Supplementary Material for
Neuromanifold-Regularized KANs for Shape-fair Feature Representations

A. Details of the cue conflict dataset

The cue-conflict dataset consists of 1536 images, is class-balanced, and is composed of 8 different categories: spider, orthopetra, garment, ladybug, barrel/chest, monkey, elephant, and citrus. The primary concerns in choosing these categories are *(i)* strong shape **and** texture information, *(ii)* agreement with the categories of both the Tiny-ImageNet and ImageNet. We used a separate set of content and style images: while content images have clear shape properties and are realistic, the style images are chosen to be rich in texture (Figure. 1). For some style images, we also used crop and pastes to remove as much of shape information as possible (Figure. 1-(c)&(d)). For each category, there are 8 content images, and 3 style images.

Our implementation is based on [1], with the main difference that we finetuned the VGG networks on TinyImageNet so that features in 64 × 64 are captured better.

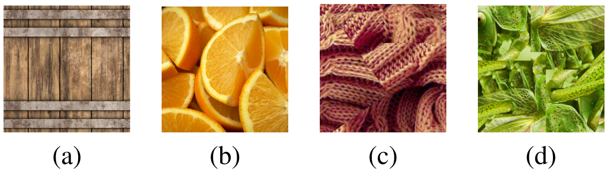
(a) (b) (c) (d)

Figure 1. Sample texture (style) sources, left to right: barrel, citrus, garment, orthopetra.

B. Potential of KANs for shape fairness

Main difference of KANs from universal approximation theorem-based neural networks is that they can learn nonlinearities directly. In order to do so, a finite parametrization of the function space is required. Usually, this is achieved by first choosing an infinite dimensional basis, followed up by choosing a feasible finite subset in which the optimization is to be done. Here, we consider polynomial basis for the ease of analysis.

The concepts of shape and texture do not have a universally accepted mathematical definition. Hence in our analyses, we focus on mathematically well-defined concepts that are shown to be related to shape and texture.

Spectral analysis of learnable and fixed nonlinearities. Although the relationship is not strict, shape information is associated with low-frequency structures that capture global form, while texture primarily resides in high-frequency details [3]. With this relation in mind, **we show that fixed nonlinearities, especially *parametric* $ReLU$, cause a more serious increase in the frequency bandwidth of the inputs, indicating repeated introduction of texture-like features.**

Polynomial nonlinearities. Decomposing a signal $x(t)$ into (WLOG) zero-phase waves as $x(t) = \sum_i a_i \cos(2\pi f_i t)$, we can see the effect of polynomial nonlinearities on the spectral components directly.

For example, for second order polynomials, $y(t) = p_2 x^2(t) + p_1 x(t) + p_0$, in view of

$$x^2(t) = \left(\sum_i a_i \cos\left(2\pi\, f_i\, t\right) \right)^2 = A(1 + g(t))$$

where $A = \sum_i \frac{a_i^2}{2}$, and

$$\begin{aligned} g(t) = {} & \frac{1}{2A} \sum_i a_i^2 \cos\left(2\pi\, 2 f_i\, t\right) \\ & + \frac{1}{A} \sum_{i>j} a_i a_j \cos\left(2\pi\left(f_i + f_j\right) t\right) \\ & + \frac{1}{A} \sum_{i>j} a_i a_j \cos\left(2\pi\left(f_i - f_j\right) t\right). \end{aligned}$$

the output signal has frequency components arising from addition/subtraction of different components. This observation can be straightforwardly generalized to n-degree polynomials, which have n-fold frequency components, **providing support for our approach of controlling the degree of the neuromanifold to emphasize shape-fairness.** That is, the frequency bandwidth increases, but as we show in the following, the severity of this effect is limited when compared with that of $PReLU$.

Parametric $ReLU$. Our approach follows and generalizes that of [2]. The Taylor series for $PReLU(x(t)) = \begin{cases} x, & \text{if } x \geq 0 \\ \alpha x, & \text{if } x < 0 \end{cases}$ can be acquired by first identifying it as

$$y(t) = PReLU(x(t)) = \frac{x(t) + \beta\sqrt{x^2(t)}}{1+\beta}, \quad \beta = \frac{1-\alpha}{1+\alpha},$$

($\beta = 1$ corresponds to $ReLU$) and identifying the input $x^2(t)$ in terms of its frequency components as

$$x^2(t) = A(1 + g(t))$$

as above. Different from polynomial activations, in addition to the 2-fold frequency components, theoretically, *infinite*-fold components are introduced:

$$y(t) = \frac{x(t) + \beta\sqrt{A(1 + g(t)}}{1+\beta}$$

can be expanded around $g(t) = 0$ as

$$y(t) = \frac{x(t)}{1+\beta} + \frac{\beta\sqrt{A}}{1+\beta}\left(\sum_{n=0}^{\infty} \frac{(-1)^n\,(2n)!}{(1-2n)(n!)^2(4^n)} g^n(t)\right).$$

Fig. 2 displays the first 1000 terms of the coefficients

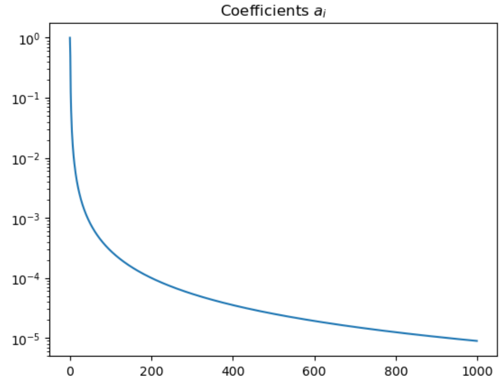

Figure 2. First 1000 Taylor coefficients coming from the nonlinear term in $PReLU$.

We report the effect of the $ReLU$ and a degree 2 polynomial in Fig. 3 and the repeated application of convolutional signal processing for both with random weights in Fig. 4.

C. Optimization of style decorrelation loss

Style decorrelation loss aims to encourage stylistic diversity between the two branches of the NMR-KAN. In Fig. 5 we show a typical progress of $\mathcal{L}_{decorr}$. The loss remains mostly unchanged in the early stages of learning, and near the half of training, goes through a rapid optimization step.

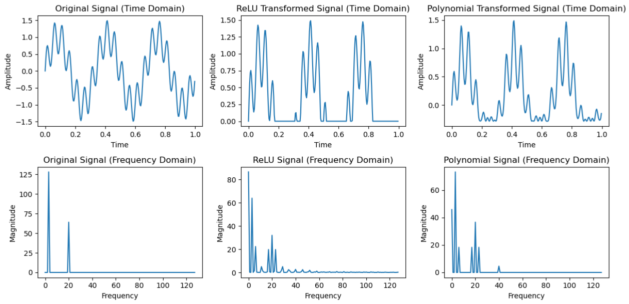

Figure 3. Nonlinearities increase the frequency bandwidth.

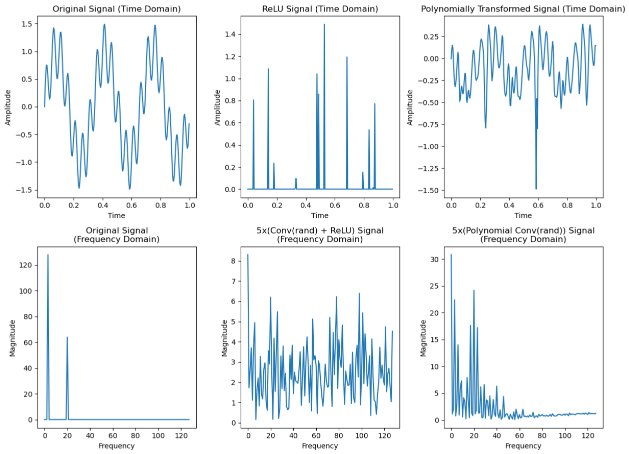

Figure 4. Nonlinearity types strongly determines the spectral components after repeated applications. After 5 layers of linear convolutional filtering, each followed up by $ReLU$ activation, the signal's frequency bandwidth is considerably more increased compared to the 5 layers of convolutional filtering with degree-2 polynomials. For both, filter size is set to 3 and the weights are randomly assigned.

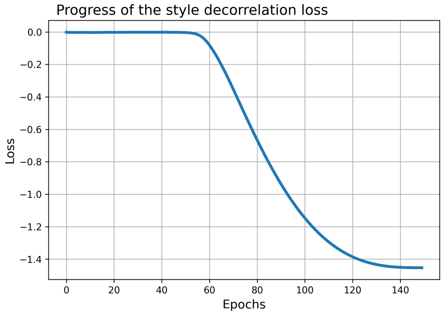

Figure 5. Typical progress of the style decorrelation loss during model training. The early stationary state is important for training stability.

References

[1] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Pro-*

ceedings of the IEEE conference on computer vision and pattern recognition, pages 2414–2423, 2016. 1

[2] Christodoulos Kechris, Jonathan Dan, Jose Miranda, and David Atienza. Dc is all you need: describing relu from a signal processing standpoint. *arXiv preprint arXiv:2407.16556*, 2024. 2

[3] Shunxin Wang, Raymond Veldhuis, Christoph Brune, and Nicola Strisciuglio. What do neural networks learn in image classification? a frequency shortcut perspective. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1433–1442, 2023. 1