

# KSSANet: KAN-Driven Spatial-Spectral Attention Networks for Hyperspectral Image Super-Resolution

Baisong Li<sup>1,2</sup>, Xingwang Wang<sup>1,2\*</sup> and Haixiao Xu<sup>1,2</sup>

<sup>1</sup>College of Computer Science and Technology, Jilin University

<sup>2</sup>Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University

lbs23@mails.jlu.edu.cn, {xww,haixiao}@jlu.edu.cn

**Abstract**—Due to the limitations of physical imaging, acquiring high-resolution hyperspectral images (HR-HSIs) has always been a significant challenge. Single hyperspectral image super-resolution (SHSR) technology aims to generate corresponding HR-HSIs by processing low-resolution hyperspectral images (LR-HSIs). Compared to multi-source data fusion methods, SHSR relies solely on a single low-resolution image and does not require additional auxiliary information or multimodal data, making it more flexible and efficient in data acquisition. Recently, Kolmogorov–Arnold Networks (KAN), which derive from the Kolmogorov–Arnold representation theorem, show great potential in modeling long-range dependencies. In this paper, we further investigate the potential of KAN for hyperspectral image restoration. Specifically, we propose a spatial-spectral attention block (SSAB) module, which includes a KAN-based spatial attention module (KAN-SpaAB) and a KAN-based spectral attention module (KAN-SpeAB), designed for the restoration of spatial and spectral information, respectively. Experimental results demonstrate that KSSANet outperforms existing methods in both quantitative evaluation and image generation quality, achieving state-of-the-art (SOTA) performance. Our code is available at: <https://github.com/Baisonm-Li/KSSANet>.

**Index Terms**—hyperspectral image, super resolution, Kolmogorov–Arnold Networks (KAN), attention module.

## I. INTRODUCTION

Hyperspectral images (HSIs) widely applied in various fields [1]–[7] due to their rich spectral and spatial information. However, the limitations of physical imaging technology make it challenging to directly acquire high resolution hyperspectral images (HR-HSIs). To address this issue, hyperspectral image super-resolution (HSI-SR) techniques have been developed, aiming to generate HR-HSIs from known low-resolution hyperspectral image (LR-HSIs). The main advantage of SHSR is that it does not require additional auxiliary data, and can enhance spatial resolution using only a single hyperspectral image, offering greater independence and flexibility in application. Currently, SHSR methods can be broadly categorized into two types: traditional approaches and deep learning (DL)-based approaches.

Traditional HSI-SR methods typically rely on handcrafted priors, such as low-rank approximation [8]–[12] and sparse coding [13]–[17], to construct mapping functions from LR-HSIs to HR-HSIs. These methods heavily depend on prior

knowledge of the images or scenes, often requiring manually designed features or assumed constraints. However, given the complex and variable nature of HSIs, acquiring accurate prior knowledge is challenging, which limits the performance of traditional approaches. Over the past decade, DL-based methods have developed rapidly, with CNN-based SR networks [18]–[22] achieving significant breakthroughs due to their strong spatial feature modeling capability. However, CNNs face the limitation when reconstructing long-range spectral information in HSIs. To address this issue, Transformer-based networks [23]–[27] have been introduced into the HSI-SR task. The self-attention mechanism of Transformers enables them to capture long-range dependencies and provide robust spectral sequence modeling capability. However, the introduction of Transformer also brings new challenges. First, Transformer-based methods typically require large amounts of training data to achieve stable convergence, but the available HSIs for training are extremely scarce. Second, for a spectral token of length  $n$ , the self-attention mechanism has a complexity of  $O(n^2)$ , which results in an unacceptable computational cost when processing complex high-dimensional HSIs.

Recently, the Kolmogorov–Arnold Network (KAN) [28], [29], a powerful new neural network architecture, has been introduced. Unlike traditional architectures, KANs apply activation functions to the connections between nodes, and these functions can be learned and adapted during training. KANs exhibit strong local plasticity and can avoid catastrophic forgetting by leveraging the locality of splines (a single sample affects only a few nearby spline coefficients).

Specifically, KSSANet achieves its super-resolution capability through multiple cascaded Spatial-Spectral Attention Blocks (SSAB). Each SSAB consists of two key submodules: a KAN-based Spatial Attention Block (KAN-SpaAB) and a KAN-based Spectral Attention Block (KAN-SpeAB). KAN-SpaAB is mainly responsible for reconstructing and optimizing spatial details, while KAN-SpeAB focuses on restoring spectral features. In SSAB, KAN-SpaAB dynamically adjusts the importance scores of spatial pixels, enabling the network to better capture and restore subtle spatial features. Meanwhile, KAN-SpeAB optimizes the importance scores of spectral channels, ensuring that the spectral information of each channel is accurately represented. This joint adjustment mechanism

\*Corresponding Author

allows KSSANet providing precise balancing of spatial and spectral information during reconstruction. Extensive experimental results demonstrate that KSSANet not only excels in subjective visual quality but also significantly surpasses existing SR methods in objective quantitative metrics.

## II. PROPOSED METHOD

### A. Overall structure

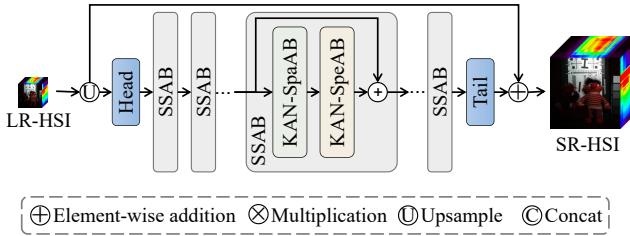


Fig. 1. Overview of the proposed KSSANet.

We denote  $\mathbf{I}_{LR} \in \mathbb{R}^{h \times w \times C}$  as the input LR-HSI,  $\mathbf{I}_{SR} \in \mathbb{R}^{H \times W \times C}$  as the output HR-HSI, and  $\mathbf{I}_{HR} \in \mathbb{R}^{H \times W \times C}$  as the ground truth (GT) for the input image  $\mathbf{I}_{LR}$  ( $C \gg c$ ). Our goal is to predict the  $\mathbf{I}_{SR}$  from the input  $\mathbf{I}_{LR}$  using the proposed HSI-SR network,

$$\mathbf{I}_{SR} = H_{Net}(\mathbf{I}_{LR}), \quad (1)$$

$H_{Net}$  denotes the mapping function of the proposed KSSANet. The overall structure of KSSANet is presented in Fig. 1. KSSANet first uses a *Head* module to project the LR-HSI into latent features. These latent features are then sequentially processed through several cascaded *SSABs*. Finally, a *Tail* module reconstructs the latent features into the shape of the HR-HSI as the predicted image,

$$\mathbf{X}_0 = f_{Head}(\mathbf{I}_{LR} \uparrow), \mathbf{X}_i = f_{SSAB}^i(\mathbf{X}_{i-1}), \mathbf{I}_{SR} = f_{Tail}(\mathbf{X}_l). \quad (2)$$

Let  $f_{Head}$  denote the mapping of the *Head* module, which include a  $3 \times 3$  convolution,  $\mathbf{I}_{LR} \uparrow$  represents the bicubic upsampling of LR-HSI. Let  $f_{SSAB}^i$  represent the mapping of the  $i$ -th *SSAB* module, and  $f_{Tail}$  denote the mapping of the *Tail* module, which consists of two  $3 \times 3$  convolutions, with the final convolution mapping the dimension to the number of spectral bands.  $\mathbf{X}_i \in \mathbb{R}^{H \times W \times D}$  represents the latent features output by the  $i$ -th *SSAB* module and the number of *SSAB* denoted as  $l$ .

### B. Kolmogorov-Arnold Networks (KAN)

KAN [28] originates from the Kolmogorov-Arnold representation theorem [30]–[32] and involves a stack of nonlinear, learnable activation functions that reshape the learning process in neural networks. In the KAN layer, each node is fully connected to every node in the subsequent layer. For each edge, an independent, trainable activation function is applied. At each node, only a summation operation is performed over all incoming edges. The learnable activation functions are defined as weighted sums of B-splines, with the B-spline [33], [34] basis functions denoted as  $B_i$ , and the fixed residual function chosen as the Sigmoid Linear Unit (SiLU),

$$f_{KAN}(x) = w_1 \cdot \text{SiLU}(x) + w_2 \cdot \sum_{i=0}^{G+k-1} c_i \cdot B_i(x). \quad (3)$$

The weights  $w_1, w_2$  and the basis function coefficients  $c_i$  are trainable parameters of spline. The basis function  $B_i$  is chosen as a  $k$ -th

degree polynomial, with the default value  $k = 3$ . The grid parameter  $G$  determines the degree of B-spline construction with the default value  $G = 5$ . Although spline-based KANs have ideal mathematical theories, original KAN still struggle with effective HSIs restoration.

### C. KAN-SpaAB

We design a KAN-based spatial attention block (KAN-SpaAB) for reconstructing spatial details of HSIs, which is expressed as:

$$\mathbf{X}_{Spa} = f_{Spa}(\mathbf{X}_{i-1}), \quad (4)$$

$f_{Spa}$  represents the mapping of the KAN-SpaAB, and  $\mathbf{X}_{Spa}$  represents the output of the KAN-SpaAB. Details as the left part of Fig. 2. To alleviate the limitations of small-kernel convolutions, such as their limited receptive field and inadequate global spatial modeling capability, KAN-SpaAB draws inspiration from the development of large-kernel convolutions [22], [35]. It employs a cascaded combination of decomposed  $\text{Conv}5 \times 5$  and  $\text{Conv}3 \times 3$  layers to approximate a  $\text{Conv}11 \times 11$ . The output features from these convolutions are concatenated along the channel dimension, then passed through a depthwise convolution ( $DWConv$ ) [36], which reduces the channel dimension to 1, resulting in an  $H \times W \times 1$  spatial score matrix. Leveraging the strong local plasticity of the KAN network, this spatial score matrix is further refined through two KAN layers using spline transformations. Finally, the spatial matrix is multiplied with the input  $\mathbf{X}_{i-1}$  along the spatial dimension, followed by a residual operation. Next, the output of KAN-SpaAB,  $\mathbf{X}_{spa}$ , is fed into KAN-SpeAB for spectral channel modeling.

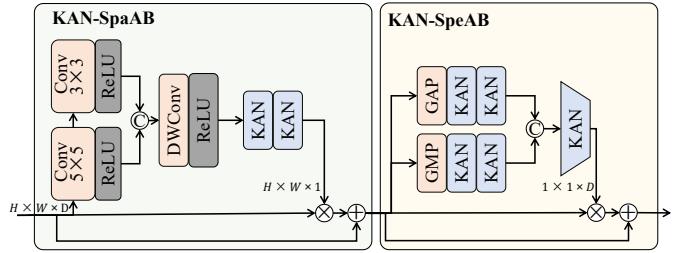


Fig. 2. Structure diagrams of KAN-SpaAB (left) and KAN-SpeAB (right). “Conv3x3” and “Conv5x5” denote 2D convolutions with kernel sizes of 3 and 5, respectively. “ReLU” refers to the ReLU activation function, while “DWConv” represents depthwise separable convolution. “GAP” and “GMP” stand for global average pooling and global max pooling, respectively.

### D. KAN-SpeAB

The KAN-based spectral attention block (KAN-SpeAB) is formulated as follows,

$$\mathbf{X}_i = f_{Spe}(\mathbf{X}_{spa}), \quad (5)$$

and the detailed structure is shown on the right side of Fig. 2. Inspired by SENet [37], we utilize globally pooled spatial features as channel excitation features. Specifically, to retain both prominent and global features of the hyperspectral image, we apply global max pooling (GMP) and global average pooling (GAP) to extract two distinct sets of channel excitation features (both size  $1 \times 1 \times D$ ). These excitation features are then refined through two KAN layers using spline adjustments. Afterward, the two features are concatenated along the channel dimension (forming a feature of size  $1 \times 1 \times 2D$ ) and fused via an alignment KAN layer (output feature of size  $1 \times 1 \times D$ ). This produces the final channel importance score matrix, which is multiplied by the input  $\mathbf{X}_{spa}$  along the channel dimension, followed by a residual operation. The output of the entire SSAB module is denoted as  $\mathbf{X}_i$ , as followed by a residual operation.



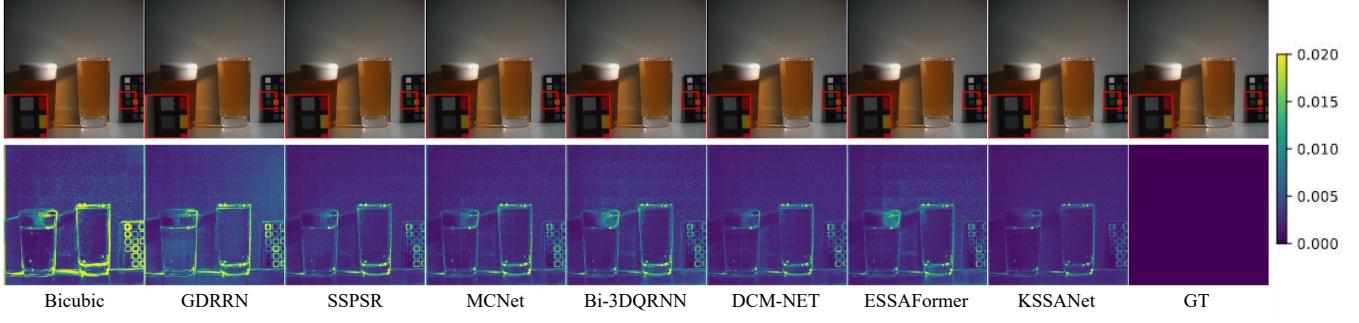


Fig. 3. Visual quality comparison on the CAVE dataset for  $\times 4$  SR. The first row presents pseudo-color images (R-29, G-12, B-4), while the second row displays the corresponding MSE heatmaps.

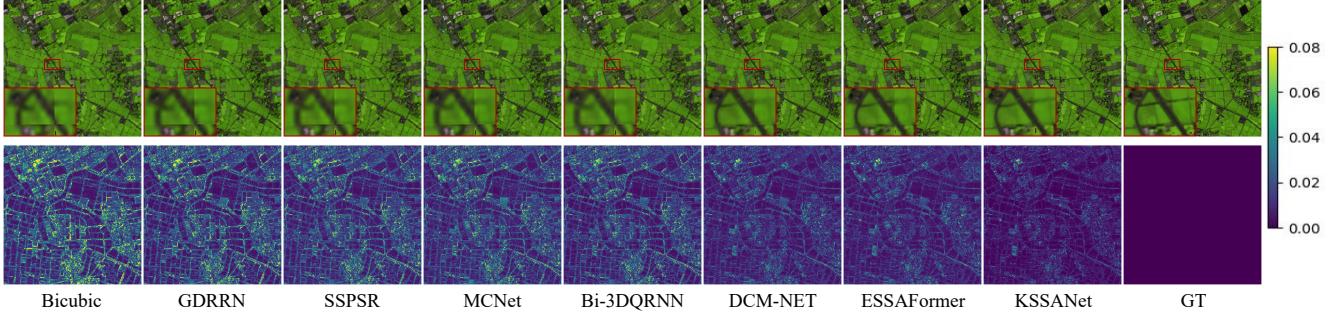


Fig. 4. Visual quality comparison on the Chikusei dataset for  $\times 4$  SR. The first row presents pseudo-color images (R-29, G-12, B-4), while the second row displays the corresponding MSE heatmaps.

two modules. Specifically, we replace KAN-SpaAB and KAN-SpeAB with the *Spatial Residual Module* and *Spectral Attention Residual Module* from SSPSR [21], respectively, and retrain and test the model on the Chikusei  $\times 4$  dataset. The test results are shown in Table III, although the proposed KAN-SpeAB has larger size and higher computational cost, it delivers significant performance improvements. With the introduction of spline-based KAN, both KAN-SpaAB and KAN-SpeAB demonstrate greater local flexibility, enabling the generation of higher-quality attention matrices, which in turn leads to enhanced performance. To validate the powerful ability

TABLE III

ABLATION COMPARISON ON THE CHIKUSEI DATASET FOR  $\times 4$  SR. THE *Spatial Residual Module* AND *Spectral Attention Residual Module* OF SSPSR [21] ARE RESPECTIVELY USED AS ABLATION REPLACEMENTS FOR KAN-SPAAB AND KAN-SPEAB.

KAN-SpaAB	KAN-SpeAB	#Params(M)	#MACs(G)	PSNR↑	SSIM↑
✓	✓	4.92	20.15	41.02	0.952
✓	✗	0.72	2.89	34.26	0.879
✗	✓	7.29	29.81	39.26	0.928
✗	✗	3.09	12.55	37.89	0.901

of KAN in generating attention matrices, we conduct further ablation experiments on Chikusei  $\times 4$  dataset. Specifically, we replace all KAN layers with linear layers to evaluate the advantage of spline-based KAN in terms of local flexibility. After replacing KAN with linear layers, the model size reaches 5.72MB and MACs reach 20.16G, while PSNR and SSIM drop to 38.14dB and 0.907, respectively. Fig. 5 further displays the performance changes during the training process of the two comparison models, showing that the KAN-driven model exhibits more significant improvements in PSNR and reductions in loss value. These experimental results demonstrate the effectiveness of the KAN-driven KSSANet in HSI-SR task.

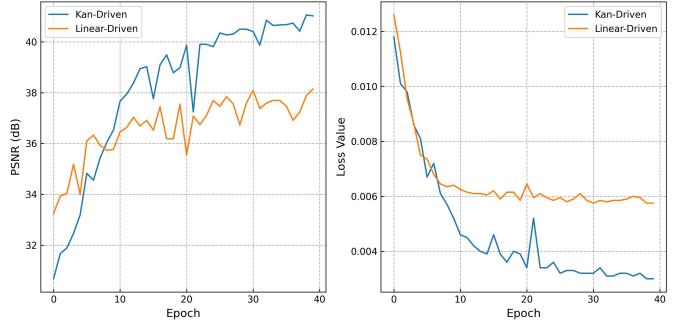


Fig. 5. Performance comparison of KAN-driven and Linear-driven models during training for the  $\times 4$  SR on the Chikusei dataset.

#### IV. CONCLUSION

This paper proposes a KAN-driven single hyperspectral image SR model called KSSANet. Specifically, we leverage the powerful capability of KAN for local adjustments in sequences to generate high-quality attention matrices. We further design a KAN-based spatial attention module (KAN-SpaAB) and a KAN-based spectral attention module (KAN-SpeAB), which are used respectively for image spatial and spectral restoration. Extensive experiments demonstrate that KSSANet not only outperforms existing methods in both quantitative metrics and visual quality.

#### V. ACKNOWLEDGMENTS

This research was supported by the National Key Research and Development Program of China (No. 2023YFB4502304)

## REFERENCES

- [1] H. Zhang, C. Gong, Y. Bai, Z. Bai, and Y. Li, "3d-anas: 3d asymmetric neural architecture search for fast hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, p. 1–19, Jan 2022.
- [2] F. Xiong, J. Zhou, and Y. Qian, "Material based object tracking in hyperspectral videos," *IEEE Transactions on Image Processing*, p. 3719–3733, Jan 2020.
- [3] P. Thenkabail, J. Lyon, and A. Huete, "Hyperspectral remote sensing of vegetation," *CRC Press eBooks,CRC Press eBooks*, Apr 2016.
- [4] M. Borensgasser, W. Hungate, and R. Watkins, "Hyperspectral remote sensing: Principles and applications," Dec 2007.
- [5] G. M. ElMasry and S. Nakuchi, "Image analysis operations applied to hyperspectral images for non-invasive sensing of food quality – a comprehensive review," *Biosystems Engineering*, p. 53–82, Feb 2016.
- [6] D. Haboudane, "Hyperspectral vegetation indices and novel algorithms for predicting green lai of crop canopies: Modeling and validation in the context of precision agriculture," *Remote Sensing of Environment*, p. 337–352, Apr 2004.
- [7] S. Qian, "Overview of hyperspectral imaging remote sensing from satellites," Nov 2022.
- [8] H. Irmak, G. B. Akar, and S. E. Yuksel, "A map-based approach for hyperspectral imagery super-resolution," *IEEE Transactions on Image Processing*, p. 2942–2951, Jun 2018.
- [9] Y. Wang, X. Chen, Z. Han, and S. He, "Hyperspectral image super-resolution via nonlocal low-rank tensor approximation and total variation regularization," *Remote Sensing*, p. 1286, Dec 2017.
- [10] T. Xu, T. Z. Huang, L. J. Deng, X. L. Zhao, and J. Huang, "Hyperspectral image superresolution using unidirectional total variation with tucker decomposition," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, 2020.
- [11] T. Xu, T. Z. Huang, L. J. Deng, X. L. Zhao, and J. Huang, "Hyperspectral image superresolution using unidirectional total variation with tucker decomposition," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, 2020.
- [12] N. Yokoya, S. Member, IEEE, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 2, pp. 528–537, 2012.
- [13] H. Huang, A. G. Christodoulou, and W. Sun, "Super-resolution hyperspectral imaging with unknown blurring by low-rank and group-sparse modeling," in *2014 IEEE International Conference on Image Processing (ICIP)*, Oct 2014.
- [14] Y. Fu, A. Lam, I. Sato, and Y. Sato, "Adaptive spatial-spectral dictionary learning for hyperspectral image restoration," *International Journal of Computer Vision*, p. 228–245, Apr 2017.
- [15] W. Dong, F. Fu, G. Shi, X. Cao, J. Wu, G. Li, and X. Li, "Hyperspectral image super-resolution via non-negative structured sparse representation," *IEEE Transactions on Image Processing*, pp. 2337–2352, 2016.
- [16] J. M. Bioucas-Dias and M. A. T. Figueiredo, "alternating direction algorithms for constrained sparse regression: application to hyperspectral unmixing," 2017.
- [17] W. Dong, F. Fu, G. Shi, X. Cao, J. Wu, G. Li, and X. Li, "Hyperspectral image super-resolution via non-negative structured sparse representation," *IEEE Transactions on Image Processing*, pp. 2337–2352, 2016.
- [18] Y. Li, J. Hu, X. Zhao, W. Xie, and J. Li, "Hyperspectral image super-resolution using deep convolutional neural network," *Neurocomputing*, p. 29–41, Nov 2017.
- [19] Q. Li, Q. Wang, and X. Li, "Mixed 2d/3d convolutional network for hyperspectral image super-resolution," *Remote Sensing*, p. 1660, May 2020.
- [20] Yang, Zhao, Chan, and Xiao, "A multi-scale wavelet 3d-cnn for hyperspectral image super-resolution," *Remote Sensing*, p. 1557, Jul 2019.
- [21] J. Jiang, H. Sun, X. Liu, and J. Ma, "Learning spatial-spectral prior for super-resolution of hyperspectral imagery," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1082–1096, 2020.
- [22] C. Zhang, M. Zhang, Y. Li, X. Gao, and S. Qiu, "Difference curvature multidimensional network for hyperspectral image super-resolution," *Remote Sensing*, vol. 13, no. 17, p. 3455, 2021.
- [23] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, *End-to-End Object Detection with Transformers*, p. 213–229, Jan 2020.
- [24] L. Zhou, Y. Zhou, J. J. Corso, R. Socher, and C. Xiong, "End-to-end dense video captioning with masked transformer," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun 2018.
- [25] J.-F. Hu, T.-Z. Huang, and L.-J. Deng, "Fusformer: A transformer-based fusion approach for hyperspectral image super-resolution," *Cornell University - arXiv,Cornell University - arXiv*, Sep 2021.
- [26] M. Zhang, C. Zhang, Q. Zhang, J. Guo, X. Gao, and J. Zhang, "Essaformer: Efficient transformer for hyperspectral image super-resolution," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 23073–23084, 2023.
- [27] Q. Ma, J. Jiang, X. Liu, and J. Ma, "Reciprocal transformer for hyperspectral and multispectral image fusion," *Information Fusion*, vol. 104, p. 102148, 2024.
- [28] Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljačić, T. Y. Hou, and M. Tegmark, "Kan: Kolmogorov-arnold networks," 2024.
- [29] Z. Liu, P. Ma, Y. Wang, W. Matusik, and M. Tegmark, "Kan 2.0: Kolmogorov-arnold networks meet science," 2024.
- [30] V. M. Tikhomirov, *On the Representation of Continuous Functions of Several Variables as Superpositions of Continuous Functions of a Smaller Number of Variables*, p. 378–382, Jan 1991.
- [31] A. N. Kolmogorov, "On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition," in *Doklady Akademii Nauk*, vol. 114, pp. 953–956, Russian Academy of Sciences, 1957.
- [32] J. Braun and M. Griebel, "On a constructive proof of kolmogorov's superposition theorem," *Constructive Approximation*, p. 653–675, Dec 2009.
- [33] G. Farin, *Curves and surfaces for computer-aided geometric design: a practical guide*. Elsevier, 2014.
- [34] T. W. Sederberg and S. R. Parry, "Free-form deformation of solid geometric models," in *Proceedings of the 13th annual conference on Computer graphics and interactive techniques*, pp. 151–160, 1986.
- [35] X. Ding, X. Zhang, Y. Zhou, J. Han, G. Ding, and J. Sun, "Scaling up your kernels to 31x31: Revisiting large kernel design in cnns," *arXiv preprint arXiv:2203.06717*, 2022.
- [36] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251–1258, 2017.
- [37] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7132–7141, 2018.
- [38] Y. Li, J. Hu, X. Zhao, W. Xie, and J. Li, "Hyperspectral image super-resolution using deep convolutional neural network," *Neurocomputing*, vol. 266, pp. 29–41, 2017.
- [39] Q. Li, Q. Wang, and X. Li, "Mixed 2d/3d convolutional network for hyperspectral image super-resolution," *Remote sensing*, vol. 12, no. 10, p. 1660, 2020.
- [40] Y. Fu, Z. Liang, and S. You, "Bidirectional 3d quasi-recurrent neural network for hyperspectral image super-resolution," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 2674–2688, 2021.
- [41] N. Yokoya and A. Iwasaki, "Airborne hyperspectral data over chikusei," Tech. Rep. SAL-2016-05-27, Space Application Laboratory, University of Tokyo, Japan, May 2016.
- [42] P. Gamba, "A collection of data for urban area characterization," in *IEEE International IEEE International Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. Proceedings. 2004*, Dec 2004.
- [43] R. Yuhas, A. Goetz, and J. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (sam) algorithm," Jun 1992.
- [44] L. Wald, "Data fusion. definitions and architectures - fusion of images of different spatial resolutions," *Le Centre pour la Communication Scientifique Directe - HAL - Diderot,Le Centre pour la Communication Scientifique Directe - HAL - Diderot*, Jan 2002.
- [45] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, p. 600–612, Apr 2004.
- [46] D. P. Kingma, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.