IMAGE DENOISING WITH KOLMOGOROV ARNOLD NETWORKS

Divesh Basina*
School of Computing and Augmented Intelligence
Arizona State University
Tempe, AZ
dbasina@asu.edu

ABSTRACT

Kolmogorov Arnold Networks have been presented as an alternative to the traditional Multi-Layer Perceptron (MLP) architecture, showcasing strong parameter efficiency to the order of magnitude 100 and higher. Nevertheless, current implementations of KANs pose a significant challenge in their efficacy in image processing deep learning models. Many works have shown that while KANs are parameter efficient and powerful, they are not worth switching from MLP architecture given the added computation costs. Our work showcases techniques that make it feasible to incorporate KANs meaningfully in image denoising architectures and have competitive results to current state-of-the-art image denoising models presented in the 2025 CVPR Image Denoising Competition organized by NTIRE, both in terms of output quality measured by PSNR and computation costs measured by time. The implementation is available at https://github.com/dbasina/KUnet$_Dn50_Competition$.

Keywords Kolmogorov-Arnold Networks · Image denoising · Depthwise-convolutions · Tokenization · UNET

1 Introduction

Kolmogorov Arnold Networks (KANs) are a recent innovation in AI/ML [1]. The paper presents an alternative to the multi-layer perceptron (MLP) used in large-scale networks and highlights its benefits. Many benefits focus on reducing model size while maintaining or improving efficiency, directly addressing the curse of dimensionality. KANs have been applied in fields such as mathematical problem solving, physics-informed networks, time series analysis and quantum physics to name a few, achieving significant gains in parameter efficiency [2, 3, 4]. Computer vision (CV) is one such field that uses large models with billions of parameters. Within CV, image denoising is a classical problem that also often validates the effectiveness of many large-scale computer vision models, leading them to find applications in image deblurring, super resolution, light normalization, diffusion denoising etc [5]. It serves as a litmus test for KANs' efficacy in computer vision tasks. Works such as [5] show that while KANs are efficient at image classification and segmentation, scaling them up reasonably to larger image models could be challenging by performing these tasks on benchmark datasets namely MNIST and CIFAR comprising of low resolution images of 64x64 and 32x32 pixels. Our work, largely inspired by UKAN [6] not only backs these claims, but also shows a collection of techniques that allow the feasibility of KANs in denoising tasks on HD images that are of 2K resolution or higher, achieving competitive results to current state of the art (SOTA) image denoising models presented at the denoise50 2025 competition organized by NTIRE for CVPR.

* *Citation*: **Authors. Title. Pages.... DOI:000000/11111.**

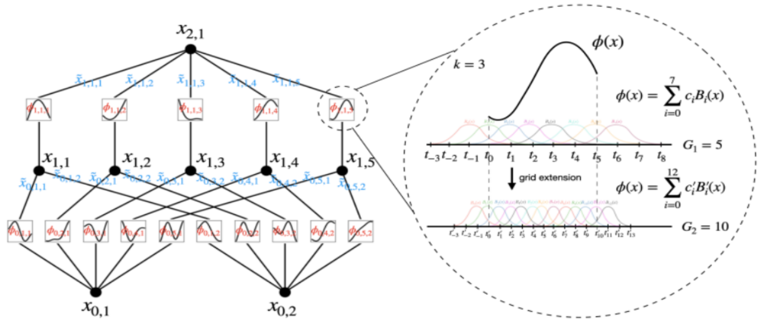

Figure 1: Left: Notations of activations that flow through the network. Right: an activation function is parameterized as a B-spline, which allows switching between coarse-grained and fine-grained grids.

2 Related Work

The model used to implement the denoiser while not novel, we have incorporated many techniques inspired from various different sources in order to make it feasible to work with KANs. The basic idea stems from UKAN as presented in [5]. While the authors from [5] present a simple KANs model tested on CIFAR and MNIST datasets, that have at most 64x64 size images, our work heavily modified the internal workings to make it feasible to train the model on large 2000x2000 pixel images. We draw inspiration from [7] and incorporate a transformer type architecture in the latent space of the UNET. The current efficient KAN implementation presented by [8] requires significant optimizatons to be able to have meaninng full training times. In order to make the model efficient, we incorporated depth wise convolutions, a technique developed to achieve the same results as regular convolutions with roughly 1/3rd number of computations. Our work is most similar to [6] who have done something similar using medical imaging data. The work showcases that KANs are infact feasible in the computer vison space. Although, given that they have only worked in the generative space and used this type of a model. Our work is more generalized and targets all kinds of images for denoising. Infact, our model was submitted to the NTIRE image denoising challenge organized by CVPR to measure the competitive quality of denoising.

3 Methodology

Our work combines the strengths of various techniques used in deep learning architectures for images. The section presents an overview of the techniques used and their impact on the performance of the model. Our implementation exploits the strengths of skip connection, depthwise convolution, tokenization and attention along with KANs proejctions of the tokens in order to downsample and process the images to denoise.

3.1 UNET and Skip Connections

U-Net is the foundation of many deep neural networks for downstream image processing tasks. Presented by Ronneberger et al. [9], U-Nets follow an encoder-decoder architecture that downsamples an image to extract its most prominent features into a latent representation. A decoder is then trained to reconstruct the image from this latent space based on the needs of the downstream task. Within the U-Net, skip connections are used to pass feature maps from the encoder directly to the decoder at corresponding levels. Without these skip connections, the decoder relies only on the latent representation, which often lacks fine-grained spatial details due to repeated downsampling. This can lead to poor reconstructions, especially in tasks requiring high precision. With skip connections, however, the decoder receives additional high-resolution features from earlier layers, enabling it to better guide the reconstruction process. The input from the skip connections acts as a structural template that complements the upsampled features, allowing for more accurate and detailed outputs.

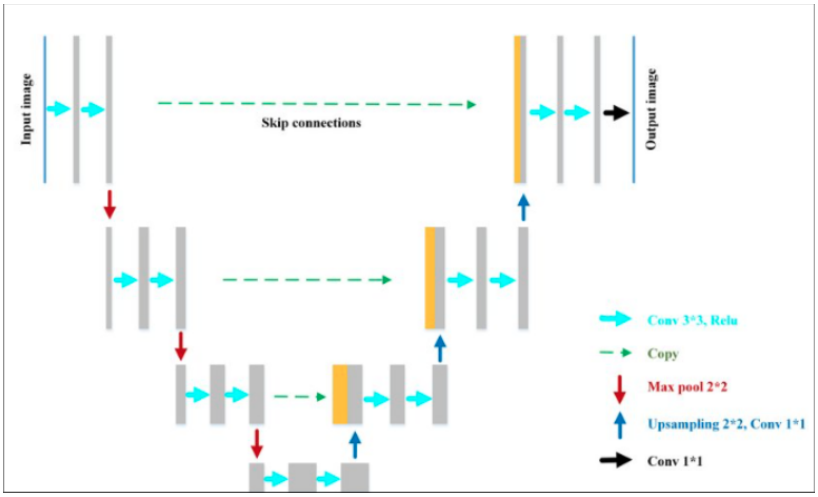

Figure 2: Skip Connections in UNets

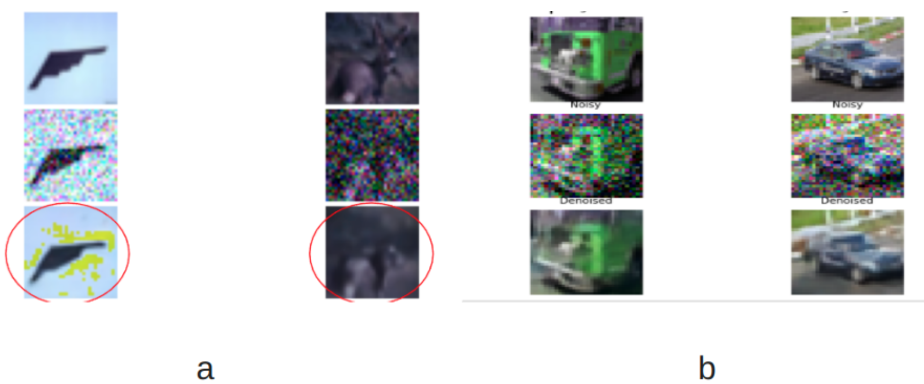

Figure 3: Performance of UNETs for denoising. Top row shows the orginial image, middle row is the noisy version of the image and the bottom row is the denoised output of the model (a) On a standard UNET we see many artifcats that appear in the denoised result. Notably, many features are also lost while upsampling as shown by the marked red circles. We can see yellow pixels appear the first denoised image where as in the second image, there are no ears in the denoised output. (b) On a UNET with skip connections and depthwise convolutions, even on a seemingly complex nature of pixels, we do not observe any artifacts and the features of the original image are well preserved

3.2 Depthwise Convolutions

The currently available efficient implementations of Kolmogorov-Arnold Network (KAN) layers still require further optimization to support large-scale models that are practical for training . To address inefficiencies elsewhere in the network, our model incorporates depthwise convolutions. While traditional convolutional layers perform 3-dimensional convolutions across both spatial and channel dimensions simultaneously, depthwise convolutions apply separate filters to each input channel independently, followed by a pointwise (1x1) convolution to combine the outputs. This approach achieves a similar representational capacity as full 3D convolutions but with significantly reduced computational cost—roughly one-third fewer multiplications. Depthwise separable convolutions, as introduced in works such as Xception [10] and MobileNets [11], offer an efficient alternative for deep architectures, especially when paired with computationally demanding components like KAN layers.

a b

Figure 4: Denoised results of the model on noisy HD 2K resolution test set images from the NTIRE denoise50 competition . The noisy image (a) consists of 50 standard deviation gaussian noise whereas (b) is the denoised output from the model

3.3 Tokenization and Attention

Tokenization is a technique mainly used in transformer architectures for processing natural language. It is used to break up a sentence into components by converting each word into a token which is a projection onto a vector. In order to process sentences, an attention mechanism is implemented which models the intermediate relationship between tokens to extract semantics. In [7], Dosovitskiy et al. showed that images can be converted to tokens, and the power of transformers can be exploited in image processing by breaking up an image into smaller patches and then processing individual pathces. In the image space, each patch of an image is treated as an independant token, and the attention layer handles the spatial relationship between patches. Given that the downstream images are HD and of 2K resolution, in order to have competitive results, we found it necessary to incorporate image tokenization. Through this, the model gains the ability to denoise at a finer grain that is determined by the patch size. Additionally, image reconstruction from the down-sampled image preserves many of the finer details of the image as we now process patches rather than the entire image at once.

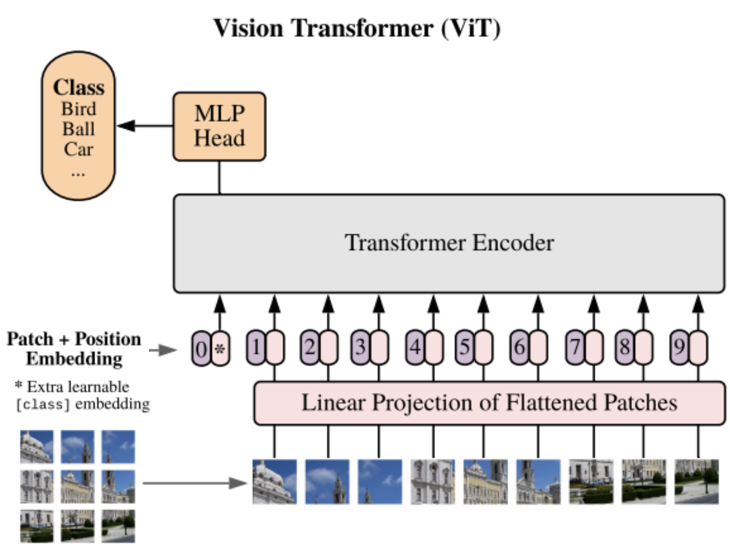

Figure 5: Vision Transformer Architecture

3.4 Kolmogorov-Arnold Linear Layers

KANs [1] were proposed as an alternative to multi-layer perceptrons (MLPs). Unlike MLPs, KANs place learnable activation functions on the edges, while the nodes function purely as summation points. These learnable activations are parameterized using third-order B-spline functions. Due to the flexibility of these learned activations, a single KAN layer offers greater expressive power compared to its MLP counterpart. KANs have also been shown to be significantly more parameter-efficient in works by Liu et al. and Ivashkov et al. [4]. However, studies such as [5] highlight that KANs can be computationally expensive, particularly for image data in datasets like MNIST and CIFAR. To balance efficiency and performance, our model integrates KAN projection layers for processing tokenized image patches, leveraging their expressiveness while limiting computational overhead.

Kolomogorov-Arnold Network Representation Theorem:

$$f(\mathbf{x}) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p}(x_p) \right) \tag{1}$$

Implementation details. Although KAN layer looks extremely simple, it is non-trivial to make it well optimizable. The key tricks are:

(1) Residual activation functions: We include a basis function $b(x)$ (similar to residual connections) such that the activation function $\phi(\mathbf{x})$ is the sum of the basis function $b(x)$ and the spline function:

$$\phi(x) = w_b b(x) + w_s \text{spline}(x). \tag{2}$$

We set

$$b(x) = \text{silu}(x) = \frac{x}{1 + e^{-x}} \tag{3}$$

in most cases $spline(x)$ is parameterized as a linear combination of B-splines such that

$$\text{spline}(x) = \sum_i c_i B_i(x) \tag{4}$$

4 Results

The development of the model was an iterative process where each technique that was picked up was incorporated with the goal of building a competitive denoiser for the NTIRE CVPR denoise50 competition by utilizing KAN linear layers to project the latent representation of a downsampled image. Earlier stages of the development for assesing the effectiveness of techniques was done on smaller datasets namely CIFAR-10. The competition was evaluated on the DIV2K [12] and LSDIR [13] datasets whose training and validation sets are publicly available. The primary error metric used in our experimentation is peak signal to noise ratio (PNSR).

4.1 Skip Connections, DW convolutions and tokenization

During our experimentation, we incrementally added skip connections and replaced the downsampling layers with depthwise convolution layers. Initially, the network took about 20 epochs to converge to a PSNR loss of -19.23 without skip connections. However, after introducing skip connections, the network converged to a PSNR loss of -19.47 within just 3 epochs, demonstrating a significant improvement in learning efficiency.

On the other hand, due to the computationally expensive nature of KANs, training on a single image from the CIFAR-10 dataset initially took approximately 3 hours per epoch. After replacing the downsampling layers with depthwise convolutions, the per-epoch training time dropped to around 20 minutes, significantly improving training efficiency. Tokenization proved substantially beneficial when applied to higher-resolution images (2K), particularly those from the DIV2K and LSDIR datasets. Without tokenization, training plateaued at a PSNR loss of 23.24 after 70 epochs. In contrast, incorporating tokenization with KAN projection layers improved the PSNR loss significantly, reaching a validation set threshold of 28.91.

4.2 Competition Results

The developed model was submitted to the 2025 CVPR NTIRE Image Denoising Challenge. During the development phase, our model ranked 9th, as shown in Table 1. This result demonstrates that KANs can be competitive, especially considering the limited development time and the high level of competition from established experts in the field.

Table 1: The development phase results of NTIRE denoise50 competition for CVPR 2025. Our results, in bold, show the effectiveness of our denoising model compared to other current state of the art models sorted in order by PSNR values on the test set provided by the organizers

#	User	PSNR	SSIM	Runtime per Image[s]
1	deepak.tyagi	29.79(1)	0.86(1)	N/R(1)
2	alwaysu	29.79(2)	0.86(2)	N/R(1)
3	Tcler_dn	29.61(3)	0.86(3)	1.81(5)
4	wangty	29.56(4)	0.86(4)	N/R(1)
5	Wedream	29.55(5)	0.86(5)	N/R(1)
6	S/W	29.48(6)	0.86(6)	N/R(1)
7	JanSeny	29.17(7)	0.85(7)	1.00(4)
8	Dennis5251	29.12(8)	0.85(8)	10.43(6)
9	**dbasina**	**28.91(9)**	**0.84(9)**	**0.05(2)**

5 Conclusion

Until this point, we have developed and improved a technique that incorporates important architectural improvements in order to provide the foundation for a high-resolution picture denoising model employing KANs. We started with a UNET backbone and used skip connections to provide more accurate reconstructions by preserving fine-grained spatial features between the encoder and decoder routes. We used depthwise convolutions, which efficiently reduced the number of parameters and accelerated training, to lower computational complexity without sacrificing performance. To improve the model's capacity to incorporate both local and global context, we also employed tokenization and attention techniques. The techniques showcased in our work show one approach to incorporate KANs into image processing models. Given the results that our work produced in a short amount of time and limited resources, our model was able to produce results within 0.88 PSNR of current SOTA image denoising models. Our work shows that KANs have a place in image processing and this needs to be further investigated with more efficient implementations.

6 Future Work

Through experimenting with the elimination of activation functions like ReLU, we hope to further improve the denoising model and maybe enhance learning dynamics. Token projection layer optimization will be investigated further in order to strike a compromise between computational economy and performance. Building on our current efforts with LSDIR denoising and motivated by recent work like TimeLens++, we are also exploring event-based picture deblurring. With a solid base already established, these strategies seek to improve our models' quality and real-time applicability, especially for real-world deployment and competitive benchmarking.

7 Workload Distribution

The workload is split up as following:
Divesh Basina: Responsible for developing and implementing the training pipeline. This involves designing the training framework, coding the necessary modules, and ensuring seamless integration of various components.

Sathwik Pillalamarri: Handles the organization of experimental results and manages the inference process. This role includes gathering performance data, maintaining logs, and ensuring that results are clearly documented for analysis.

Sudhanshu Kadam and Ajesh Nadar: Tasked with reviewing existing computer vision (CV) models and research papers. They will tune and implement promising techniques from the literature into our current work, aiming to enhance model performance and contribute innovative ideas.

References

[1] Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljačić, Thomas Y. Hou, and Max Tegmark. Kan: Kolmogorov-arnold networks, 2025.

[2] Remi Genet and Hugo Inzirillo. Tkan: Temporal kolmogorov-arnold networks, 2024.

[3] Zavareh Bozorgasl and Hao Chen. Wav-kan: Wavelet kolmogorov-arnold networks, 2024.

[4] Petr Ivashkov, Po-Wei Huang, Kelvin Koor, Lirandë Pira, and Patrick Rebentrost. Qkan: Quantum kolmogorov-arnold networks, 2024.

[5] Basim Azam and Naveed Akhtar. Suitability of kans for computer vision: A preliminary investigation, 2024.

[6] Chenxin Li, Xinyu Liu, Wuyang Li, Cheng Wang, Hengyu Liu, Yifan Liu, Zhen Chen, and Yixuan Yuan. U-kan makes strong backbone for medical image segmentation and generation, 2024.

[7] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale, 2021.

[8] Alexander Dylan Bodner, Antonio Santiago Tepsich, Jack Natan Spolski, and Santiago Pourteau. Convolutional kolmogorov-arnold networks, 2025.

[9] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

[10] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.

[11] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[12] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.

[13] Yawei Li, Kai Zhang, Jingyun Liang, Jiezhang Cao, Ce Liu, Rui Gong, Yulun Zhang, Hao Tang, Yun Liu, Denis Demandolx, Rakesh Ranjan, Radu Timofte, and Luc Van Gool. Lsdir: A large scale dataset for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 1775–1787, June 2023.