



# Special Paper

방송공학회논문지 제29권 제7호, 2024년 12월 (JBE Vol. 29, No. 7, December 2024)

<https://doi.org/10.5909/JBE.2024.29.7.1168>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

## Levels-of-Experts based on Hard Segmentations for Improving Implicit Neural Networks

Myeong-Jun Oh<sup>a)</sup> and Sung-Ho Bae<sup>a)\*</sup>

### Abstract

Implicit neural representation is used in various applications such as super-resolution, 3D reconstruction, and image compression. Because continuous signals can be generated at an arbitrary scale through the interpolation ability inherent in neural networks by storing signals in neural networks. This paper aims to alleviate the problem that existing implicit neural representation neural networks have poor ability to reconstruct high-frequency characteristics. In other words, this paper proposes a new technique to increase high-frequency reconstruction performance by applying explicit characteristics using hard segmentation to the implicit neural representation model. This paper is based on a level-of-Experts method that selects a specific weight among several weights according to the position. In this paper, we propose a method to assign different weights according to each segmentation area by using hard segmentation based on luminance in an image rather than the existing even-sized tile format. Through this, the proposed method induces each weight in the layer to predict data values with similar characteristics, so that the high-frequency signal can be trained. As a result of experiments on the Kodak dataset, it was shown that the average PSNR was significantly improved by 0.84dB compared to the previous one when the proposed method was used.

Keyword : Deep learning, Implicit neural representation, Hard segmentation

## I. Introduction

Recently, many studies have been conducted to in-

corporate deep neural network technologies into the field of image compression. Among them, image compression research using Implicit Neural Representation (INR) has been actively studied<sup>[2,3,4,5]</sup> where images are stored in the neural networks in the form of weights.

Traditionally, image compression tasks are mainly based on standardized algorithms such as Run-Length Encoding<sup>[29]</sup>, Entropy Encoding<sup>[30]</sup>, Discrete Cosine Transform and Quantization<sup>[31]</sup>. However, the methods using INR have attracted attention as they reached a level that exceeds the compression rate of conventional JPEG methods<sup>[5]</sup>. Compressed images using INR not only achieve a compression

a) Department of Artificial Intelligence at Kyung Hee University

‡ Corresponding Author : Sung-Ho Bae

E-mail: shbae@khu.ac.kr

Tel: +82-31-201-2593

ORCID: <https://orcid.org/0000-0002-3389-1159>

※ This research was supported by the three MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support programs (IITP-2023-RS-2023-00258649), (IITP-2024-RS-2023-00259004) and (IITP-2024-RS-2024-00438239), supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

• Manuscript October 29, 2024; Revised December 2, 2024; Accepted December 3, 2024.

rate comparable to or better than that of JPEG method, but also theoretically allow for infinite resolution, which gives them an advantage over traditional methods<sup>[2,4]</sup>.

INR represents images by using a neural network model that takes the coordinates of pixels within the image as in-

puts and outputs the corresponding pixel values. This is possible because the neural network can predict continuous pixel value outputs when continuous position coordinates are taken<sup>[6,7,15]</sup>.

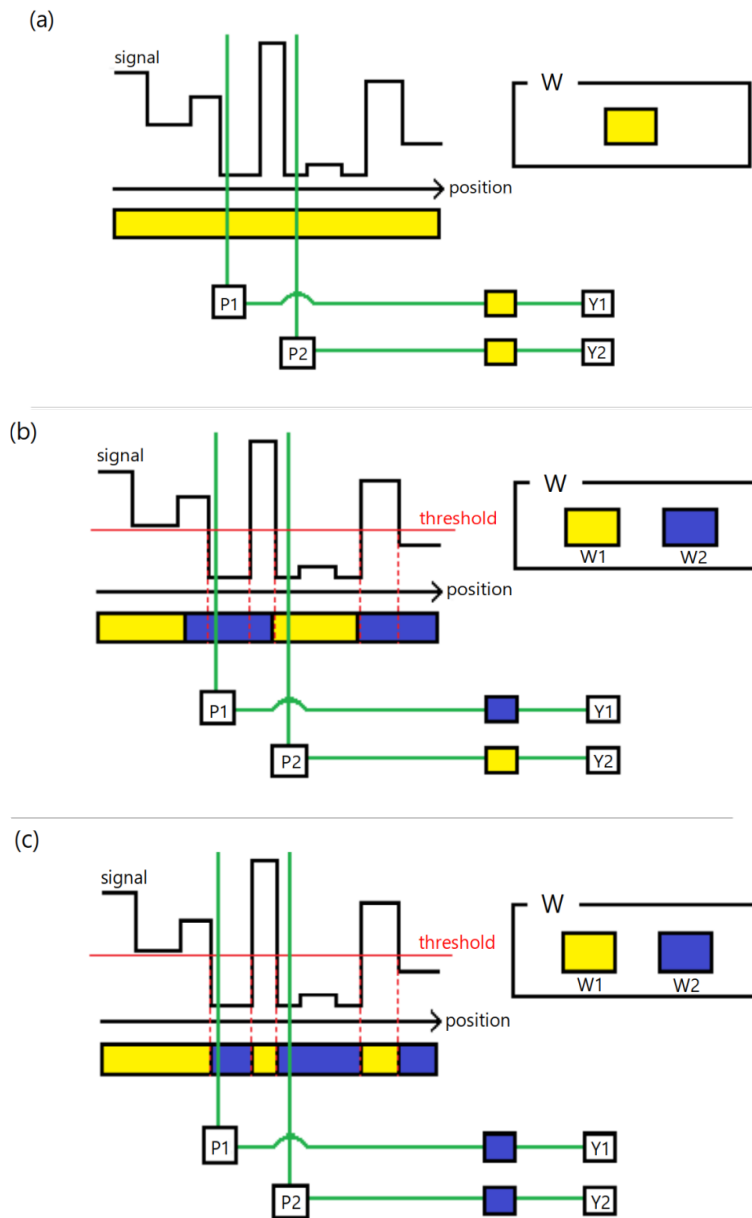


Fig. 1. Comparison in overview of the linear layer of (a) normal MLP used in INR<sup>[8]</sup>, (b) Level-Of-Experts, and (c) the proposed hard segmentation-based Level-Of-Experts

To improve performance of INR models, a positional-dependent weight layer approach, which is called Level-of-Experts (LoE), has been proposed<sup>[1]</sup>. Fig. 1 illustrates a normal MLP's linear layer which is used in INR model<sup>[8]</sup> (a), a model's using the LoE method (b), and the proposed model's using a LoE method based on hard-segmentation (c). In Fig. 1, the signal has values based on their positions (coordinates), and  $W1$  and  $W2$  represent to sets of weights (experts) shared in a layer.

The standard INR model in Fig. 1 (a) is designed with a single weight per node in the layer, which uses the same weight for computations ( $W$ ) regardless of the position input. However, the INR model using LoE method shown in Fig. 1 (b) introduces the set of two experts ( $W1$ ,  $W2$ ) in a layer, mapping weights evenly by region. And the proposed method in Fig. 1 (c) divides and assigns the experts by characteristics of signal. This method allows the positions of signals that have similar characteristics to be assigned to the same expert, achieving overfitting and enhancing the reconstruction performance of the INR model.

Fig. 1 demonstrates which weights are selected for output in each model when coordinates  $P1$  and  $P2$  are taken as inputs. In Fig. 1 (a), there is only single weight in the layer, so both  $P1$  and  $P2$  use the yellow weight ( $W$ ) regardless of their positions. In Fig. 1 (b), weights are assigned based on a predefined grid, so  $P1$  and  $P2$  use blue weight ( $W2$ ) and yellow weight ( $W1$ ) each in the same layer. By applying the LoE method to the model, it allows selection of different weights based on position, achieving more varied frequency representations instead of applying a single frequency across all areas.

However, this method has some limits. As shown in Fig. 1 (b), experts are mapped in a uniform (rectangular or cuboidal) region in the signal domain, which may not fully leverage the characteristics of the input signal. That is, the boundaries of the regions tend to generate discontinuity in continuous regions of signals. Since overfitting to the input signal is important for INR tasks, we determined that

grouping pixels with similar characteristics into the same region and using it for expert (i.e., the weight) mapping grid would be much more advantageous for signal reconstruction.

Thus, as shown in Fig. 1 (c), we propose the hard segmentation-based LoE. The proposed method first sets thresholds for hard segmentation, treating the resulting segmentation map as a grid similar to that in Fig. 1 (b) for training. Using this method, both  $P1$  and  $P2$ , coordinates of which signal values located below the threshold, are computed using the same blue weight in Fig. 1 (c), unlike in Fig. 1 (b). This approach enhances the specialization of the experts by assigning signal locations with similar characteristics to the same expert, unlike the conventional LoE method.

This paper proposes a hybrid neural representation method, as shown in Fig. 1 (c), where weights are selected based on an explicit feature, segmentation generated by dividing the image based on luminance. By training with the characteristics of the input image, our method enables to accelerate overfitting, thus expecting significantly higher reconstruction performance than conventional methods. Moreover, the proposed method exploits the discontinuities between grid regions by separating by segments that are not continuous with each other originally. Experimental results show that the PSNR is increased by 0.84dB in average compared to the existing LoE method.

## II. Related Work

### 1. Implicit Neural Representation

INR is a method of representing signal data as a continuous function through neural networks. This approach takes implicit features such as coordinates of the signal as input and trains a neural network that outputs the data values at the corresponding location. It can represent various signals, such as audio<sup>[8,14]</sup>, images<sup>[8,15,16]</sup>, and 3D mod-

els<sup>[8,17,18]</sup> with continuous functions. Unlike discrete representation methods like pixels or voxels, INR allows for resolution-independent representation<sup>[6,7]</sup>. The neural network structure primarily uses a Multi-Layer Perceptron (MLP), which has applications across fields such as super-resolution<sup>[6,7,15]</sup>, data compression<sup>[2,3,4,5]</sup>, and 3D rendering<sup>[8,17,18]</sup>.

To improve INR, many studies have been researched such as providing additional information to the input<sup>[18,21]</sup>, modifying the activation function<sup>[8,19,20]</sup>, or changing the network structure<sup>[23]</sup>. Among these, SIREN (Sinusoidal Representation Network)<sup>[8]</sup>, which uses a sine function as the activation function, effectively represents complex signals and patterns by leveraging its periodic characteristics. SIREN particularly captures high-frequency information more effectively than methods that use non-periodic functions such as ReLU<sup>[20,22]</sup>. Moreover, it provides a continuous and smooth representation and makes the entire network infinitely differentiable, demonstrating high performance in approximating complex functions and representing signals. This paper also aims to enhance performance with reconstruction of high frequency details by applying the sine activation function in SIREN.

INR with Level-of-Experts<sup>[1]</sup> is an extension of the conventional INR method. This approach utilizes multiple experts to effectively represent information across various frequency bands. Each expert is specialized in a specific frequency range, allowing it to capture diverse characteristics of the input signal. This method enables more accurate representation of both fine details and the overall structure of the image, showing good performance in high-resolution data or signals with multi-scale characteristics. However, this method allocates regions to each expert based on predefined formulas, resulting in a structure where all experts are evenly assigned to represent the information. Thus, in this paper, we introduce a non-uniform region allocation method for LoE that assigns signal regions with similar characteristics to the same expert.

## 2. Hard Segmentation

Hard Segmentation refers to a technique that divides an image based on similarity in local signal characteristics within the image, such as pixel colors or line segments. The Watershed algorithm<sup>[25]</sup> is a method that segments an image by treating it as a topographic surface. In this approach, pixel brightness values are treated as elevations, and it is assumed that water starts filling from markers, designated either directly or through other algorithms. As water rises, watersheds will be formed where the waters from different makers merge, dividing the image into multiple regions. This method is particularly effective for separating objects that are in contact with each other.

The EM algorithm<sup>[26]</sup> assumes that pixels in an image follow different distributions in a six-dimensional space based on color and texture, and segments the image using a probabilistic model. This method iteratively calculates and updates the probability of each pixel belonging to a specific Gaussian cluster until convergence, resulting in the final image segmentation.

Statistical Region Merging (SRM)<sup>[27]</sup> is a method that performs image segmentation by merging visually similar segments based on semi-supervised statistical perceptual grouping principles and an image generation model.

There are many ways to segment images, but for fast computation, our paper uses A simple segmentation with luminance-based thresholding which is a basic and simple image segmentation method. This method divides an image into several groups based on pixel values. This method sets a specific threshold, either arbitrarily or through a specific algorithm<sup>[10,11]</sup>, and classifies pixels into different groups depending on whether their values exceed the threshold. Thresholding has the advantage of being easy to implement and fast to compute, making it useful in various applications such as medical image analysis and satellite image processing.

In this paper, we use the Multi-Otsu method<sup>[11]</sup> based on

luminance. The Multi-Otsu Method is an extension of Otsu's binarization method, the Otsu Method<sup>[10]</sup>, to classify an image into more than two classes. The Otsu Method is an automatic threshold selection algorithm for image binarization, used to split an image into two or more classes. It analyzes the histogram of the image and finds the optimal threshold that maximizes inter-class variance and minimizes intra-class variance when dividing the image into two classes. The Multi-Otsu Method extends this to divide the image into multiple classes, thereby automatically determining thresholds even for complex images.

### III. Proposed Method

Fig. 1 (b) illustrates the conventional LoE method alongside the proposed method.

An  $L$  layers INR model that takes 2D coordinates  $X$  as input and outputs a 3D RGB color value  $Y$  can be expressed mathematically as follows:

$$Z^0 = X^0 \quad (1)$$

$$Z^i = \Phi(W^i Z^{i-1} + b^i), i = 1, \dots, L-1 \quad (2)$$

$$Y = W^L Z^{L-1} + b^1 \quad (3)$$

$W^i$  and  $b^i$  represent the weight and bias of the  $i$ -th layer, respectively, and  $\Phi$  means activation function. In the case of SIREN, a sine function is used here.

In the LoE paper, when a  $2 \times 2$  allocation structure is applied to a model for 2D images, they can be represented as  $W^i = \{W_{0,0}^i, W_{0,1}^i, W_{1,0}^i, W_{1,1}^i\}$ ,  $b^i = \{b_{0,0}^i, b_{0,1}^i, b_{1,0}^i, b_{1,1}^i\}$ ,  $W^i$  can be expressed mathematically as follows:

$$W^i = W_{\lfloor x \rfloor \bmod 2, \lfloor y \rfloor \bmod 2}^i, \text{ where } \begin{bmatrix} x \\ y \end{bmatrix} = A^i x + b^i \quad (4)$$

At this point,  $A^i, b^i$  are predefined, and the values used

in this paper are 128, 0, 8, 0, 4, 0 for  $A^1, b^1, A^2, b^2, A^3, b^3$ . In this paper, we propose an approach to assign regions based on thresholding segmentation from the segmented images, instead of using the last layer of (4) to assign regions, so that it could allocate them in a non-uniform shape.

The overview of the proposed method is shown in Fig. 1 (c). It is nearly identical to the Fine-to-Coarse tiling method in the conventional INR with LoE, but in the first tiling step, we replaced the conventional tiling method with a Multi-Otsu method based on luminance to divide the image into a total of four regions. The function used in the model is the sine function, as employed in SIREN.

### IV. Experiment

For the experiments, the KODAK dataset<sup>[12]</sup> was used. The KODAK dataset consists of 24 images of 24-bit color with a resolution of  $768 \times 512$ , and it is primarily used for performance analysis of image processing or compression algorithms. The data were normalized so that the pixel coordinates and values of the images fall within the range of -1 to 1 for training purposes. The experimental results were evaluated using PSNR, which compares the quality difference between the original and processed images. The model structure has an input dimension of 2 and an output dimension of 3, with an MLP containing three hidden layers of 256 nodes each, trained over a total of 5,000 iterations. We used Mean Squared Error for loss and Adam with  $\beta_1 = 0.9, \beta_2 = 0.999$  for optimization. We implemented models with PyTorch<sup>[28]</sup>, and experimented on a RTX3090 GPU environment.

Table 1 summarizes the experimental results on the KODAK dataset for various methods. In the LoE method, FtC and QT refer to "Fine to Coarse" and "Quad Tree," respectively, where the grid size increases from small to large, and decreases from large to small<sup>[1]</sup>, 128, 8, 4, 2,

Table 1. PSNR Comparison Table for Different Methods using KODAK Dataset

Method	Average PSNR(dB)	Highest PSNR(dB)	Lowest PSNR(dB)
SIREN	30.28	35.00	23.75
LoE(FtC)	38.41	41.15	35.39
LoE(QT)	36.35	39.42	31.98
Ours	39.25	42.04	36.45

0 for each  $A^1, A^2, A^3, A^4, b^{1,2,3,4}$  in FtC and 2, 4, 8, 128, 0 for each  $A^1, A^2, A^3, A^4, b^{1,2,3,4}$  in QT, so that it could assign with smallest( $3 \times 2$ ) and second biggest( $192 \times 128$ ) tiles in one model. And we use LoE method based on SIREN since it performs better than models in [1], using LeakyReLU for activation function.

In Table 1, Result shows our method's image reconstruction perform the best among methods in this table. Compare with LoE (FtC) method, which is best from existing methods in this table, our method exceeds PSNR with 0.84dB in average, and 1.06dB in lowest.

In Fig. 2., SIREN model shows slow training speed than other methods, And 'Fine to Coarse' and 'Quad Tree' method shows struggle to train edge region in images. However, Our method can train model fast and evenly even in edge regions.

In the ablation studies, we experiment to analyze position of segmentation layer and method of segmentation.

Table 2 shows the results based on the position of the segmentation layer applied. When changing the position of

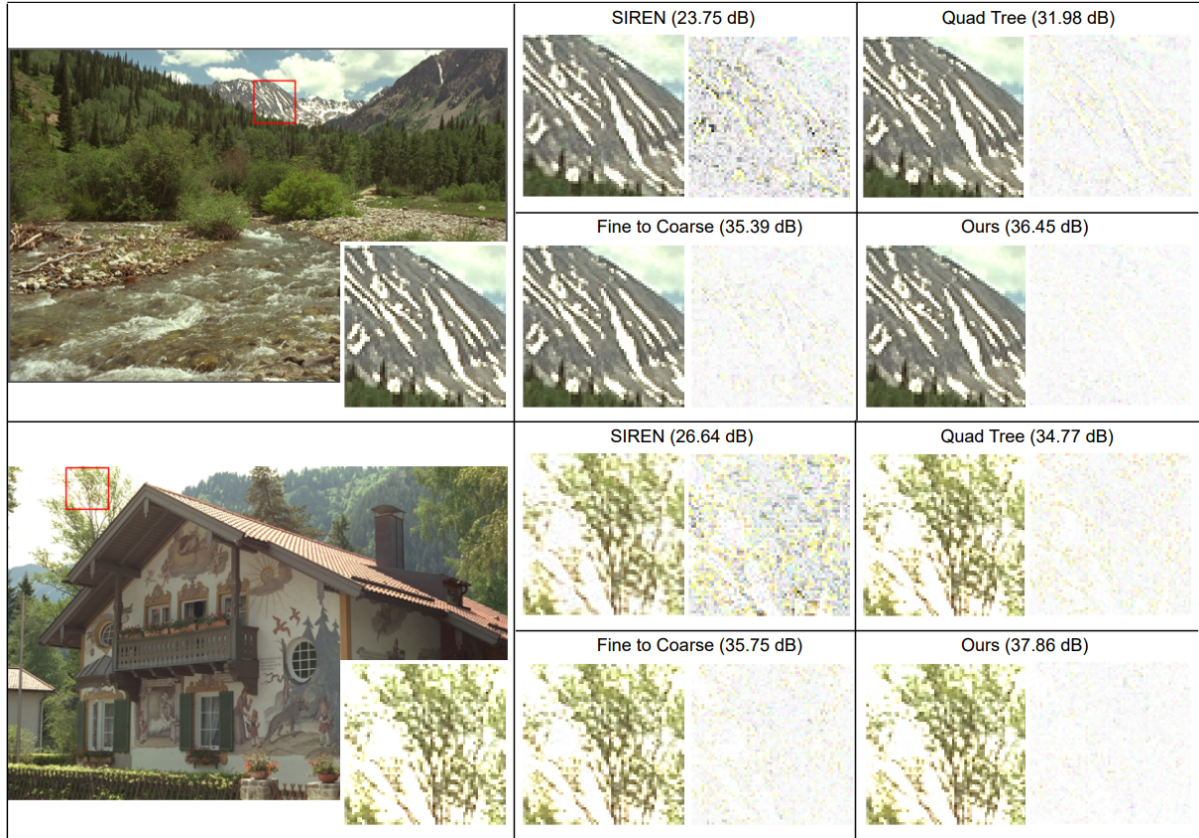


Fig. 2. Visualization of experimental result. Left figure is full and cropped original image, and each cell in right figure shows cropped reconstructed image(left), and normalized RGB channel error map(right)



Table 2. PSNR Comparison Table for Different positions of segmentation layer using KODAK Dataset

Layer pos.	Average PSNR(dB)	Highest PSNR(dB)	Lowest PSNR(dB)
First layer	37.51	40.72	34.31
Second layer	37.53	40.73	34.03
Third layer	38.34	41.49	34.53
Fourth layer	39.25	42.04	36.45

segmentation layer, except for the first layer, the rest were implemented by replacing the layers at that location in the LoE (FtC) method. About first layer, we replaced second layer to first layer first, and then we replaced first layer to segmentation layer. The experimental results show that when the segmentation layer is located in fourth layer

Table 3. PSNR Comparison Table for Different segmentation methods of segmentation layer using KODAK Dataset

Seg. method	Average PSNR(dB)	Highest PSNR(dB)	Lowest PSNR(dB)
Equal bin	39.02	42.05	35.82
Ours	39.25	42.04	36.45

performs best. We think it is because big chunks of segmentations in segmentation layer affects more to results. In Fig. 3., We can get much less errors noticeably in error maps when we use Otsu method layer in the last layer.

Table 3 shows the results based on the position of the segmentation layer applied. Equal bin means setting thresholds to separate pixels in groups almost equally, and ours is Multi-Otsu method. The experimental results show that

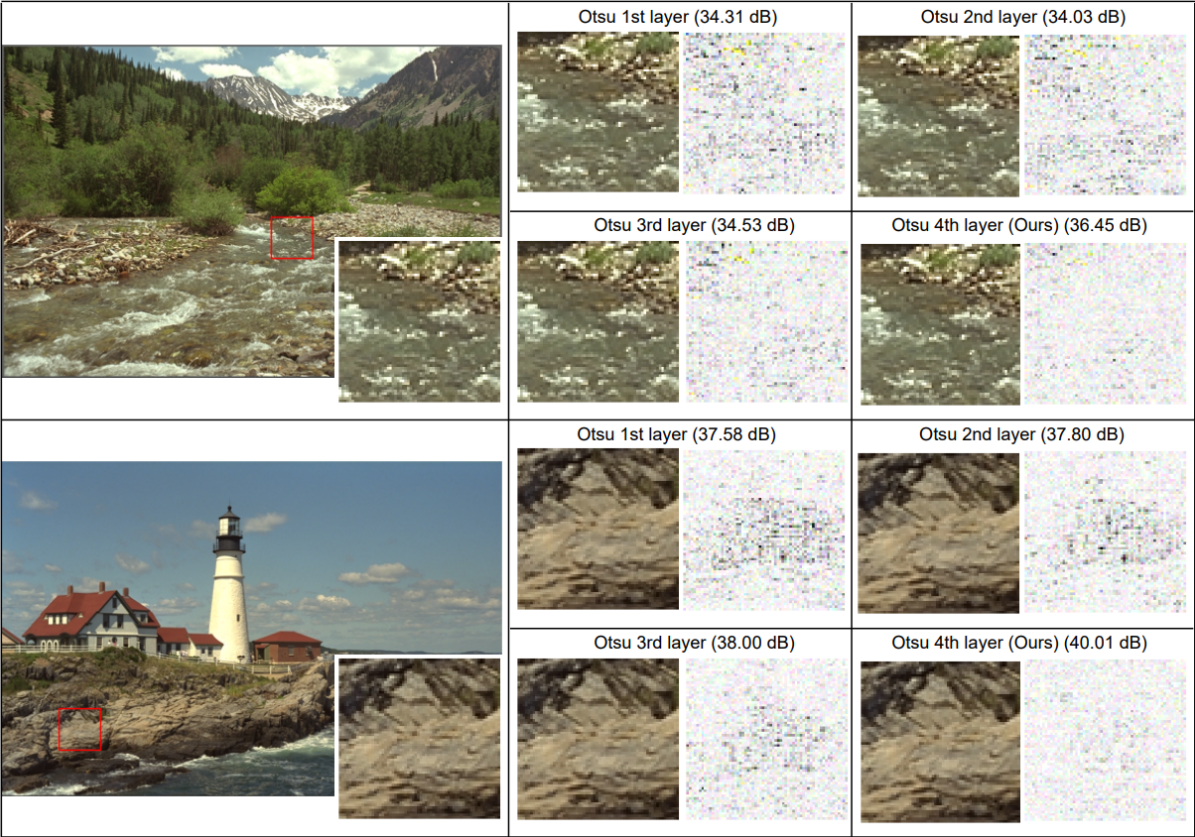


Fig. 3. Visualization of ablation experimental result in table 2. Left figure is full and cropped original image, and each cell in right figure shows cropped reconstructed image(left), and normalized RGB channel error map(right)

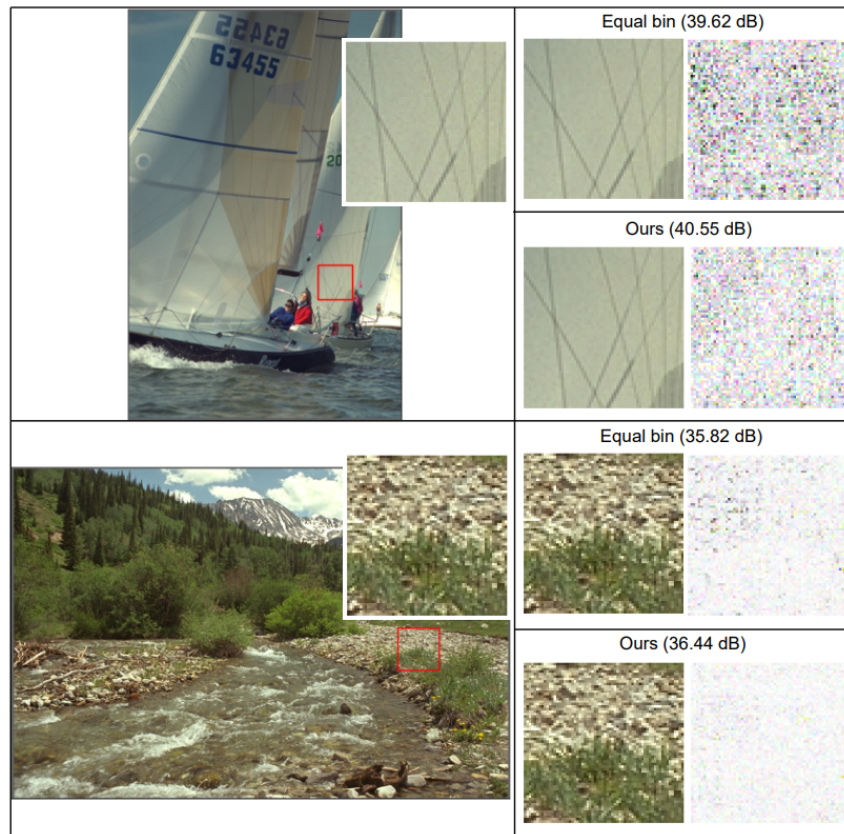


Fig. 4. Visualization of ablation experimental results in table 3. Left figure is full and cropped original image, and each cell in right figure shows cropped reconstructed image(left), and normalized RGB channel error map(right)

equal bin shows highest PSNR, but it shows lower PSNR than ours overall, so its average PSNR is lower than ours. In Fig. 4., We can see 2 method shows similar patterns in error maps, but ours mispredict pixel values evenly, so ours can get higher ‘Lowest PSNR’ than equal bin method.

## V. Conclusion

In this study, we propose a hybrid neural representation method that uses the luminance characteristics of images for segmentation, thereby leveraging explicit features. Through this method, we were able to train neural representation model that shows higher performance compared to existing

methods.

Based on experimental results, we found that inducing rapid overfitting through our method led to enhanced image reconstruction performance. Also, experiments using additional images confirmed the importance of inducing overfitting in INR. Therefore, we expect that if we can apply methods that induce greater overfitting on the signals than methods used in this paper, It is possible to achieve even better results.

## Reference

- [1] Zekun Hao, Arun Mallya, Serge Belongie, and Ming-Yu Liu, “Implicit Neural Representations with Levels-of-Experts,” *Advances in Neural*



- Information Processing Systems, Vol.35, pp. 2564-2576, 2022.  
<https://openreview.net/forum?id=St5q10aqLTO>
- [2] Yannick Strümpfer, Janis Postels, Ren Yang, Luc van Gool, and Federico Tombari, "Implicit Neural Representations for Image Compression," European Conference on Computer Vision, Cham Springer Nature Switzerland, pp. 74-91, October 2022.  
doi: [https://doi.org/10.1007/978-3-031-19809-0\\_5](https://doi.org/10.1007/978-3-031-19809-0_5)
- [3] Zongyu Guo, Gergely Flamich, Jiajun He, Zhibo Chen, José Miguel and Hernández-Lobato, "Compression with Bayesian Implicit Neural Representations," Advances in Neural Information Processing Systems, Vol.36, pp. 1938-1956, 2023.  
doi: <https://doi.org/10.48550/arXiv.2305.19185>
- [4] Emilien Dupont, Hrushikesh Loya, Milad Alizadeh, Adam Goliński, Yee Whye Teh, and Arnaud Doucet, "COIN++: Neural Compression Across Modalities," arXiv preprint, arXiv:2201.12904, 2022.  
doi: <https://doi.org/10.48550/arXiv.2201.12904>
- [5] Tuan Pham, Yibo Yang, and Stephan Mandt, "Autoencoding implicit neural representations for image compression," ICML 2023 Workshop Neural Compression: From Information Theory to Applications, 2023.  
<https://openreview.net/forum?id=bZn0XOm37w>
- [6] Yi Ting Tsai, Yu Wei Chen, Hong-Han Shuai, and Ching-Chun Huang, "Arbitrary-Resolution and Arbitrary-Scale Face Super-Resolution with Implicit Representation Networks," Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 4270-4279, 2024.  
doi: <https://doi.org/10.1109/WACV57701.2024.00422>
- [7] Zeyuan Chen, Yinbo Chen, Jingwen Liu, Xingqian Xu, Vidit Goel, Zhangyang Wang, Humphrey Shi, and Xiaolong Wang, "Videoinr: Learning video implicit neural representation for continuous space-time super-resolution," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2047-2057, 2022.  
doi: <https://doi.org/10.48550/arXiv.2206.04647>
- [8] Vincent Sitzmann, Julien N. P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein, "Implicit neural representations with periodic activation functions," Advances in neural information processing systems, Vol.33, pp. 7462-7473, 2020.  
doi: <https://doi.org/10.48550/arXiv.2006.09661>
- [9] Arthur Jacot, Franck Gabriel, and Clément Hongler, "Neural tangent kernel: Convergence and generalization in neural networks," Advances in neural information processing systems, Vol.31, pp. 8580-8589, 2018.  
doi: <https://doi.org/10.48550/arXiv.1806.07572>
- [10] Nobuyuki Otsu, "A threshold selection method from gray-level histograms," Automatica, Vol.11, pp. 23-27, 1975.  
doi: <https://doi.org/10.1109/TSMC.1979.4310076>
- [11] Ping-Sung Liao, Tse-Sheng Chen, and Pau-Choo Chung, "A fast algorithm for multilevel thresholding," Journal of Information Science and Engineering, Vol.17, No.5, pp. 713-727, 2001.  
doi: <https://doi.org/10.6688/JISE.2001.17.5.1>
- [12] Eastman Kodak, "Kodak Lossless True Color Image Suite (PhotoCD PCD0992)," 1993, Retrieved from <http://r0k.us/graphics/kodak/>
- [13] Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred A. Hamprecht, Yoshua Bengio, and Aaron Courville, "On the spectral bias of neural networks," International conference on machine learning, PMLR, pp. 5301-5310, May, 2019.  
doi: <https://doi.org/10.48550/arXiv.1806.08734>
- [14] Kun Su, Mingfei Chen, and Eli Shlizerman, "Inras: Implicit neural representation for audio scenes," Advances in Neural Information Processing Systems, Vol.35, pp. 8144-8158, 2022.  
<https://openreview.net/forum?id=7KBzV5IL7W>
- [15] Yinbo Chen, Sifei Liu, and Xiaolong Wang, "Learning continuous image representation with local implicit image function," Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 8628-8638, 2021.  
doi: <https://doi.org/10.1109/CVPR46437.2021.00852>
- [16] Ivan Anokhin, Kirill Demochkin, Taras Khakhulin, Gleb Sterkin, Victor Lempitsky, and Denis Korzhnikov, "Image generators with conditionally-independent pixel synthesis," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14278-14287, 2021.  
doi: <https://doi.org/10.1109/CVPR46437.2021.01405>
- [17] Vincent Sitzmann, Michael Zollhoefer, and Gordon Wetzstein, "Scene representation networks: Continuous 3d-structure-aware neural scene representations," Advances in Neural Information Processing Systems, Vol.32, 2019.  
doi: <https://doi.org/10.48550/arXiv.1906.01618>
- [18] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," Communications of the ACM, Vol.65, No.1, pp. 99-106, 2021.  
doi: <https://doi.org/10.1145/3503250>
- [19] Yitong Xia, Hao Tang, Radu Timofte, and Luc Van Gool, "Sinerf: Sinusoidal neural radiance fields for joint pose estimation and scene reconstruction," arXiv preprint, arXiv:2210.04553, 2022.  
doi: <https://doi.org/10.48550/arXiv.2210.04553>
- [20] Kenneth O. Stanley, "Compositional pattern producing networks: A novel abstraction of development," Genetic programming and evolvable machines, Vol.8, pp. 131-162, 2007.  
doi: <https://doi.org/10.1007/s10710-007-9028-8>
- [21] Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng, "Fourier features let networks learn high frequency functions in low dimensional domains," Advances in neural information processing systems, Vol.33, pp. 7537-7547, 2020.  
doi: <https://doi.org/10.48550/arXiv.2006.10739>
- [22] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove, "Deepsdf: Learning continuous signed distance functions for shape representation," Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 165-174, 2019.  
doi: <https://doi.org/10.1109/CVPR.2019.00025>
- [23] Rizal Fathony, Anit Kumar Sahu, Devin Willmott, and J. Zico Kolter, "Multiplicative filter networks," International Conference on Learning Representations, 2020.

<https://openreview.net/forum?id=OmtmcPkkhT>

- [24] David B. Lindell, Dave Van Veen, Jeong Joon Park, and Gordon Wetzstein, “Bacon: Band-limited coordinate networks for multiscale scene representation,” Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 16252-16262, 2022.  
doi: <https://doi.org/10.1109/CVPR52688.2022.01577>
- [25] Luc Vincent and Pierre Soille, “Watersheds in digital spaces: an efficient algorithm based on immersion simulations,” IEEE Transactions on Pattern Analysis & Machine Intelligence, Vol.13, No.06, pp. 583-598, 1991.  
doi: <https://doi.org/10.1109/34.87344>
- [26] Serge Belongie, Chad Carson, Hayit Greenspan, and Jitendra Malik, “Color-and texture-based image segmentation using EM and its application to content-based image retrieval,” Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271), pp. 675-682, IEEE, 1998.  
doi: <https://doi.org/10.1109/ICCV.1998.710790>
- [27] Richard Nock and Frank Nielsen, “Statistical region merging,” IEEE Transactions on pattern analysis and machine intelligence, Vol.26, No.11, pp. 1452-1458, 2004.  
doi: <https://doi.org/10.1109/TPAMI.2004.110>
- [28] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala, “Pytorch: An imperative style, high-performance deep learning library,” Advances in neural information processing systems, Vol.32, 2019.  
doi: <https://doi.org/10.48550/arXiv.1912.01703>
- [29] A. Harry Robinson and Colin Cherry, “Results of a Prototype Television Bandwidth Compression Scheme,” Proceedings of the IEEE, Vol.55, No.3, pp. 356 - 364, 1967.  
doi: <https://doi.org/10.1109/PROC.1967.5493>
- [30] David A. Huffman, “A method for the construction of minimum-redundancy codes,” Proceedings of the IRE, Vol.40, No.9, pp. 1098-1101, 1952.  
doi: <https://doi.org/10.1109/JRPROC.1952.273898>
- [31] Gregory K. Wallace, “The JPEG still picture compression standard,” Communications of the ACM, Vol.34, No.4, pp. 30-44, 1991.  
doi: <https://doi.org/10.1109/30.125072>
- [32] Gary J. Sullivan, Jens-Rainer Ohm, Woo-Jin Han, and Thomas Wiegand, “Overview of the high efficiency video coding (HEVC) standard,” IEEE Transactions on circuits and systems for video technology, Vol.22, No.12, pp. 1649-1668, 2012.  
doi: <https://doi.org/10.1109/TCSVT.2012.2221191>

---

## Introduction Authors

---



**Myeong-Jun Oh**

- Mar. 2018 ~ Feb. 2024 : Bachelor's degree in the Department of Computer Engineering at Kyung Hee University
- Mar. 2024 ~ Present : Master's degree in the Department of Artificial Intelligence at Kyung Hee University
- ORCID : <https://orcid.org/0009-0001-3945-1824>
- Research interests : Implicit Neural Representation, Inverse problems in image processing



**Sung-Ho Bae**

- Mar. 2004 ~ Feb. 2011 : Bachelor's degrees in the Department of Computer Engineering and Electronic Engineering (dual major) at Kyung Hee University
- Feb. 2011 ~ Aug. 2016 : Ph. D. in the Department of Electrical Engineering at KAIST
- Jul. 2016 ~ Aug. 2017 : Postdoc. Associate in MIT Computer Science and Artificial Intelligence Laboratory (CSAIL)
- Sep. 2017 ~ Present : Associate Professor at Kyung Hee University in school of Computing
- Research interests : Inverse problems in image processing, video compression, generative AI, and model compression