Topological Data Analysis and Topological Deep Learning: with Applications to Image Data

Haiyu Zhang

Department of Mathematics
Southern University of Science and Technology

May 23, 2024

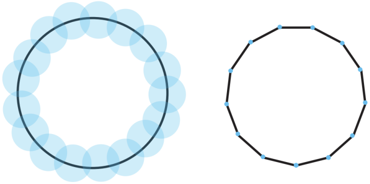
Data have shape
shape has meaning
meaning brings value.

Outline

- Topological Data Analysis:Topological Features Discovery
  Theoretical Foundation and Techniques of TDA
  Local Behavior of Feature Space in Natural Images
- Topological Deep Learning:Topological Features Embedding
  Topological Features of CNN Trained for Image Classification
  Topological Convolutional Neural Network

Theoretical Foundation and Techniques of TDA

- Filtered Simplicial Complex and Persistent Homology
  A filtration of simplicial complexes Σ_ϵ is given by the sequence $\emptyset \subseteq \Sigma_{\epsilon_1} \subseteq \cdots \subseteq \Sigma_{\epsilon_n}$ with $\epsilon_1 < \epsilon_2 \ldots < \epsilon_n$.
  We can construct a persistent chain complex $C_*(\Sigma_*)$ by setting $C_*^i(\Sigma_*) = C_*(\Sigma_{\epsilon_i})$ and the chain map $x : C_*^i(\Sigma_*) \hookrightarrow C_*^{i+1}(\Sigma_*)$.
  For $i < j$, the (i,j)-persistent homology of C, denoted $H_*^{i\to j}(\mathrm{C})$, is defined to be the image of the induced homomorphism $x_* : H_*(C_*^i(\Sigma_*)) \to H_*(C_*^j(\Sigma_*))$.



Local Behavior of Feature Space in Natural Image

We consider the 3 by 3 high contrast local patches as features of natural images.

How to obtain the feature space of a natural image?

- Extract 3 by 3 local patches from the image and regard them as 9-dimensional vectors.
- Apply Discrete Cosine Transform (DCT) to obtain patches with high contrast.
- Apply KNN Density Filtration to get the core set $X(k, p)$ with high density.

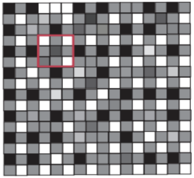


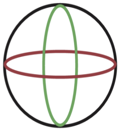
Figure: Extract 3 by 3 local patches from a image.

Local Behavior of Feature Space in Natural Image

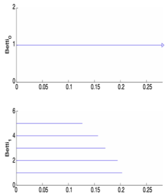
Apply TDA on the feature space, we obtain......

- Three Circle Model: $X(15, 30)$ [Gabrielsson and Carlsson, 2008]

a) Point cloud of $X(15, 30)$



b) Abstract model



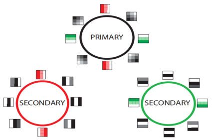


Figure: Three Circle Model

Local Behavior of Feature Space in Natural Image

Is there a 2-dim surface containing these three circle?

- Klein bottle [Gabrielsson and Carlsson, 2008]

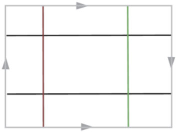


Figure: Three model embeds naturally in Klein bottle.

Interpretation: As the density estimation parameter (k) decreases, the space of high contrast local patches with high density gradually fills out a 2-dimensional manifold: firstly the primary circle of **linear gradients** and then two secondary circles of **quadratic gradients in horizontal and vertical directions** (two preferences of natural images). Finally, it fills all inter-midiate directions except **the purely quadratic gradients in non-horizontal and non-vertical directions**.

Theoretical Proof of the Existence of Klein Bottle: Polynomial Space [Carlsson, 2009]

Regard the 3 by 3 patches as the result vectors obtained by applying a smooth real-valued function of two variables on nine grid points of the unit disc. Now we consider a subspace of the space of smooth real-valued function P_0 , the set of all functions in the form

$$f(x, y) = q(\lambda x + \mu y)$$

where $\lambda^2 + \mu^2 = 1$ and q is a quadratic function with single variable. Let A be the space of single variable quadratic function $q(t) = c_0 + c_1 t + c_2 t^2$ with two restrictions

$$\int_{-1}^{1} q(t)\, dt = 0 \; \& \int_{-1}^{1} q(t)^2\, dt = 1$$

which indicates that A is an ellipse which is homeomorphic to a circle.

Theoretical Proof of the Existence of Klein Bottle

For any $q \in A$ and any unit vector $\vec{v}$ in $\mathbb{R}^2$, define $q_{\vec{v}} : \mathbb{R}^2 \to \mathbb{R}$ by $q_{\vec{v}}$ $(\vec{\omega}) = q(\vec{v} \cdot \vec{\omega})$ where $\vec{\omega}$ is a 2-dimensional variable. We have that

$$\int_D {q_{\vec{v}}}^2 \neq 0 \,\&\, \int_D q_{\vec{v}} = 0$$

Define a continuous map ϕ from $A \times S^1$ to P_0 by the formula

$$(q, \vec{v}) \mapsto \frac{q_{\vec{v}}}{\|q_{\vec{v}}\|_2}$$

Let (c_0, c_1) be the representative of q and it is easy to check that $((c_0, c_1) \times \vec{v}) \sim ((c_0, -c_1) \times (-\vec{v}))$ in ϕ, i.e.

$$\phi((c_0, c_1) \times \vec{v})) = \phi((c_0, -c_1) \times (-\vec{v})))$$

Notice that it is an 2-fold covering map and the orbit space is homeomorphic to a Klein bottle.

Local Behavior of Feature Space in Natural Image

There is a large portion of the space of patches which is topologically equivalent to a well-known two-manifold, the Klein bottle.

Topological Features of CNN Trained for Image Classification

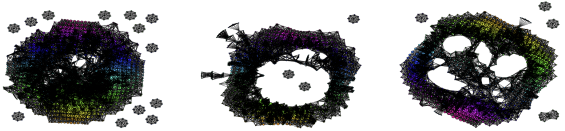
Regard a 3 by 3 kernel in a CNN as a 9-dimensional vector, which is called the weight vector. The distribution space of weight vectors on a given convolutional layer is called the weight space.

Motivations

- Filters of CNNs have the same size of local patches. Can we try to put similar analysis on the space of weight vectors?
- We know that neurons in the primary visual cortex and the weight vectors and the filters in CNNs are reflecting responses or functions on the space of patches. There is a conjecture that weight vectors or filters are also distributed in similar structures and can be regarded as a function on patches via inner product constructions to better extract features.

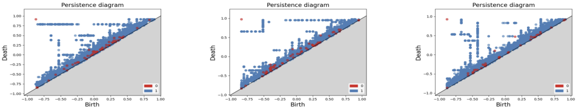
Topological Data Analysis on Weight Space of CNNs

Train 100 CNNs to high test accuracy. These 100 trained CNNs give us 64 Ã 100 = 6400 9-dimensional points (first layer spatial filters) which we standarize.



a) untrained weights b) 5-epochs trained weights c) 20-epochs trained weights

Figure: Learning process of CNN trained on MNIST



a) untrained weights b) 5-epochs weights c) 20-epochs weights

Figure: Persistence diagrams

Topological Data Analysis on Weight Space of CNNs

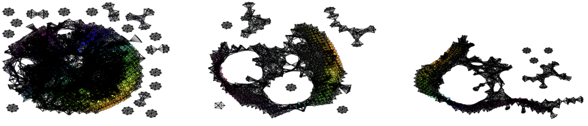


Figure: Learning process of CNN trained on CIFAR-10

It improved interpretability of neural networks.

- It shows that the spaces of spatial filters learn simple global structures.
- Both underfitting and overfitting can cause the disappearance of this simple structure.

Correlation between the Feature Space of Natural Images and the Weight Space of CNN

The core topological features of well-trained weight space is consistent with the space of high contrast patches.

Feature Extraction: A kernel act on the vector of local patch as the inner product to extract the features.The weight vectors have similar distribution so that the result of inner product is larger than that of low contrast patches, which makes high-contrast patches have a greater impact in subsequent classification.

Topological Covolutional Neural Networks

Klein Filters and Circle Filters [Love et al. 2023]

- $$f(x, y) = q(\lambda x + \mu y)$$

  where $\lambda^2 + \mu^2 = 1$ and q is a quadratic function with single variable satisfying that the quadratic coefficient is a multiple of the constant term.

- $$F_{\mathcal{K}}(\theta_1, \theta_2)(x, y) = \sin(\theta_2)(\cos(\theta_1)x + \sin(\theta_1)y)$$

  $+ \cos(\theta_2)Q(\cos(\theta_1)x + \sin(\theta_1)y)$, where $Q(t) = 2t^2 - 1$.

- To define $F_{\mathcal{S}^1}$, consider the composite of $F_{\mathcal{K}}$ and the inclusion map from S^1 to $\mathcal{K}$ which maps θ to $(\theta, \pi/2)$:

$$F_{\mathcal{S}^1}(\theta)(x, y) := F_{\mathcal{K}}(\theta, \pi/2)(x, y) = \cos(\theta)x + \sin(\theta)y,$$

Klein Filters and Circle Filters

$$Filter(\kappa)(n, m) = \int_{-1+\frac{2m}{2s+1}}^{-1+\frac{2(m+1)}{2s+1}} \int_{-1+\frac{2n}{2s+1}}^{-1+\frac{2(n+1)}{2s+1}} F_M(\kappa)(x, y)dxdy$$

- It aims to discrete the kernel by calculating the average of integration.
- The kernel in Klein Filters is the combination of the first and second-order differential operators.

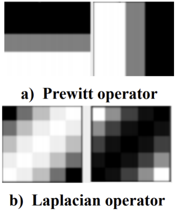
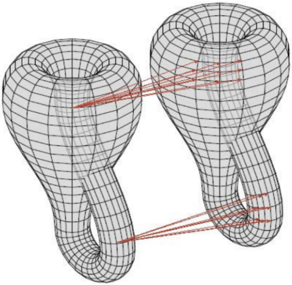


Figure: Kernels in Klein Filters

Klein One Layer and Circle One Layer



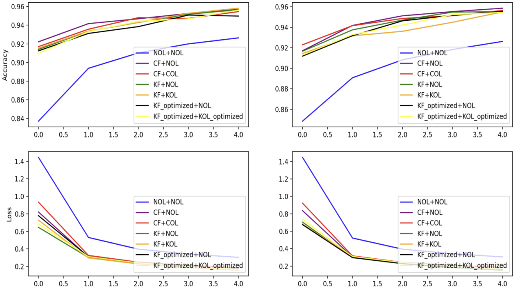
Regularization: In COL and KOL, the filter space are parameterized by a discretization of a circle or Klein Bottle (topological features embedding) and during the entire training process, any weights connecting slices beyond a fixed threshold distance on the Klein bottle are maintained as zero. The locality conditions promote the network to learn local features both within the raw pixels and in the tangential directions in the geometry of a circle or Klein bottle.

Optimized Klein One Layer

The theoretical analysis reveals that the points of highest density in the Klein bottle are located on the three-circle model, whereas points of lowest density are generated by pure quadratic functions on non-vertical and non-horizontal directions. Hence we simulate the true Klein bottle distribution by placing a varying number of convolution kernels in different regions.

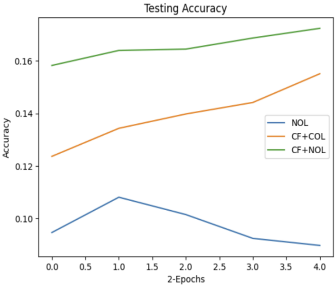
TCNN vs CNN

- Training Performance



TCNN vs CNN

- Generalizability



Future Work

- Establish a connection between geometric properties and intrinsic properties of neural networks, which makes the topological features a criterion to measure the performance of neural networks.
- Geometric Analysis on other areas.

Thanks for your listening!