Appendix

A Encoding signals

Embedding	Image type	PSNR
ReLU	Natural	20.42
Tanh	Natural	16.91
SoftPlus	Natural	16.03
SiLU	Natural	17.59
Gaussian	Natural	33.43
Laplacian	Natural	33.01
Quadratic	Natural	32.90
Multi-Quadratic	Natural	33.11
ExpSin	Natural	32.99
Super-Gaussian	Natural	33.12
ReLU	Text	18.49
Tanh	Text	16.19
SoftPlus	Text	15.77
SiLU	Text	17.43
Gaussian	Text	36.17
Laplacian	Text	36.29
Quadratic	Text	35.55
Multi-Quadratic	Text	36.20
ExpSin	Text	35.89
Super-Gaussian	Text	35.57
ReLU	Noise	10.82
Tanh	Noise	9.66
SoftPlus	Noise	9.71
SiLU	Noise	11.21
Gaussian	Noise	11.78
Laplacian	Noise	11.01
Quadratic	Noise	11.67
Multi-Quadratic	Noise	11.29
ExpSin	Noise	11.45
Super-Gaussian	Noise	11.33

Table 2. Quantitative comparison between activations in 2D signal encoding on [27] after 3000 epochs. The proposed activations yield high PSNRs. The noise signals are difficult to be encoded with high fidelity due to limited redundancy.

B Norms of the layer outputs.

Our empirical results strongly suggested that the coordinate networks control the local Lipschitz constant of the encoded signal primarily via the angle between

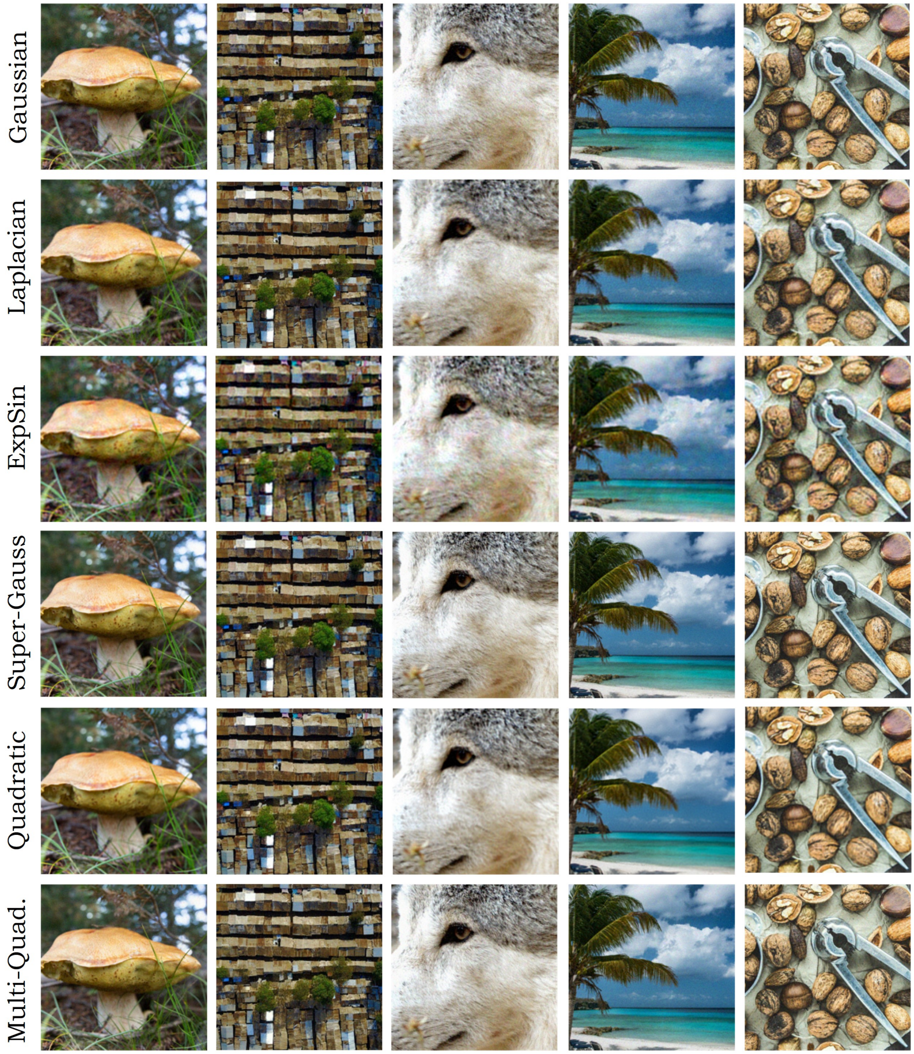

Fig. 10. Qualitative examples of 2D signal encoding using the proposed activations on the natural images by [27].

Fig. 11. A qualitative comparison between ReLU and Gaussian activations (w/o positional embedding) in 3D view synthesis.

Fig. 12. A qualitative comparison between ReLU and Gaussian activations (w/ positional embedding) in 3D view synthesis.

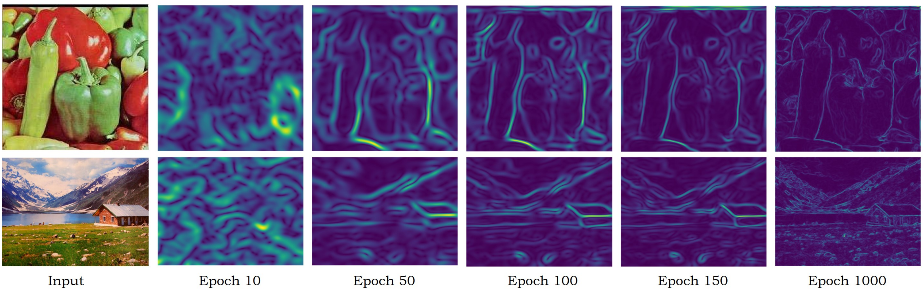

Fig. 13. $\|\mathbf{J}_f(\mathbf{x})\|_F$ convergence as the training progresses.

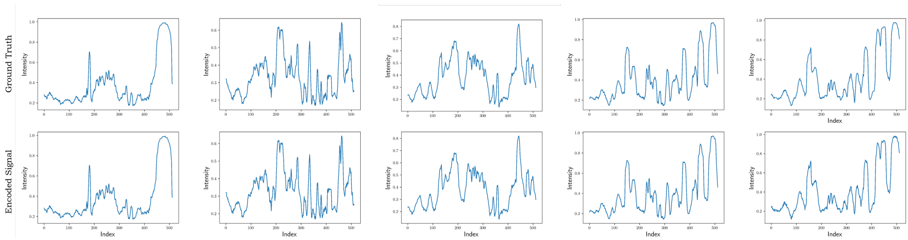

Fig. 14. 1D signal encoding using Gaussian activated MLPs.

the layer outputs. In this section, we conduct another experiment to validate this behavior further. We measure the deviations of the layer output norms within patches of images (Fig. 15). As shown, within a local area, the norms of the layer outputs do not significantly deviate from their maximum, which backs up our previous experiments.

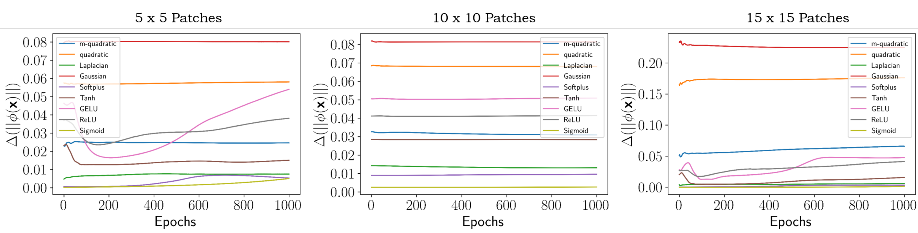

Fig. 15. The norms of the layer outputs remain approximately locally constant. For each patch obtained via an overlapping sliding window over an image, we measure $\frac{\|\max(\phi(\mathbf{x}))-\min(\phi(\mathbf{x}))\|}{\|\max(\phi(\mathbf{x}))\|}$. The above measure is then averaged over all the patches, layers, and a subset of 10 images to obtain $\Delta(\|\phi(\mathbf{x})\|)$ while training. As depicted, the norms of the vectors do not significantly deviate from their maximum within a subset.

In Sec 3.4, the desired properties of the activations are derived under the restriction that the norms of the outputs of the layers remain approximately constant over a local subset. However, even in the case where the vector norms are not approximately constant, our conclusions still hold: First, let us formally define the local Lipschitz smoothness.

Definition 1. *A function $f : \mathbb{R}^m \to \mathbb{R}^n$ is locally Lipschitz around $\boldsymbol{x}_0 \in \mathbb{R}^m$ if for all $\boldsymbol{x} \in \mathbb{B}^m_\delta$ there exists a constant C such that $\|f(\boldsymbol{x}) - f(\boldsymbol{x}_0)\| \leq C\|\boldsymbol{x} - \boldsymbol{x}_0\|$ where $\boldsymbol{x}_0$ is the center $\boldsymbol{x}_0$ of $\mathbb{B}^m_\delta$. Then, the smallest C for which the above inequality is satisfied is called the Lipschitz constant of f around $\boldsymbol{x}_0$, and is denoted as $C_{\boldsymbol{x}_0,\delta}(f)$.*

A hidden-layer is a composition of an affine function g and a non-linearity ψ. Hence, the composite local Lipschitz constant of a hidden-layer is upperbounded with $C_{\mathbf{x}_0,\delta}(\psi \circ g) \leq C_{\mathbf{x}_0,\delta}(g)C_{\mathbf{x}_0,\delta}(\psi)$. We study these properties as δ approaches zero. As shown in Sec.3.4, the affine transformation g cannot vary the local Lipschitz smoothness across the signal. Thus, we focus on the activation function ψ. Since ψ is a continuously differentiable function, applying the Taylor expansion gives

$$\psi(\mathbf{x}) = \psi(\mathbf{x}_0) + \mathbf{J}(\psi)_{\mathbf{x}_0}(\mathbf{x} - \mathbf{x}_0) + \Theta(\|\mathbf{x} - \mathbf{x}_0\|), \tag{10}$$

where $\Theta(\|\mathbf{x} - \mathbf{x}_0\|)$ is a rapidly decaying function as $\mathbf{x} \to \mathbf{x}_0$. Rearranging Eq. 10 we get

$$\|\psi(\mathbf{x}) - \psi(\mathbf{x}_0)\| \leq \|\mathbf{J}(\psi)_{\mathbf{x}_0}\|_o \|(\mathbf{x} - \mathbf{x}_0)\| + \|\Theta(\|\mathbf{x} - \mathbf{x}_0\|)\|, \tag{11}$$

$$\lim_{\delta \to 0}\left[\sup_{\mathbf{x} \in \mathbb{B}_\delta^m} \frac{\|\psi(\mathbf{x}) - \psi(\mathbf{x}_0)\|}{\|\mathbf{x} - \mathbf{x}_0\|}\right] \leq \lim_{\delta \to 0}\left[\|\mathbf{J}(\psi)_{\mathbf{x}_0}\|_o + \frac{\|\Theta(|\mathbf{x} - \mathbf{x}_0|)\|}{\|\mathbf{x} - \mathbf{x}_0\|}\right]. \tag{12}$$

The left-hand side of Eq. 12 is the point-wise Lipschitz constant of $\psi(\cdot)$ at $\mathbf{x}_0$ by definition. Again, the quantity $\lim_{\delta \to 0} \frac{\|\Theta(|\mathbf{x} - \mathbf{x}_0|)\|}{\|\mathbf{x} - \mathbf{x}_0\|}$ is zero by deifinition. Thus, denoting the point-wise Lipschitz constant of $\psi(\cdot)$ at $\mathbf{x}_0$ as $C_{\mathbf{x}_0}(\psi)$, we get

$$C_{\mathbf{x}_0}(\psi) \leq \|\mathbf{J}(\psi)_{\mathbf{x}_0}\|_o. \tag{13}$$

Consider,

$$\|\mathbf{J}(\psi)_{\mathbf{x}_0}\|_o = \sup_{\mathbb{B}_\delta^m \ni \|\mathbf{x}\| = 1} \|\mathbf{J}(\psi)\mathbf{x}\|$$

Here the operator norm is defined with the vector-norm on the vector space of $\mathbf{x}$. Further, the Frobenius norm and the vector-norm are the same for vectors. Therefore,

$$\|\mathbf{J}(\psi)_{\mathbf{x}_0}\|_o = \sup_{\mathbb{B}_\delta^m \ni \|\mathbf{x}\| = 1} \|\mathbf{J}(\psi)\mathbf{x}\|_F$$

Frobenius norm is submultiplicative. Thus,

$$\begin{aligned}
\|\mathbf{J}(\psi)_{\mathbf{x}_0}\|_o &\leq \sup_{\mathbb{B}_\delta^m \ni \|\mathbf{x}\| = 1} \|\mathbf{J}(\psi)\|_F \|\mathbf{x}\|_F \\
&= \sup_{\mathbb{B}_\delta^m \ni \|\mathbf{x}\| = 1} \|\mathbf{J}(\psi)\|_F \|\mathbf{x}\| \\
&= \|\mathbf{J}(\psi)\|_F \sup_{\mathbb{B}_\delta^m \ni \|\mathbf{x}\| = 1} \|\mathbf{x}\| \\
&= \|\mathbf{J}(\psi)\|_F.
\end{aligned}$$

Hence, with Eq. 13, we have

$$C_{\mathbf{x}_0}(\psi) \leq \|\mathbf{J}(\psi)_{\mathbf{x}_0}\|_F. \tag{14}$$

.

Since $\mathbf{J}(\psi)$ is diagonal,

$$C_{\mathbf{x}}(\psi) \leq \sqrt{D(\max_{x \in \mathbb{R}} \psi'(x))^2}, \tag{15}$$

Where D is the width of the network layer. One can see that in order to have a larger Lipschitz constant over some interval, the maximum first-order derivative of the activation should be higher over the same interval. Furthermore, in order to obtain varying local Lipschits smoothness, the maximum first-order derivatives of the activation within finite intervals should vary across the domain.