Poster ID: THU-AM-195

清華大學
Tsinghua University

TINC: Tree-structured Implicit Neural Compression

Runzhao Yang[1,2]

1 Department of Automation, Tsinghua University, Beijing 100084, China
2 yangrz20@mails.tsinghua.edu.cn

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

1 Introduction

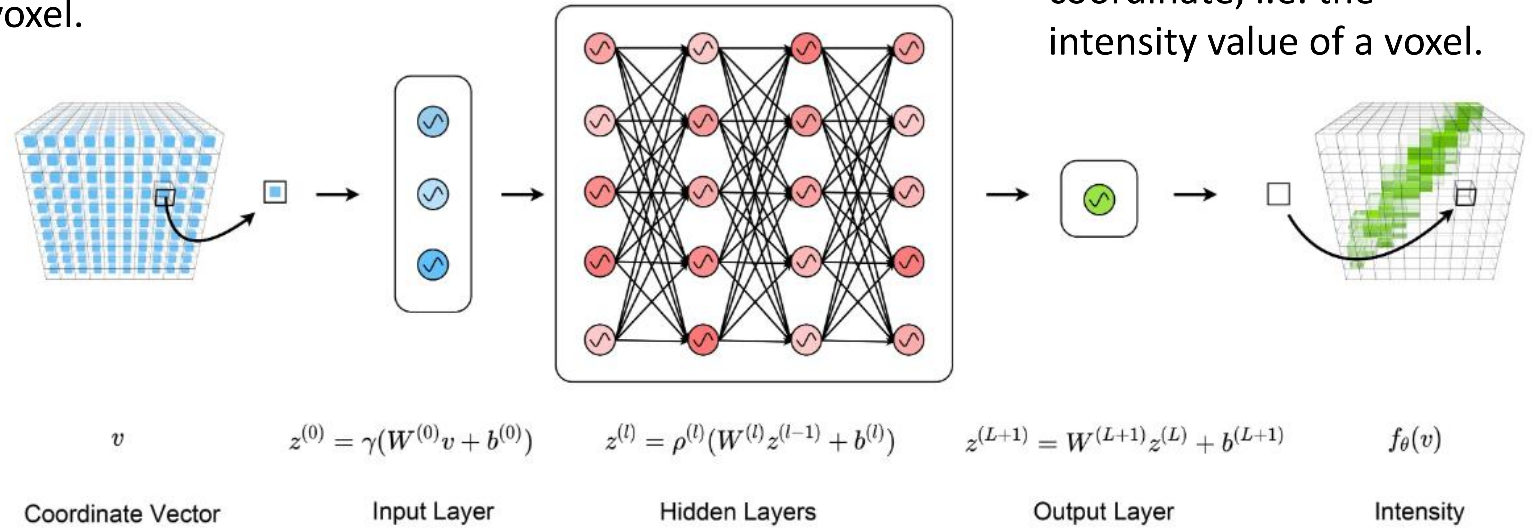
□Implicit Neural Representation, INR, is a promising compressor

- ➢ Treat the data as the result of sampling a continuous function.
- ➢ Use a neural network to parameterize the function to represent the data.

Input: Any coordinate in the imaging data coordinate system, i.e. the coordinate of a voxel.

Parameterized Neural Networks

Output: The value of the imaging data corresponding to this coordinate, i.e. the intensity value of a voxel.



characteristics

Traditional discrete grid representation

⬇

Continuous parameterization

- ✓ Not limited by grid resolution
- ✓ Simulation of details in the signal
- ✓ Modeling the higher order derivative information contained in the natural signal

1 Introduction

☐ INR is limited confronted with large sized data

- ➢ INR is intrinsically of limited spectrum coverage and cannot envelop the sp… of the target data.
- ➢ Two pioneering works using INR for data compression, including **NeRV** and **SCI** have attempted to handle this issue in their respective ways.

NeRV

- ➢ Introduces the convolution operation into INR.
- ✓ Reduces the required number of parameters using the weight sharing mechanism.
- X Convolution is spatially invariant and thus limits NeRV's representation accuracy on complex data with spatial varying feature distribution.

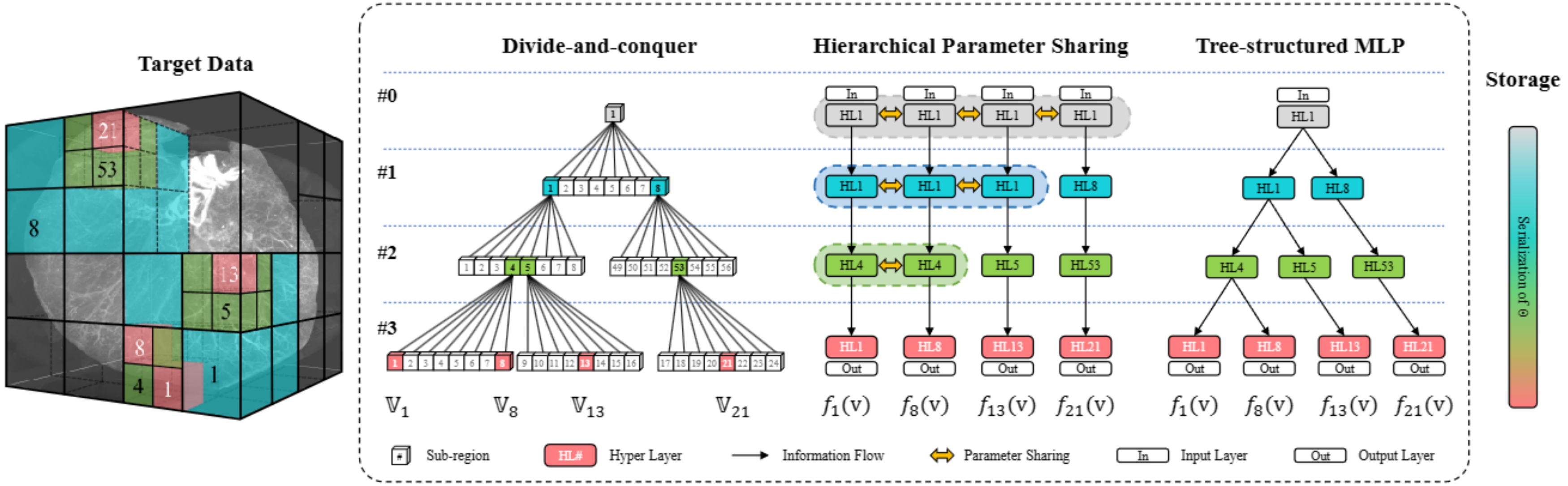
SCI

- ➢ Adopts divide-and-conquer strategy and partitions the data into blocks within INR's concentrated spectrum envelop.
- ✓ Improves the local fidelity.
- X Cannot remove non-local redundancies for higher compression ratio and tend to cause blocking artifacts.

1 Introduction

□We introduce TINC: Tree-structured Implicit Neural Compression

- ➢ We propose to build a tree-structured Multi-Layer Perceptrons (MLPs), w consists of a set of INRs to represent local regions in a compact manner and organizes them under a hierarchical architecture for parameter sharing and higher compression ratio.



1 Introduction

□TINC outperforms the SOTAs under high compression ratios

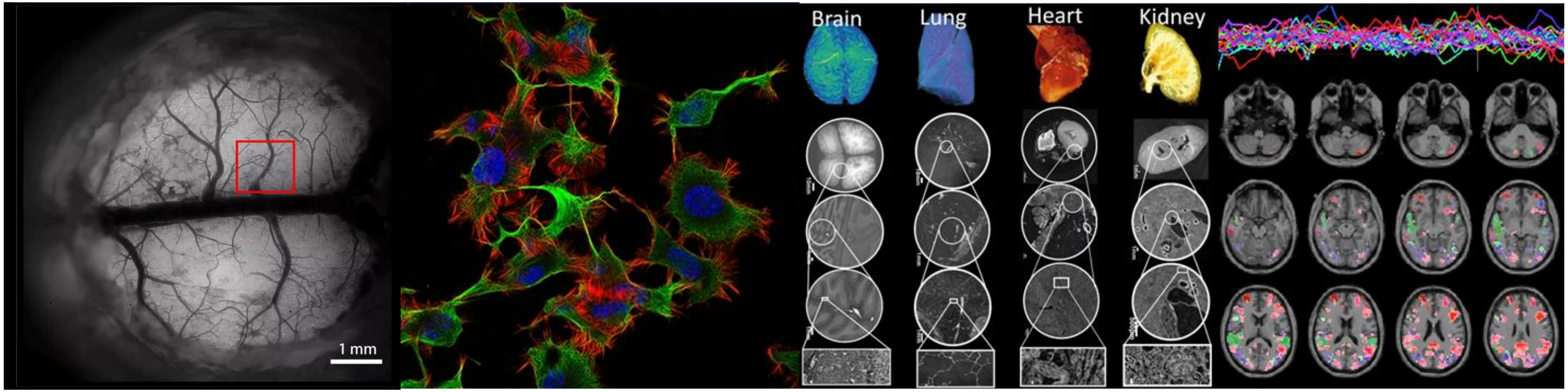
➢ Using the massive and diverse biomedical data, we conduct extensive experiments to validate that TINC greatly improves the capability of INR and even outperforms the commercial compression tools (H.264 and HEVC) under high compression ratios.

Method	Medical data				Biological data			
	PSNR All. (dB)	SSIM All.	PSNR High. (dB)	SSIM High.	Acc.200 All.	Acc.500 All.	Acc.200 High.	Acc.500 High.
TINC (ours)	•52.02	•0.9897	#50.59	#0.9878	•0.9945	•0.9970	#0.9934	#0.9958
JPEG	41.41	0.9722	30.49	0.9374	0.6612	0.9834	0.0197	0.9882
H.264	51.18	0.9896	47.28	0.9860	0.9919	0.9959	0.9860	0.9926
HEVC	#52.31	#0.9903	•50.51	•0.9877	#0.9955	#0.9975	0.9917	0.9930
SCI	51.90	0.9894	50.39	0.9876	0.9943	0.9965	•0.9921	•0.9951
NeRF	50.93	0.9875	49.66	0.9863	0.9935	0.9962	0.9903	0.9940
NeRV	47.11	0.9859	40.11	0.9800	0.9815	0.9901	0.9732	0.9867
DVC	47.39	0.9865	45.74	0.9840	0.9827	0.9900	0.9692	0.9789
SGA+BB	46.56	0.9836	43.02	0.9808	0.8038	0.9883	0.4817	0.9798
SSF	46.25	0.9807	43.70	0.9773	0.7221	0.9603	0.7790	0.9542

1 Introduction

☐Biomedical Imaging

- ➢ Visualization of organisms at different scales of cells, tissues and organs us[...] various imaging techniques.



1 Introduction

Tsinghua University

JUNE 18-22, 20

□Biomedical imaging data characteristics, needs and challenges of compre

characteristics

- **High sampling rate:** for capturing minute structural details, providing higher spatial resolution.
- **High imaging speed:** for capturing rapid dynamic changes, providing higher temporal resolution.
- **High dimensionality:** for representing information including spatial location, time series, etc.
- **Large volume:** terabytes or even petabytes of data.

needs

- **Storage:** reduce storage costs, and avoid experimental data loss
- **Transmission:** reduce transmission costs, promote experimental data sharing.
- **Analysis:** reduce I/O pressure, reduce storage space during analysis, improve analysis efficiency, and accelerate scientific discovery.

challenges of compressors

How to design high compression rate biomedical imaging data compressor, for efficient storage, transmission and analysis of biomedical data?

1 Introduction

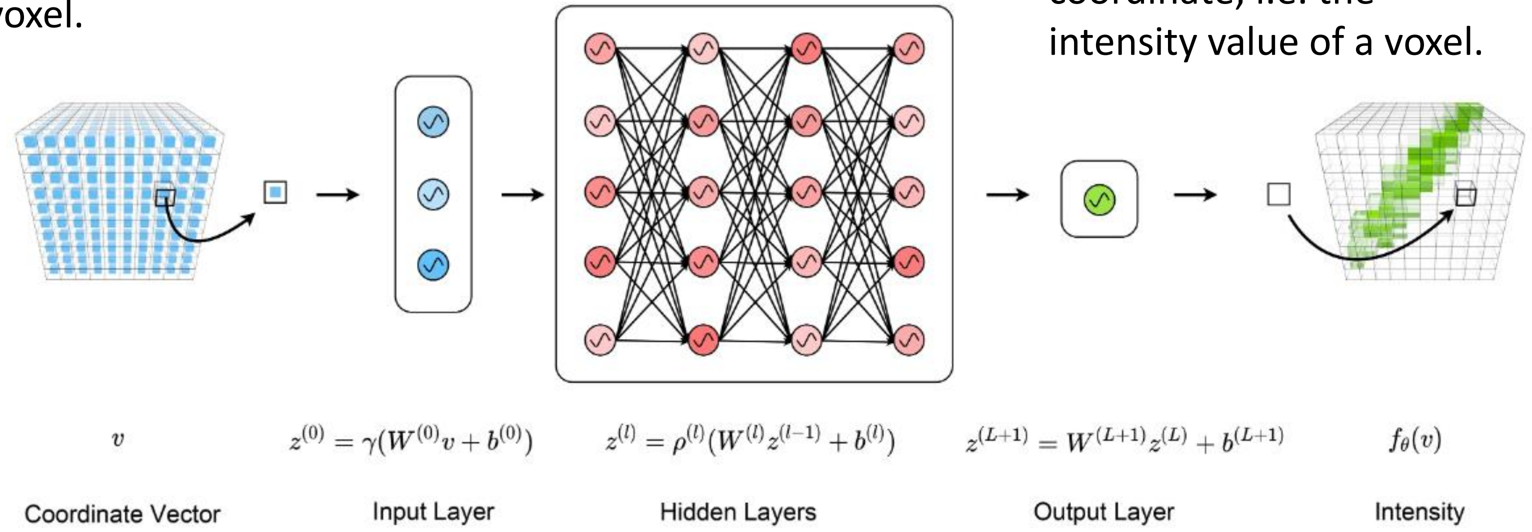
☐Implicit Neural Representation, INR, is a promising compressor

- ➢ Treat the data as the result of sampling a continuous function.
- ➢ Use a neural network to parameterize the function to represent the data.

Input: Any coordinate in the imaging data coordinate system, i.e. the coordinate of a voxel.

Parameterized Neural Networks

Output: The value of the imaging data corresponding to this coordinate, i.e. the intensity value of a voxel.



characteristics

Traditional discrete grid representation

⬇

Continuous parameterization

- ✓ Not limited by grid resolution
- ✓ Simulation of details in the signal
- ✓ Modeling the higher order derivative information contained in the natural signal

1 Introduction

☐ INR is limited confronted with large sized data

- ➢ INR is intrinsically of limited spectrum coverage and cannot envelop the sp… of the target data.
- ➢ Two pioneering works using INR for data compression, including **NeRV** and **SCI** have attempted to handle this issue in their respective ways.

NeRV

- ➢ Introduces the convolution operation into INR.
- ✓ Reduces the required number of parameters using the weight sharing mechanism.
- X Convolution is spatially invariant and thus limits NeRV's representation accuracy on complex data with spatial varying feature distribution.

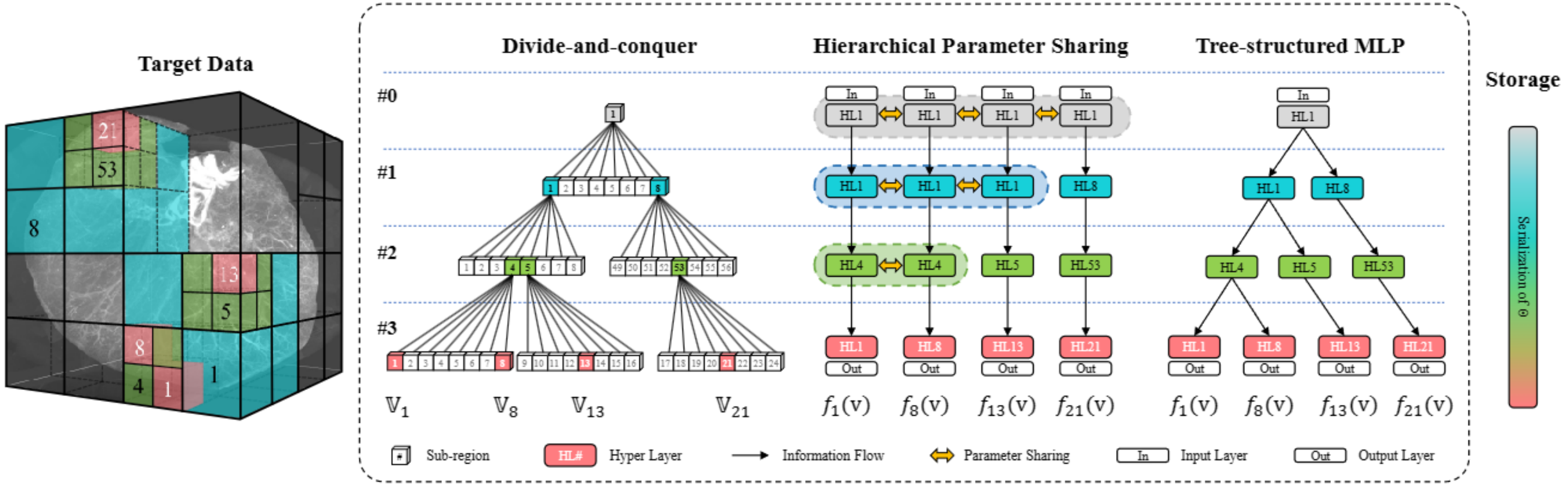
SCI

- ➢ Adopts divide-and-conquer strategy and partitions the data into blocks within INR's concentrated spectrum envelop.
- ✓ Improves the local fidelity.
- X Cannot remove non-local redundancies for higher compression ratio and tend to cause blocking artifacts.

1 Introduction

We introduce TINC: Tree-structured Implicit Neural Compression

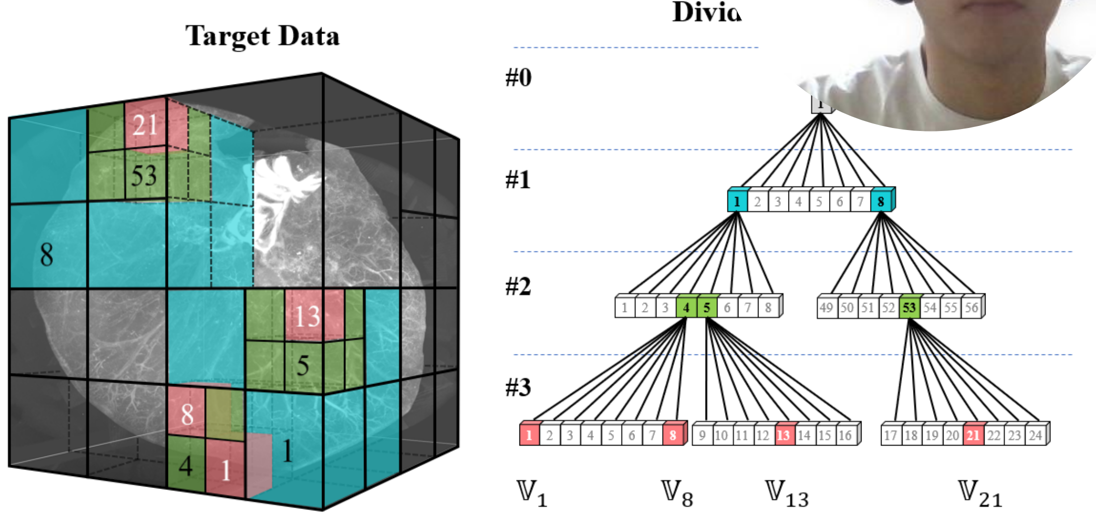
- We propose to build a tree-structured Multi-Layer Perceptrons (MLPs), w… consists of a set of INRs to represent local regions in a compact manner and organizes them under a hierarchical architecture for parameter sharing and higher compression ratio.



2 Method

☐Ensemble of Implicit Neural Compressors

- We borrow the idea of ensemble learning to partition the target volume into blocks and use multiple less expressive $f_k(\cdot, \boldsymbol{\Theta}_k)$ to achieve a powerful representation.
- We adopt the divide-and-conquer strategy to ensemble all implicit functions that represents data at its corresponding coordinate region.



$$f(\mathbf{v}; \boldsymbol{\Theta})$$

$$\Downarrow$$

$$f_k(\cdot, \boldsymbol{\Theta}_k)$$

$$\{f_k, k = 1, \cdots, K\}$$

optimization problem

$$\min_{\boldsymbol{\Theta}} \int_{\mathbf{v}\in\mathbb{V}} \mathcal{L}(f(\mathbf{v}; \boldsymbol{\Theta}), \mathbf{d}(\mathbf{v}))$$

$$\begin{cases} f(\mathbf{v}; \boldsymbol{\Theta}) := \sum_{k=1}^{K} \mathbf{1}_{\mathbb{V}_k}(\mathbf{v}, f_k(\mathbf{v}, \boldsymbol{\Theta}_k)), \\ \mathbf{1}_{\mathbb{V}_k}(\mathbf{v}, x) = \begin{cases} x, \mathbf{v} \in \mathbb{V}_k \\ 0, else \end{cases} \\ |\boldsymbol{\Theta}| = \sum_{k=1}^{K} |\boldsymbol{\Theta}_k| \end{cases}$$

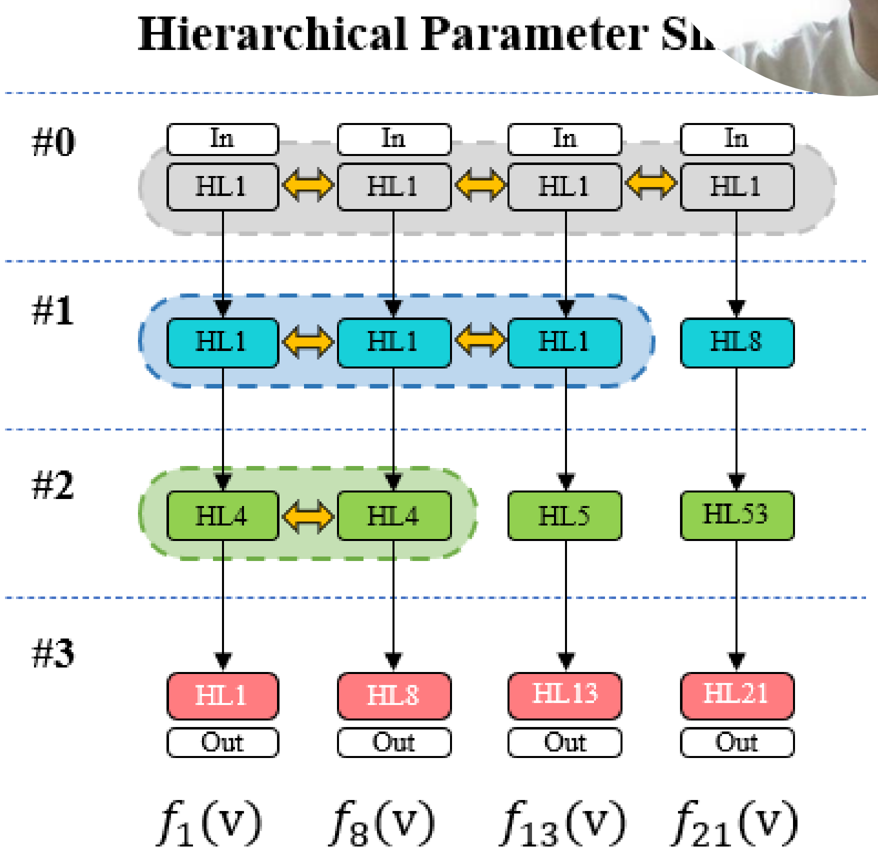
2 Method

Tsinghua University

JUNE 18-22, 20

◻Hierarchical Parameter Sharing Mechanism

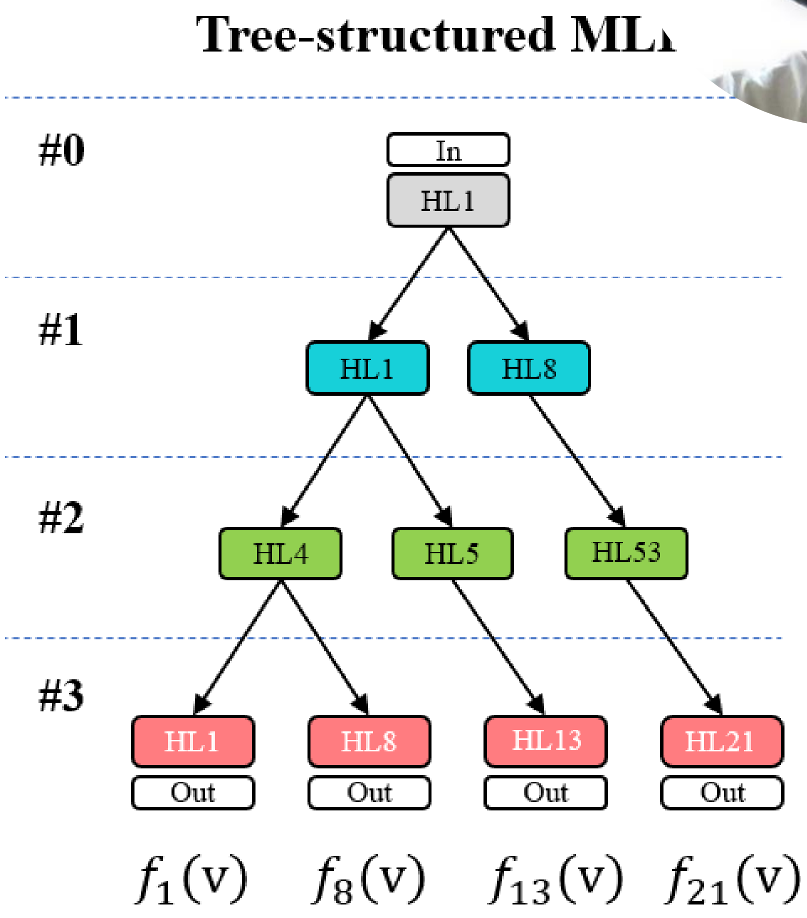
- We let these $\{f_k\}$ share their neural network parameters hierarchically with each other according to the spatial distance between corresponding regions.
- for a leaf node at level l, its corresponding MLP-implemented hidden layers can be divided into l segments, i.e. $f_k = f_k^{out} \circ f_k^{l} \circ f_k^{l-1} \circ \cdots \circ f_k^{1} \circ f_k^{in}$
- The sharing mechanism is defined on the octree structure. For example, if f_i and f_j share the same ancestor nodes at 1~3 levels, three pairs of hidden layer segments $(f_i^1, f_j^1), (f_i^2, f_j^2), (f_i^3, f_j^3)$ will share the same parameters.



2 Method

☐Tree-structured Network Architecture

- ➢ We propose a tree-structured MLP based on the L level octree partitioning.
- ➢ Each node contains a hyper layer consisting of some fully connected layers and takes the output of its parent node's hyper layer as input.
- ➢ Root node and leaf nodes additionally contain the input and output layers respectively.
- ✓ The output information of the leaf node is processed by the hyper layers in its ancestor nodes.
- ✓ At the same level, all sibling nodes share the same parent node and thus take the same information as input.



3 Experiments

Tsinghua University

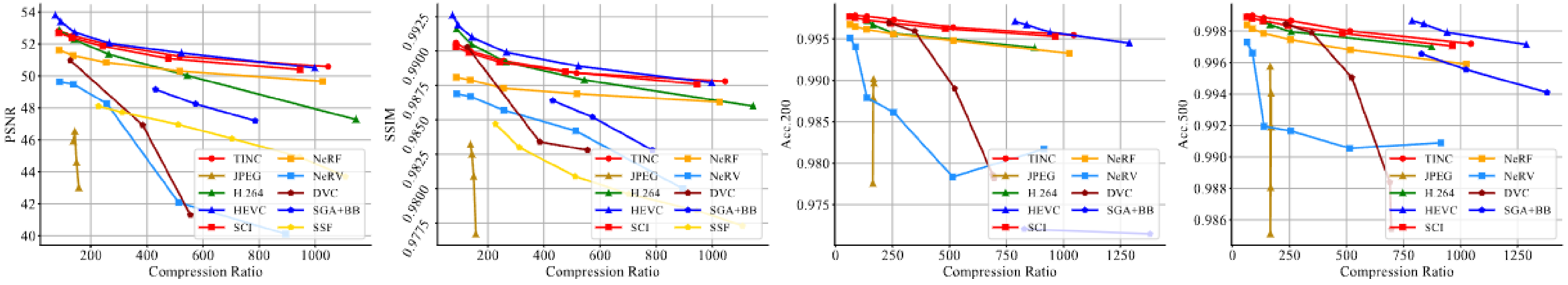
JUNE 18-22, 20

☐Performance Comparison with SOTAs

Method	Medical data				Biological data			
	PSNR All (dB)	SSIM All	PSNR High (dB)	SSIM High	Acc 200 All	Acc 500 All	Acc 200 High	Acc 500 High
TINC (ours)	•52.02	•0.9897	#50.59	#0.9878	•0.9945	•0.9970	#0.9934	#0.9958
JPEG	41.41	0.9722	30.49	0.9374	0.6612	0.9834	0.0197	0.9882
H.264	51.18	0.9896	47.28	0.9860	0.9919	0.9959	0.9860	0.9926
HEVC	#52.31	#0.9903	•50.51	•0.9877	#0.9955	#0.9975	0.9917	0.9930
SCI	51.90	0.9894	50.39	0.9876	0.9943	0.9965	•0.9921	•0.9951
NeRF	50.93	0.9875	49.66	0.9863	0.9935	0.9962	0.9903	0.9940
NeRV	47.11	0.9859	40.11	0.9800	0.9815	0.9901	0.9732	0.9867
DVC	47.39	0.9865	45.74	0.9840	0.9827	0.9900	0.9692	0.9789
SGA+BB	46.56	0.9836	43.02	0.9808	0.8038	0.9883	0.4817	0.9798
SSF	46.25	0.9807	43.70	0.9773	0.7221	0.9603	0.7790	0.9542

\# Best

● Second best

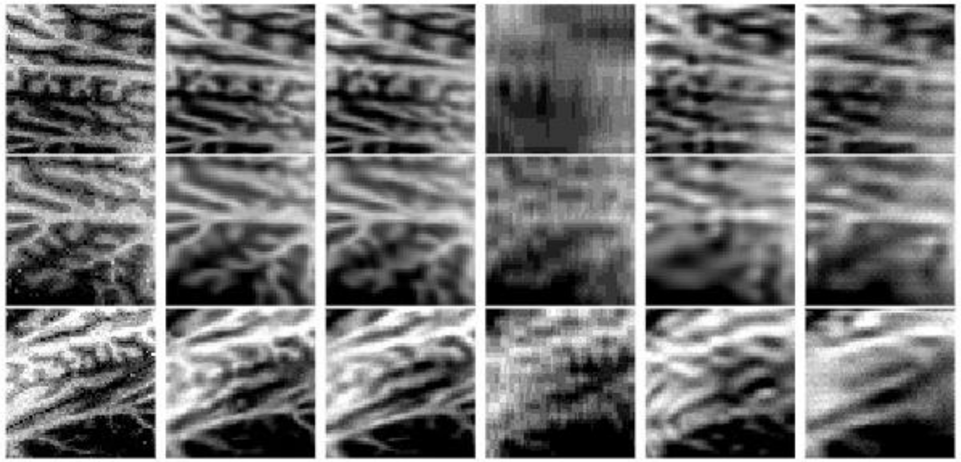


(a) (b) (c) (d)

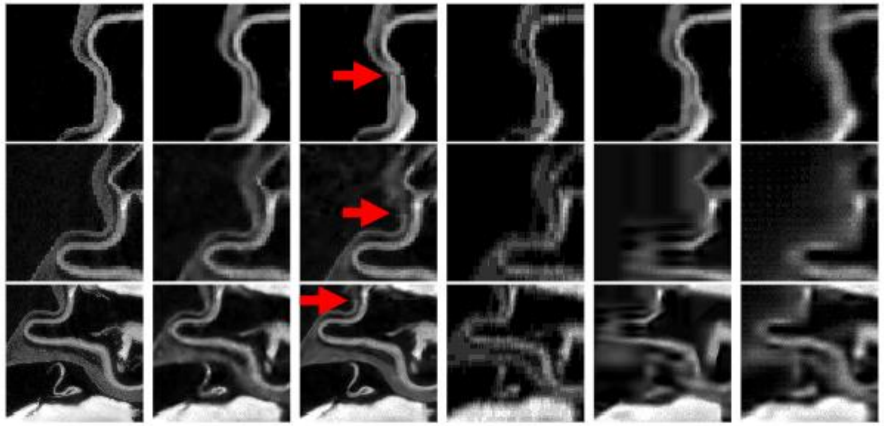
TINC outperforms the SOTAs under high compression ratios

3 Experiments

☐Performance Comparison with SOTAs



(a) Three ROIs from *Brain* data; compression ratio: ~87×



(b) Three ROIs from *Heart* data; compression ratio: ~87×

Ground Truth TINC (ours) SCI NeRV HEVC SGA+BB

- ✓ Outperforms the SOTAs
- ✓ Avoids blocking artifacts at the boundary

3 Experiments

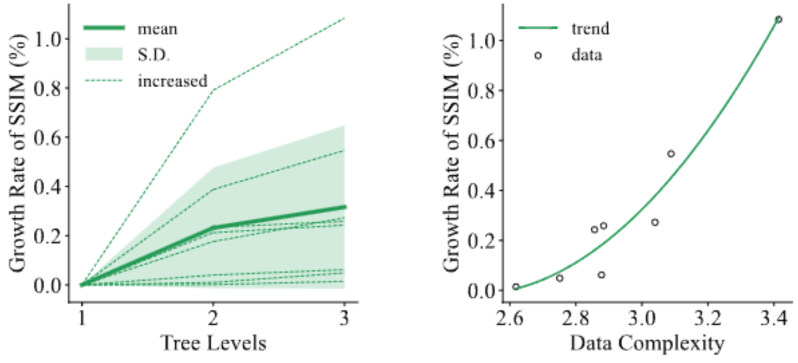
Tsinghua University

JUNE 18-22, 20

CV

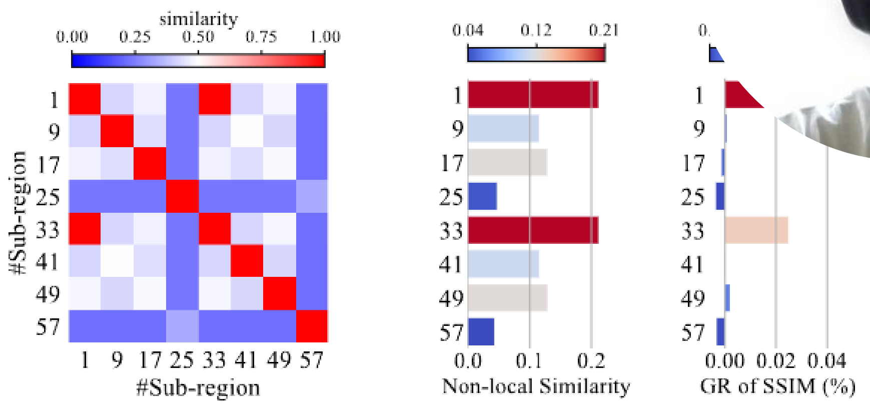
Flexibility Settings for Different Data

- We also analyze TINC's flexibility to different cases via experimentally studying the effect of three key settings:
  1. number of tree levels
  2. intra-level parameter allocation
  3. inter-level parameter allocation



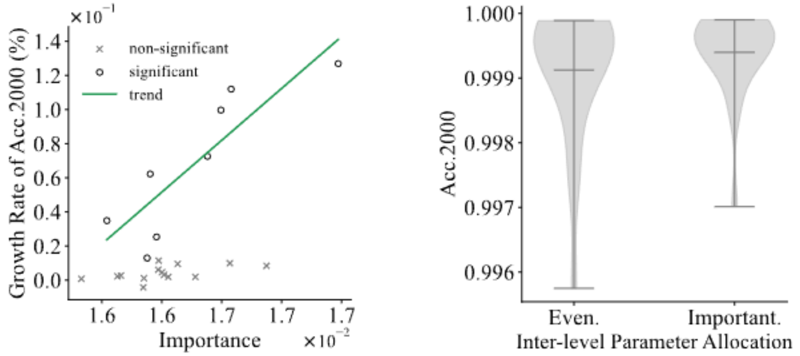
(a) (b)

✓ Setting of Tree Levels L



(a) (b) (c)

✓ Inter-level Parameter Allocation



(a) (b)

✓ Intra-level Parameter Allocation

4 Conclusion

☐Limitations and Future Extensions

- Similar to all current INR based compression methods, TINC is of high decompression speed but slow in compression, since it takes time to pursue the MLPs matching the target data.
- We plan to combine meta-learning to find the best initialization parameters for each organ to speed up TINC.