Methods of Algebraic Topology in the Study of Neural Network Architecture

Zhiwang Yu

Department of Mathematics
Southern University of Science and Technology
Advisor: Fuquan Fang and Yifei Zhu

May 10, 2025

Content

1. Background
2. Preliminaries
3. Main Results
4. Future Works
5. References

Evolution of Neural Networks

Historical Breakthroughs

- 1968: Hubel & Wiesel's visual cortex studies (biological basis for CNNs) [8]
- 1980: Fukushima's neocognitron (first CNN prototype with simple/complex cells) [6]
- 1998: LeCun's LeNet-5 (first successful application for digit recognition) [10]
- 2012: AlexNet revolution (deep learning breakthrough using GPUs) [9]
- 2015: ResNet innovations (enabled training of 100+ layer networks) [7]

Core CNN Components

- Convolutional layers (learn local features through filters) [14]
- Pooling layers (reduce spatial dimensions) [17]
- Skip connections (enable gradient flow in deep networks) [18]

The Emergence of Topological Data Analysis (TDA)

From Algebraic Topology to Data Science

- Traditional tools: Homology groups & Betti numbers (measure connectivity)
- Early applications: Theoretical studies using simplicial complexes (2000s)

Birth of Persistent Homology

- Edelsbrunner et al. introduced multi-scale analysis framework [5]
- Zomorodian & Carlsson improved noise stability [19]
- Became mainstream after Carlsson's 2009 review [2]

Algorithmic Breakthroughs

- Chazal: Stability guarantees [4]
- Oudot: Efficient computation [13]
- Software: Gudhi/Ripser libraries
- Visualization: Mapper algorithm

Construction of Topological Neural Networks

Discovery of Image Features

- De Silva & Carlsson found image patches form Klein bottle structure [16, 3]

Neural Network Verification

- Gabrielsson & Carlsson: CNN kernels preserve Klein bottle topology

Topological CNN Development

- Love et al. created:
  - Circular convolution kernels
  - Klein bottle convolution kernels
- Outperformed traditional CNNs on MNIST [12]

Phonemes

- We are highly motivated to apply the work of Love et al. to speech signals. Since convolutional neural networks primarily operate at the word level at most, we chose to focus on phonemes as the minimal linguistic unit.
- For phonemes, we provide only a brief introduction. Broadly, they can be categorized into two types: *voiced* (exhibiting clear periodicity) and *voiceless* (resembling white noise).
- Our core experiments involve phoneme classification, primarily using datasets such as SpeechBox, LJSpeech, and TIMIT. We extract all phoneme-level data from these datasets (excluding stressed segments).

Topological Audio Processing

Key Methods

- **Speech Processing:**
  - MFCCs + Persistent Homology (Brown and Knudson[1])
  - Preserves harmonic structures in noise
  - Enhances CNN spectral analysis
- **Music Signals:**
  - Topological persistence + CNNs (Liu et al.[11])
  - Effective feature extraction for classification

Core Advantages

- Reveals structural patterns via persistent homology (Robinson[15])
- Bridges topology with signal dynamics
- Robust to noisy environments

Date below

Topological Neural Networks for Speech

Speech Spectrograms

- Generated via STFT/MFCC (similar to grayscale images)
- Example visualizations will follow

Key Differences from Images

- Axes have physical meaning:
  - Horizontal: Time domain
  - Vertical: Frequency domain
- Diagonal directions lack semantic meaning

Research Motivation

- Build speech-adapted topological neural networks
- Leverage axis-specific information

Waveform and Spectrogram

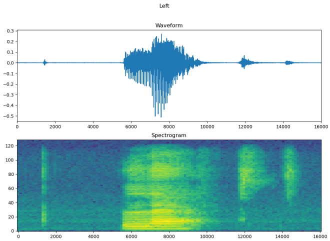


Figure 1 Waveform and Spectrogram

- Shows the word "left":
  - Waveform: Amplitude vs Time
  - Spectrogram: Frequency vs Time (color = energy)
- Each spectrogram column represents:
  - Short-time Fourier Transform (STFT) or
  - Mel-Frequency Cepstral Coefficients (MFCC)

Wave

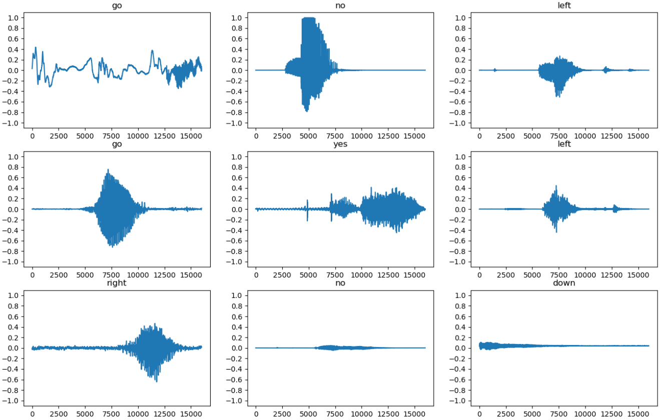


Figure 2 Wave

Spectrogram

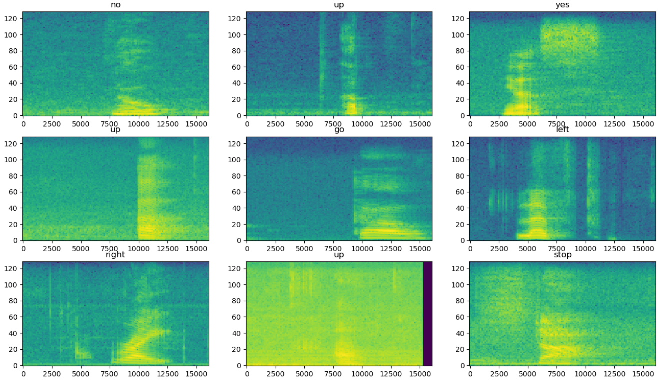


Figure 3 Spectrogram

Convolutional Neural Networks and Operations

$$\begin{bmatrix}1 & 2 & 3\\4 & 5 & 6\\7 & 8 & 9\end{bmatrix} \xrightarrow{\begin{bmatrix}1 & -1\\1 & -1\end{bmatrix}} \begin{bmatrix}-2 & -2\\-2 & -2\end{bmatrix}$$

- Example shows a 3×3 image convolved with 2×2 kernel producing 2×2 output
- Kernel operation:
  - Slides top-to-bottom, left-to-right
  - Element-wise multiplication then summation
  - Output size smaller than input
- CNNs excel at local feature extraction using:
  - Multiple parallel kernels
  - Shared weight architecture

The Structure of CF and KF

The convolutional kernels in CF have the following form

$$\begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}$$

up to a rotation. The convolutional kernels in KF add the following matrix

$$\begin{bmatrix} 1 & 1 & 1 \\ -2 & -2 & -2 \\ 1 & 1 & 1 \end{bmatrix}$$

with some linear relations.

Theoretical Framework

Theorem 1: Moduli Space of Convolution Kernels

For normalized 3×3 kernels (centered and normalized), the moduli space M is homeomorphic to S^5 (excluding zero matrices after centering).

Definition 1: Contrast

The **contrast** of a convolution kernel $\boldsymbol{A} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3] \in M$ is defined by

$$\mathrm{con}(\boldsymbol{A}) = \sqrt{||\boldsymbol{v}_1 - \boldsymbol{v}_2||^2 + ||\boldsymbol{v}_2 - \boldsymbol{v}_3||^2}.$$

Group Action $\mathrm{SO}(3)$

- For the moduli space M, there exists a natural left multiplication action by $\mathrm{SO}(3)$ which preserves contrast and maintains the norm of every vector.

Theorem 2: Orbit Space $M/\mathrm{SO}(3)$

Considering the left multiplication action of $\mathrm{SO}(3)$, the quotient space $M/\mathrm{SO}(3)$ is homeomorphic to the disk D^2.

- The left multiplication action of $\mathrm{SO}(3)$ is considered here because each column of the speech signal represents a coherent unit (short-time information). Thus, we use the left group action to traverse the entire space.

Kernel Construction

- Decompose kernel space into:
  - D^2 (elliptical disk component)
  - $\mathrm{SO}(3)$ (rotation group component)
- Sampling strategy:
  - Random selection from each component
  - Uniform and well-distributed initialization
- Special cases:
  - Boundary cases
  - $x = y$ cases (degenerate $\mathrm{SO}(3)$ action)
- Current implementation does not handle these special cases

Confusion matrix

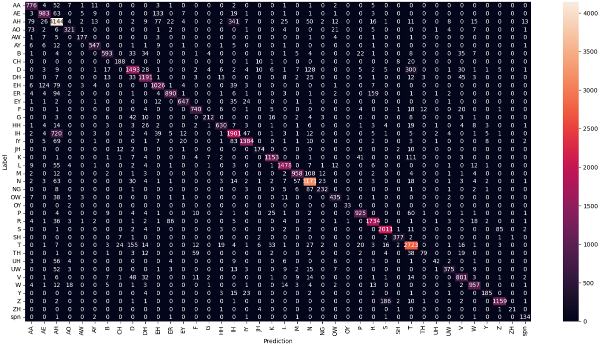


Figure 4 Confusion Matrix

Experimental Results

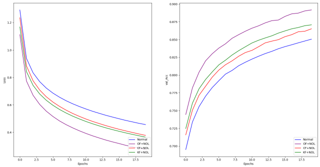


Figure 5 Comparisions of Loss and Accuracy on LJSpeech

- Dataset: LJSpeech
- Methodology:
  - Segmented raw audio into phonemes
  - **Selected 500 samples per category for training**
- Results comparison:
  - Baseline: Normal NN
  - Prior work: CF+NOL, KF+NOL (Love et al.)
  - Our method: OF+NOL (optimal performance)
- Metrics: Loss and Accuracy vs Epochs

Experimental Results

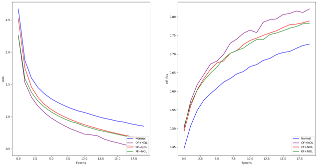


Figure 6 Comparisions of Loss and Accuracy on LJSpeech(New)

- Dataset: LJSpeech
- Methodology:
  - Segmented raw audio into phonemes
  - **Training directly**
- Results comparison:
  - Baseline: Normal NN
  - Prior work: CF+NOL, KF+NOL (Love et al.)
  - Our method: OF+NOL (optimal performance)
- Metrics: Loss and Accuracy vs Epochs

SNR=0

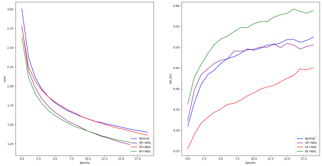


Figure 7 Comparisions of Loss and Accuracy on LJSpeech (SNR=0)

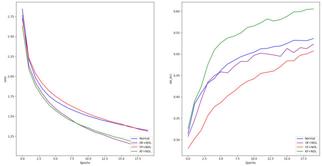


Figure 8 Comparisions of Loss and Accuracy on LJSpeech (New, SNR=0)

Results on Words and Images

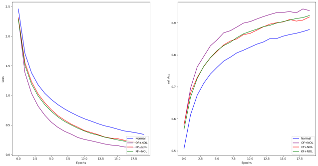


Figure 9 Comparisions of Loss and Accuracy on SpeechCommands

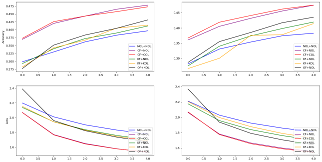


Figure 10 Comparisions of Loss and Accuracy on CIFAR10

Sketch of Proof

Proof of Theorem 1:

- Centralization: Column vectors sum to 0
- Normalization: Quotient by Frobenius norm $\Rightarrow M \cong S^5$

Proof of Theorem 2:

- Define map $\phi : M \to \mathbb{R}^3$ by $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3) \mapsto (x, y, z)$, where

$$x = \|\boldsymbol{v}_1\|^2, \quad y = \|\boldsymbol{v}_3\|^2,$$
$$z = \langle \boldsymbol{v}_1, \boldsymbol{v}_3 \rangle$$

with the following key relations:

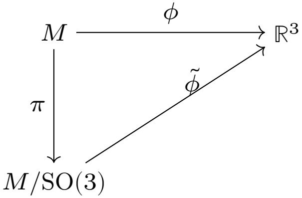
$$x + y + z = \frac{1}{2} \text{ (matrix norm)}$$
$$z^2 \leq xy \text{ (inner product)}$$

- $\mathrm{SO}(3)$ invariance: Coordinates well-defined on $M/\mathrm{SO}(3)$

(To be continued...)

Sketch of Proof (Continued)

(2) We define $\tilde{\phi} : M/\mathrm{SO}(3) \to \mathbb{R}^3$ with commutative diagram:



Eliminating z yields elliptical disk in (x, y): $(3x + 3y - 2)^2 + 3(x - y)^2 \leq 1$
This represents $\phi(M) = \tilde{\phi}(M/\mathrm{SO}(3))$. To prove $\tilde{\phi}$ is injective:

- $M/\mathrm{SO}(3)$ is compact
- $\mathbb{R}^3$ is Hausdorff
- Equivalent to transitivity of kernel configurations under group action
- Any two vectors with given norms and angle can be rotated to match

Thus $\tilde{\phi}$ is injective. □

Short-term Goals

- Further clarify the necessity and advantages of the $\mathrm{SO}(3)$ action.
- Explore convolution kernels with more diverse shapes.
- Conduct in-depth research on model adaptation to enhance its robustness.

Long-term Goals

- At the phoneme level, capture local features as comprehensively as possible, and subsequently establish a topological structure (though Transformers may be more suitable for this task).
- When addressing different practical problems, emphasize multimodality and the integration of topological and geometric methods.

References I

[1] Kenneth A Brown and Kevin P Knudson. "Nonlinear statistics of human speech data". In: *International Journal of Bifurcation and Chaos* 19.07 (2009), pp. 2307–2319.

[2] Gunnar Carlsson. "Topology and data". In: *Bulletin of the American Mathematical Society* 46.2 (2009), pp. 255–308.

[3] Gunnar Carlsson et al. "On the local behavior of spaces of natural images". In: *International journal of computer vision* 76 (2008), pp. 1–12.

[4] Frédéric Chazal et al. *The Structure and Stability of Persistence Modules*. 2012. ISBN: 978-3-319-42543-6.

[5] Edelsbrunner, Letscher, and Zomorodian. "Topological persistence and simplification". In: *Discrete & computational geometry* 28 (2002), pp. 511–533.

[6] Kunihiko Fukushima. "Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position". In: *Biological Cybernetics* 36.4 (1980), pp. 193–202. ISSN: 0340-1200, 1432-0770.

[7] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[8] DH Hubel and TN Wiesel. "Receptive Fields and Functional Architecture of Monkey Striate Cortex". In: *The Journal of physiology* 195.1 (1968), pp. 215–243. ISSN: 0022-3751.

[9] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Communications of the ACM* 60.6 (2017), pp. 84–90. ISSN: 0001-0782, 1557-7317.

[10] Yann LeCun et al. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324.

References II

[11] Jen-Yu Liu, Shyh-Kang Jeng, and Yi-Hsuan Yang. "Applying topological persistence in convolutional neural network for music audio signals". In: *arXiv preprint arXiv:1608.07373* (2016).

[12] Ephy R Love et al. "Topological convolutional layers for deep learning". In: *Journal of Machine Learning Research* 24.59 (2023), pp. 1–35.

[13] Steve Y Oudot. *Persistence theory: from quiver representations to data analysis*. Vol. 209. American Mathematical Society Providence, 2015.

[14] Talha Qaiser et al. "Fast and accurate tumor segmentation of histology images using persistent homology and deep convolutional features". In: *Medical image analysis* 55 (2019), pp. 1–14.

[15] Michael Robinson. *Topological signal processing*. Vol. 81. Springer, 2014.

[16] Vin de Silva and Gunnar Carlsson. "Topological estimation using witness complexes". In: *SPBG'04 Symposium on Point - Based Graphics 2004*. Ed. by Markus Gross et al. The Eurographics Association, 2004. ISBN: 3-905673-09-6.

[17] F. Sultana, A. Sufian, and P. Dutta. "A Review of Object Detection Models Based on Convolutional Neural Network". In: *Intelligent Computing: Image Processing Based Applications*. Ed. by J. K. Mandal and Soumen Banerjee. Singapore: Springer Singapore, 2020, pp. 1–16. ISBN: 978-981-15-4288-6.

[18] Nida M. Zaitoun and Musbah J. Aqel. "Survey on Image Segmentation Techniques". In: *Procedia Computer Science* 65 (2015), pp. 797–806. ISSN: 18770509.

[19] Afra Zomorodian and Gunnar Carlsson. "Computing Persistent Homology". In: *Discrete Comput. Geom* 33 (2005), pp. 249–274.

Thanks for your listening!