}. && (40)\\
& && (41)
\end{aligned}
$$

This has mean

$$
\begin{aligned}
\mathbf{E}[y] &= \int_0^1 y f_Y(y)\mathrm{d}y && (42)\\
&= \int_0^1 \frac{\sigma_r}{\sqrt{\pi}\sqrt{-\log(y)}}y^{\sigma_r^2}\,\mathrm{d}y && (43)\\
&= \frac{\sigma_r}{\sqrt{\pi}}\int_0^1(-\log(y))^{-0.5}\,y^{\sigma_r^2}\,\mathrm{d}y && (44)\\
&= \frac{\sigma_r}{\sqrt{\pi}}\frac{\Gamma(-0.5+1)}{(\sigma_r^2+1)^{-0.5+1}} && (45)\\
&= \frac{\sigma_r}{\sqrt{\pi}}\frac{\sqrt{\pi}}{\sqrt{\sigma_r^2+1}} && (46)\\
&= \frac{\sigma_r}{\sqrt{\sigma_r^2+1}} && (47)
\end{aligned}
$$

and variance

$$
\begin{aligned}
\mathbf{Var}(y) &= \int_0^1 y^2 f_Y(y)\mathrm{d}y - \mathbf{E}[y]^2 && (48)\\
&= \int_0^1 \frac{\sigma_r}{\sqrt{\pi}\sqrt{-\log(y)}}y^{\sigma_r^2+1}\,\mathrm{d}y - \mathbf{E}[y]^2 && (49)\\
&= \frac{\sigma_r}{\sqrt{\sigma_r^2+2}} - \frac{\sigma_r^2}{\sigma_r^2+1} && (50)
\end{aligned}
$$

so

$$
\mu^2(x_i)+\sigma^2(x_i) = \frac{\sigma_r}{\sqrt{\sigma_r^2+2}}. \quad (51)
$$

Thus by Eq. (4) we initialize our i^{th} layer weights with variance

$$
\begin{aligned}
\sigma^2(W_i) &= \frac{\sigma_p^2}{M_i\left(\mu^2(x_i)+\sigma^2(x_i)\right)} && (52)\\
&= \frac{\sigma_p^2\sqrt{\sigma_r^2+2}}{M_i\sigma_r} && (53)
\end{aligned}
$$

which is equivalent to by $\mathcal{U}([-c,c])$ where

$$
\begin{aligned}
c &= \sigma_p\sqrt{\frac{3}{M_i\left(\mu^2(x_i)+\sigma^2(x_i)\right)}} && (54)\\
&= \sigma_p\sqrt{\frac{3\sqrt{\sigma_r^2+2}}{M_i\sigma_r}}. && (55)
\end{aligned}
$$

7. SIREN's initialization

Note that SIREN [23] also uses sine activations and specifically uses $\sigma_p = 1$. In their derivation they use $\sin\left(\frac{\pi}{2}x\right)$ in order to only consider the monotonic region of sine. Our method also gives $\sigma^2(W_i) = \frac{2}{M_i}$ with that activation function. In the code, SIREN actually use $\sin(30x)$ which our method also gives $\sigma^2(W_i) = \frac{2}{M_i}$. However, for $\sin(x)$ our method gives $\sigma^2(W_i) = \frac{2.31}{M_i}$.

8. Comparison to Xavier and Kaiming init.

Our initialization for Gaussians is

$$
\begin{aligned}
\sigma^2(W_i) &= \frac{\sigma_p^2}{M_i\left(\mu^2(x_i)+\sigma^2(x_i)\right)} && (56)\\
&= \frac{\sigma_p^2\sqrt{\sigma_r^2+2}}{M_i\sigma_r} && (57)
\end{aligned}
$$

where $\sigma_r = \frac{\sigma_a}{\sigma_p}$, while Xavier initialization is of the form (for middle layers)

$$
\sigma^2(W_i) = \frac{1}{M_i} \quad (58)
$$

and Kaiming initialization is of the form

$$
\sigma^2(W_i) = \frac{2}{M_i}. \quad (59)
$$

Thus for a fixed σ_a (the Gaussian activation function parameter), Xavier and Kaiming are equivalent to our Gaussian initialization for some σ_p. For example, $\sigma_a = 0.05$, $\sigma_p = 0.33$ with our Gaussian init gives $\sigma^2(W_i) = \frac{1.02}{M_i}$ matching Xavier initialization, and $\sigma_a = 0.05$, $\sigma_p = 0.41$ gives $\sigma^2(W_i) = \frac{1.96}{M_i}$ matching Kaiming.

For fixed σ_a (so a fixed Gaussian activation function), it is unlikely that the σ_p that Xavier or Kaiming correspond to are optimal (Tab. 8 Middle). However, if we grid search on σ_a (*i.e.* on the activation function) then it is possible that there will be a σ_a such that Xavier and Kaiming will correspond to the optimal σ_p for that σ_a (Tab. 8 Bottom). In fact, the observed trend in Fig. 2b makes it quite likely.

9. Image Comparison.

We compare image reconstruction with Gaussian activation with the three different types of initializations in Fig. 4.

Figure 4. **Image comparison.** Left to right: random normal init, our MC init, our init.

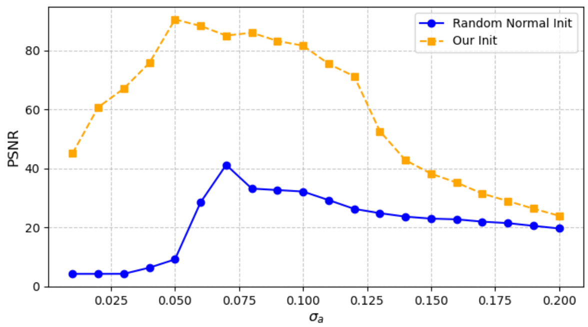

Figure 5. **Performance gap vs. σ_a.** As σ_a decreases, performance drops for both inits, but a significant gap remains.

10. Audio Reconstruction Implementation details

The audio reconstruction results presented in the main paper differed from the image and SDF reconstruction setups in several key aspects. Specifically, the network architecture used three hidden layers, each containing 256 elements, and the bias terms were initialized identically to the weights. These modifications were consistently applied across all initialization methods examined and proved essential for achieving convergence.

11. Improvement gap dependence on activation function parameters

We give results for small σ_a in Fig. 5 Left. The results show that the proposed initialization outperforms random init for smaller σ_a values while a degradation in performance is observed for both.