

Topological Persistence of the Neural Embedding of the Archetypal Subspace

S. Suresh, V. Abrol *Senior Member, IEEE*, K. Kalyanaraman, S. Dutta Roy *Senior Member, IEEE*

Abstract—Persistent homology (PH) has recently emerged as a topological tool to analyse the learnability in the layer-by-layer feature embedding space of a well-trained deep neural network. However, the computation of the PH is computationally impractical in higher dimensions or for large-scale datasets. To address this, we analyse the homological features of the embedding space of the principal convex hull of the data. This is achieved using Archetypal Analysis (AA), a factor analysis method that expresses the data as a convex combination of extremal points on the convex hull. Through homological inference of point clouds, we show that, due to geometric properties, adding archetypes to the data does not change the homology of the underlying space approximated by the data. We use this analysis to prove that the homological complexity of archetypes of a class can be used as the minimal homological expressivity needed for a classifier to avoid misclassification. Since inferring the homological complexity of the underlying space for real-world data can be ambiguous, we attempt to quantify the transformation of the homological complexity of the embedding space of the class archetypes through the topological persistence of their embedding space. For this, we employ “topological average life,” which measures the average persistence of the homologies in the point cloud. We find a negative correlation between the fine tuned model’s performance and the topological average life of the archetypal subspace.

Index Terms—Archetypes, Persistent homology, Expressivity, Deep neural networks

I. INTRODUCTION

A popular method of interpreting and visualising the learning behaviour of a deep neural network (DNN) is through its embedding space [1]–[3]. One way of carrying this out is by studying the “persisting shape” of the data through its embedding across the many layers of a DNN. In the language of algebraic topology, the persisting shapes or structures which, in turn, characterise the shape of the underlying or embedding space, are homological structures such as connected components, holes, and higher dimensional voids. We consider the embedding space of the data to be a geometric object of interest, and it is transformed through maps (which may be

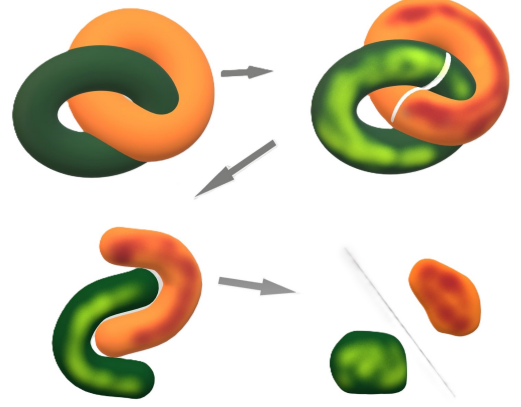


Fig. 1. Transformation of a two-class manifold as it passes through a DNN classifier. The first arrow (grey) denotes a non-homeomorphic map, followed by homeomorphic maps that smoothens the decision boundaries. The DNN “breaks” the homology in the data as well as manifold of each of the classes and reduces them all to a singleton point or equivalently a cluster. Without breaking the void, finding a hyperplane (decision boundary) that separates the two classes is impossible. This behaviour can also be approximated using class archetypes that lie on the convex hull of the manifold (shown as red points). Archetypes influence the breaking of homology as well as the final decision boundary.

homeomorphic) based on a particular task being performed by the DNN. For instance, segmentation of an image would require a homeomorphic map in order to preserve the homology [4] while classification would be a more general map that collapses the subspace of images (in each class) to a singleton element representing that class [5]. To illustrate, consider the topology of input data shown in Fig. 1 as it transforms through the layers to perform a binary classification task. Ideally, to achieve a reasonable separation between classes, each of the topological structures (the different homology classes) should be transformed across the various layers of the DNN and finally be reduced to a singleton point at the output layer representing its category. The number of such structures in the embedded manifold of the data at the input and output of each layer may be taken to be a measure of the nature of the mathematical map that is a DNN layer; topological structure preservation across this layer would indicate that the map is a homeomorphism, and otherwise it is not. In algebraic topology, the number of topological structures helps define a homological complexity measure, which can be quantified as the sum of the Betti numbers of the embedding manifold. In [1], the authors have provided a theoretical upper and lower bound connecting the homological complexity with the depth of a DNN. Furthermore, [2] relates expressivity of a

Manuscript received June, 2024; revised xxx; Accepted xxx. This work is supported by the SERB-Core Research Grant (CRG/2023/003352), Government of India and in part by Infosys Foundation via the Infosys Centre for AI, IITD. The associate editor coordinating the review of this manuscript and approving it for publication was xxx. (Corresponding author: Surya Suresh.)

Surya Suresh and Vinayak Abrol are with Infosys Centre for AI and CSE Department, IIT Delhi, India. (e-mail: [suryakas, abrol]@iitd.ac.in).

Kaushik Kalyanaraman is with the Mathematics Department, IIT Delhi, India. (e-mail: [kaushik@iitd.ac.in]).

Sumantra Dutta Roy is with the EE Department, IIT Delhi (e-mail: sumantra@ee.iitd.ac.in).

Digital Object Identifier xxx/xxxx

classifier to the homology of data such that an architecture incapable of attaining a particular homological complexity would misclassify inputs; see also Figs. 1 and 5 for illustration.

In this work, we use persistent homology (PH) [6], a computational method for extracting the homology intrinsic to the space in which a data set thought of as a point cloud resides. A major drawback of PH is its computational complexity (informally, cubic in the size of the data [2], [7], [8]), which can be alleviated by approximating a subsample of the data while still preserving the underlying homology. A witness complex [9] is one such tool for data sampling; however, it does not appear to scale well to higher dimensional data. As an alternative, the authors in [10] proposed to approximate the topology of the underlying manifold through repeated random sampling, which then requires statistical analysis without any computational gain. In contrast, we analyse the homological features of the embedding space through the principal convex hull of the data. Towards this end, we employ archetypal analysis (AA), a factor analysis method to determine representative points on the principal convex hull of a point cloud dataset, and follow this up by computing the PH over the space spanned by the archetypes. Intuitively, due to a geometric property that archetypes of a class or cluster archetypes are extremal points, it is possible that the archetypes do not necessarily capture the correct homology for their generating data, thereby leading to incorrect classifications or misinterpretations of a task by a DNN [11]. Therefore, it is important to consider that the homology of the decision region is influenced by the homology (and therefore homological complexity) of the archetypes of each class (as illustrated in Fig. 1) in the layers of a DNN.

Furthermore, in PH, the relevance of a homology is determined by the length of its persistence or its lifespan. That is, the longer a homological class persists relative to another, the more likely it is to truly represent the homology of the underlying space. However, this computation is fraught with ambiguities, especially for high dimensional real-world data. Considering therefore a recently proposed topological measure, the “topological average life” [12], we prove that the change in the homological complexity of data representing a class across the DNN can be quantified using the average topological lifetime of its class archetypes. We then empirically verify the applicability of our results and characterise various practical DNN architectures. In addition, we also study the correlation between the performance of such deep models and the topological average life across layers at depth. Our results shall provide practitioners with an alternative topological tool for understanding the learning behaviour of DNNs.

A. Contributions

- 1) *How do the archetypes’ geometric properties relate to a data’s homology?*
 - a) We prove that augmenting data with its archetypes does not change the underlying homology (Theorem 1).
 - b) We prove that archetypes can be used to approximate (up to a constant factor) an upper bound for the homological complexity of data (Theorem 2).

- c) We extend the result in Theorem 2 to archetypes of data classes (Corollary 1).

2) *Topological characterisation of a DNN classifier*

- a) We show that the homology of class archetypes having the correctly homological complexity is necessary to avoid misclassification (Theorem 3).
- b) We show that the topological average life of the neural embedding space of a class across the layers needs to be reduced in size (as a result, the homological complexity also reduces) in a DNN classifier (Theorem 4).

3) *Empirical evaluation of practical DNN models*

- a) We empirically study the persistent homology of neural embedding space of class archetypes in relation to the transitions of data within the classifier.
- b) We demonstrate a negative correlation between a fine-tuned model’s performance and the topological average life of the archetypal subspace.
- c) We verify our claims on various popular DNN architectures and real image datasets.

II. RELATED WORK

Embedding spaces of a neural network are a rich representation of the features of the point cloud pertinent to an underlying task, such as classification and object detection. Recently, methods from topology [5], [12] have been used to characterise the network using topological invariants of persistent homology (PH) as a measure, such as Betti numbers β_i and Euler characteristics [6], [13]. Preserving the topological structure of the data has led to a number of applications such as image segmentation [4], image fusion [14], and dimensionality reduction [15]. In addition, a series of works have also demonstrated how topological features serve as a potential measure of generalization in DNNs [2], [9]. For instance, [2] showed how the homological complexity of the data can be used to estimate the complexity of a neural architecture. Our work extends this idea to the archetypal subspace for real-world datasets. In [12], authors proposed a topological score based on the topological average life of the DNN during training to measure the performance gap of a model between the training and test datasets. In [5], the authors established a link between the topological transformation of decaying Betti numbers through depth and the expressivity of a neural network with different activation functions. The measures of persistence of the homology derived from PH have been employed to study the expressivity of neural networks by analysing weight connectivity and comparing network complexity [16]. Further [12] shows that the persistence of the homology calculated during training of the weight matrix can be used to generalise to the test score. However, most existing studies are restricted to simple feed-forward networks and simulated/toy datasets. In contrast, the computational complexity around scaling a topological method to higher-dimensional multi-class problems is tackled in this work by exploiting the embedding space of the archetypes instead of the whole dataset.

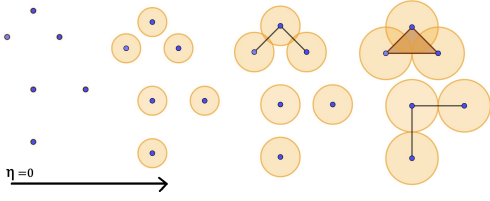


Fig. 2. Persistence homology on a point cloud [6]. As η increases, the subcomplexes are formed by connecting points, and when all the points are connected ($\eta = \infty$), there is only one connected component.

III. PRELIMINARIES

A. Topology

- **Topological invariants** are mathematical quantities associated with a **topological space** X which do not change under continuous deformations [6], [13].
- The **Euler characteristic** χ is an integer and one of the simplest topological invariants for X .
- A **simplicial complex** Γ is a set whose elements are simplices. A 0-simplex is a point, a 1-simplex is a line segment, a 2-simplex is a filled triangle, and a k -simplex is the k -dimensional counterpart. The characterizing properties of Γ are that (i) for a simplex $\sigma \in \Gamma$, all its faces are also in Γ , and (ii) for simplices $\sigma, \tau \in \Gamma$, either $\sigma \cap \tau = \phi$ or it is a shared face of σ and τ .
- Given simplicial complexes Γ, Γ' , a **simplicial map** $g : \Gamma \rightarrow \Gamma'$ linearly maps simplices of Γ to simplices of Γ' .
- **Homology** of a topological space X is defined to be $H(X) := \{H_k(X)\}_{k=0}^{\infty}$ where H_k is the k^{th} homology. The homology H_0 represent the connected components, H_1 holes, H_2 voids, and H_k their k dimensional counterparts. Further, for a topological space X , the k^{th} **Betti number** (β_k) is the dimension of k^{th} homology or $\dim H_k(X)$. In this work, we denote the **data** as $D \subset \mathbb{R}^p$ and its associated topological space by X_D . In our mathematical proofs, the point cloud D is embedded in some appropriate Euclidean space. Furthermore, D_c denotes a class in D for $c \in C$ where C is the index set of all classes. The **cardinality** or the number of elements in D is denoted $|D|$. The **Euler characteristic** of X can also be obtained as $\chi(X) = \sum_{k=0}^{\infty} (-1)^k \beta_k$.
- A **filtration** is essentially a function $\mathcal{F} : \mathbb{R} \rightarrow \Gamma$ that maps to (equivalently filters out) a subcomplex of Γ . It has the property that for $\eta_1, \eta_2 \in \mathbb{R}$ such that $\eta_1 \leq \eta_2$, $\Gamma_{\eta_1} := \mathcal{F}(\eta_1)$ is a subcomplex of $\Gamma_{\eta_2} := \mathcal{F}(\eta_2)$, that is, we have the inclusion $\Gamma_{\eta_1} \subseteq \Gamma_{\eta_2} \subseteq \Gamma$. $\eta \in \mathbb{R}$ is usually referred to as the **threshold** of the filtration function. **Persistent homology (PH)** is a topological tool that uses filtration to construct subcomplexes Γ_η to extract the homology of a point cloud D (see Figure 2). Note that because of the inclusion property not all features extracted through PH is a homology of the topological space X_D . We will consider that the PH was computed on the point cloud till only a single connected component remains.
- **Homological features** from PH $\hat{H}$ can be represented in the form of barcodes or persistence diagrams (PD). **Persistence diagram (PD)** for a point cloud D is a

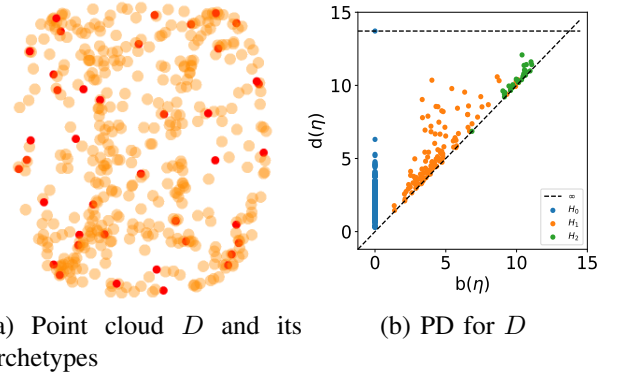


Fig. 3. Torus is the underlying topological space X with Betti number $\{\beta_k\}_{k=0}^{\infty} = \{1, 2, 1, 0, 0, \dots\}$; (a) Noisy point cloud D with archetypes $\hat{A}_D$ (red), and (b) PD for D ,

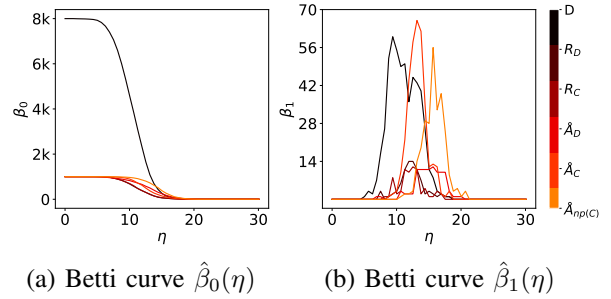


Fig. 4. Betti curves (a) $\hat{\beta}_0(\eta)$ and (b) $\hat{\beta}_1(\eta)$ for the STL10 dataset (D), archetypes ($\hat{A}_D$), classwise archetypes ($\hat{A}_C$), nearest points to $\hat{A}_C$ classwise ($\hat{A}_{np}$), randomly sampled subset (R_D) and class-wise randomly sampled subset of dataset (R_C).

series of points in the extended plane $\mathbb{R}^2$ that represents the birth and death information of the homology that appears in the subcomplex during filtration. Further, the η at which i^{th} homology $\hat{H}^i$ appears in the PD is the **birth**, denoted as b_i or $b_i(\eta)$ and the η at which the homology disappear is the **death** denoted as d_i or $d_i(\eta)$. We denote all the homology that arises in the PD as $\hat{H}(PD)$. The significant persisting homology is the possible homology of the underlying topological space X_D of the point cloud D . The dotted line in the PD (see Fig. 3 (b)) represents ∞ and the points at this line represent homology that never dies.

- The k^{th} **Betti-curve** $\hat{\beta}_k$ (Fig. 4) maps PD to an integer-valued curve for different values of η . In the Betti-curve diagram, the x -axis represents η , and the y -axis represents the number of k^{th} -homologies for the subcomplex Γ_η that are still persisting at a particular η . The k^{th} **Betti number** at Γ_η is the value of k^{th} Betti-curve at η denoted by $\beta_k(\eta)$. For a N -point cloud, we have $\beta_0(0) = N$ and $\beta_0(\eta) \rightarrow 1$ as $\eta \rightarrow \infty$. Similarly, in a N -point cloud for some η' , $\beta_k(\eta') \geq 0$ and $\beta_k(\eta) \rightarrow 0$ as $\eta \rightarrow \infty$ (see Fig. 4). Ideally, the smaller the max η for which β_0 dies, the closer the points are in the point cloud.
- We define the **homological/topological complexity** of a topological space X as the sum of Betti numbers, that is, $\omega(X) := \sum_k \beta_k(X)$. We denote the homological

complexity at Γ_η of the PD of D as $\omega_\eta(D) = \sum_k \hat{\beta}_k^D(\eta)$.

- The **lifespan** of a homology $\hat{H}^i(PD)$ that determine its persistence is given by $|b_i - d_i|$. The **topological average life** of the PD of D is given by $\lambda_D = \frac{\sum_i |b_i - d_i|}{|\hat{H}(PD)|}$ for i^{th} homology that appear in the PD with finite lifespan. For points that never die, their lifespan is ∞ , and we have excluded them from our definition. Further, note that both lifespan and λ are functions of η .
- A **homological feature** of a topological space X denoted $hfs(X)$ is the smallest positive homological critical value [17] of η . Homological critical values are the values of η that induce a change in the homology.
- Consider a metric space equipped with the metric ρ then for nonempty subsets U, V of the metric space. We define $\rho(v, U) = \inf_{u \in U} \rho(v, u)$ and $e(V, U) = \sup_{v \in V} \rho(v, U)$. Then $e(V, U)$ is called excess of V over U . The **Hausdorff distance** between V and U is defined as $\rho_e(V, U) = \max\{e(V, U), e(U, V)\}$.

For ease of reference, we list the various symbols we have used in Table I.

B. Archetypes

Archetypes ($\hat{A}$) are extremal points represented as convex combinations of data itself [18]–[20] (see Fig. 3(a)). Three main useful properties of archetypes are: (1) they lie on the principal convex hull (or boundary or edge) of the data; (2) they closely follow the data but don't necessarily belong to the data; (3) they form a vector basis for the vector space of the data, that is, they span the space of data. Since archetypes are not exact points of the original data D , one can find the closest point in the point cloud ($\hat{A}_{np(D)}$) to represent the archetypes without replacement. Let D be a point cloud, then for $\forall x_j \in D$ and $\hat{x}$, the center of D .

$$\text{Objective function: } \min_{\hat{A}} F(\hat{A}), \quad (1)$$

$$\rho(a_i, \hat{x}) \geq \rho(x_j, \hat{x}) \quad \forall a_i \in \hat{A}, \quad (2)$$

$$\sum_i^m \alpha_i a_i = x_j, \quad a_i \in \hat{A} \text{ and } \sum_i \alpha_i = 1, \quad (3)$$

where m is the rank of the vector space of D , $F(\hat{A}) = [1/|D| \sum_{i=1}^N \rho^2(x_i, \text{Hull}(D))]^{1/2}$ [20] and ρ is the metric for the underlying metric space. We will denote the archetype of a point cloud D as $\hat{A}_D$. As the number of archetypes $n \rightarrow N$, $\bar{F}(\hat{A}) \rightarrow 0$, since $\hat{A}_D = D$ at $n = N$.

IV. HOMOLOGY OF ARCHETYPES

In this section, we theoretically study the relation of archetypes to the homology of data and establish that archetypal analysis is a valuable practical tool to scale a topological method to higher dimensional and large-scale datasets/models.

A. Homology inference for archetypes

Archetypes are extremal points on the convex hull of data, and due to this intrinsic geometric property, augmenting data with its archetypes should not change the underlying topology.

TABLE I
SYMBOLS

Symbol	Meaning
D	Dataset
$ D $ or $N = D $	Cardinality of D
X_D	Topological space of D
$H(X_D)$	Homology of space X_D
$\beta_k(X)$	k^{th} Betti number of X
$\hat{A}_D$	Archetypes of D
η	Threshold associated with filtration
Γ_η	Subcomplex at η in the filtration
$\hat{\beta}_k(\eta)$ or $\hat{\beta}_k$	k^{th} Betti number of Γ_η
$\omega_\eta(D)$	Homological complexity of Γ_η
f	Classifier on D
$H_s f$	Support homology of f
λ	Topological average life

Intuitively, this is a direct consequence of the homology inference theorem [6], [17], [21], which states that the quality with which a point cloud approximates the homology of an underlying space depends on the Hausdorff distance between them. Here, we prove that the difference in the homology of the underlying spaces for the point clouds $D \cup \hat{A}_D$ and D depends only on the objective function $F(\hat{A})$ for computing the archetypes [20].

Theorem 1. *Let $\bar{F}(\hat{A})$ represent the value of the objective function in Eq.(1) after optimisation. Let X' and X be the topological spaces associated with the point clouds $D \cup \hat{A}_D$ and D , respectively. Then, $H(X') \simeq H(X)$ (that is, the homology groups of X' and X are isomorphic) given $\bar{F}(\hat{A})$ is sufficiently small.*

Proof. Let $D' = D \cup \hat{A}_D$ and D be point clouds which approximate the homology groups of their underlying topological spaces X and X' with $\eta \geq \rho_e(D', D)$. Then by the homology inference theorem, we have $\dim \hat{H}_k(D') = \dim \hat{H}_k(D) \quad \forall \eta > \rho_e(D', D)$, and thus X' and X have isomorphic homology groups provided $\rho_e(D', D)$ is sufficiently small. We complete the proof by actually showing that $\rho_e(D', D) = \bar{F}(\hat{A})$. To see this, consider the quantity $e(D, D') = \sup_{x \in D} \rho(x, D')$ where $\rho(x, D') = \inf_{a \in D'} \rho(x, a) = 0$ which follows since $D \subset D'$. This implies that $e(D, D') = 0$. Now consider $e(D', D) = \sup_{a \in D'} \rho(a, D)$. Here $\rho(a, D) = 0 \quad \forall a \in D$ and $\rho(a, D) = \bar{F}(\hat{A}) \quad \forall a \in \hat{A}$. This implies that $e(D', D) = \bar{F}(\hat{A})$. Thus $\rho_e(D', D) = \max\{e(D, D'), e(D', D)\} = \bar{F}(\hat{A})$. Thus for $\bar{F}(\hat{A})$ sufficiently small $H(X) \simeq H(X')$. $\square$

This important result also helps us prove extensions of our main results in Section V for DNNs.

B. Upper bound for homological complexity

Our main result here establishes a connection between the homology of $\hat{A}$ and the homology of D at each η . To motivate this, observe the Betti curves in Fig. 4 for the STL10 dataset. We see that: 1) $\hat{\beta}_0^{\hat{A}_D}(\eta)$ values are persisting longer for $\hat{A}_D$, and 2) $\hat{\beta}_1^{\hat{A}_D}(\eta)$ best resemble $\hat{\beta}_1^D(\eta)$. This suggests a relation between the homological complexity ω of the data and its archetypes. Formally, we quantify $\hat{\beta}$ using ω_η and prove that

$\omega_\eta(\hat{A}_D)$ upper bounds $\omega_\eta(D)$. In order to do so, we use the following result from [22].

Theorem (1.2, [22]). Let $g : \Gamma \rightarrow \Gamma'$ be a simplicial map of finite simplicial complexes such that $\chi(g^{-1}(\gamma')) = \kappa \forall \gamma' \in \Gamma'$ where $\chi(\Gamma)$ is the Euler characteristic of Γ and g^{-1} denotes the preimage of $\gamma' \in \Gamma'$. Then, $\chi(\Gamma) = \kappa \cdot \chi(\Gamma')$.

Now, we provide a similar characterization for the homological complexity ω of the simplicial complex $\Gamma_\eta(D)$ obtained for a threshold η in a PD for D , and $\Gamma_\eta(\hat{A}_D)$ is the simplicial complex for the same η in a PD for the archetypes $\hat{A}_D$. In the following, we shall consider g to be any surjective simplicial map between $\Gamma_\eta(D)$ and $\Gamma_\eta(\hat{A}_D)$.

Theorem 2. Suppose there are N points in $D \subset \mathbb{R}^p$, and we have n archetypes $\hat{A}_D$. Provided $n \geq m + 1$ where m is the minimum number of $\hat{A}_D$ needed to span D , for those threshold η in the PD such that there exists a surjective map between $\Gamma_\eta(D)$ and $\Gamma_\eta(\hat{A}_D)$, we have that:

$$\omega_\eta(D) \leq \left\lceil \frac{N}{n} \right\rceil \omega_\eta(\hat{A}_D). \quad (4)$$

Proof. Here we demonstrate that $\kappa = \left\lceil \frac{N}{n} \right\rceil$. This is readily inferred by considering that any simplicial map from $\Gamma_\eta(D)$ to $\Gamma_\eta(\hat{A}_D)$ needs to map N vertices of D to n vertices of $\hat{A}_D$. Thus, a preimage of the simplicial map can contain at most $\left\lceil \frac{N}{n} \right\rceil$ vertices. As a result, for the base case of the proof in [22], $\kappa = \left\lceil \frac{N}{n} \right\rceil$, and we have that:

$$\hat{\beta}_0^D(0) = \left\lceil \frac{N}{n} \right\rceil \hat{\beta}_0^{\hat{A}_D}(0).$$

Next, for any finite η , due to the defining property of distances for archetypes as in Eqn. (1), we have that:

$$\hat{\beta}_0^D(\eta) = \left\lceil \frac{N}{n} \right\rceil \hat{\beta}_0^{\hat{A}_D}(\eta).$$

The rest of the proof follows from an induction argument for $\omega_\eta(D)$ and $\omega_\eta(\hat{A}_D)$ following a similar line of reasoning as in [22] for $\chi(\Gamma)$ and $\chi(\Gamma')$. $\square$

Observe that as $n \rightarrow N$, $\lim_{n \rightarrow N} \left\lceil \frac{N}{n} \right\rceil \omega_\eta(A) = \omega_\eta(X) \forall \eta$.

When $n = N$, $\hat{A}_D = D$, with reference to Fig. 6 in the Section VII, the lighter blue lines associated with the $\hat{A}_D$ are upper bounding the $\omega_\eta(D)$ shown in darker blue. In Theorem 2 and its corollary we attempt to generalise this property to satisfy for all η and $n \leq N$.

Corollary 1. Suppose we have a finite point cloud $D = \bigcup_{c \in C} D_c$ where $D_c \subset \mathbb{R}^p$, $D_{c_i} \cap D_{c_j} = \emptyset$, and $\hat{A}_C = \bigcup_{c \in C} \hat{A}_c$ is the union of n archetypes from each c . Provided $n \geq m + 1$, where $m = \max_{c \in C} \{m_c\}$ such that m_c is the minimum $\hat{A}_c$ needed to span D_c and $N = n \sum_c |D_c|$. Then we have that:

$$\omega_\eta(D) \leq N_C \omega_\eta(\hat{A}_C) \text{ where } N_C = \left\lceil \frac{N}{(n|C|)} \right\rceil \quad (5)$$

The proof follows directly from Theorem 2, where here we define a simplicial map that takes every point in the D_c to $\hat{A}_c$ such that the N vertices of D maps to the $n|C|$ vertices of $\hat{A}_C$. The result in Theorem 2 and Corollary 1 helps us to derive a relation between the homological complexity of D and $\hat{A}_C$. Later in Section V, we relate the homology of the X_D and that of a classifier f . Since it is ambiguous to compute $H(X_D)$ exactly for real-world datasets, our result relies on Corollary 2 that can at most provide a loose upper bound for $\omega(X_D)$. Nevertheless, one can use it to study some interesting properties, such as the impact of sample complexity and input size on the persistence of homology (see Section VII).

Corollary 2. Let $M = \max_\eta \{\omega_\eta(\hat{A}_D)\}$, then for the topological space X_D of the point cloud D with at most $\hat{m}$ non-zero homology, we have $\omega(X_D) \leq \kappa M \hat{m}$ where $N = |D|$, $n = |\hat{A}_D|$ and $\kappa = \left\lceil \frac{N}{n} \right\rceil$.

The proof is a direct consequence of theorem 2 and the insight that there is at most $\hat{m}$ non zero Betti numbers $\beta_k(X_D)$ and each of these are upper bounded at some η by $\kappa \omega_\eta(\hat{A}_D)$.



Fig. 5. Transformation of a two-class manifold. Here, the underlying space has a higher homological complexity in both classes compared to Fig.1. Best viewed in colour.

V. HOMOLOGY OF NEURAL EMBEDDINGS

In this section, we study how the topology of feature embedding space changes as it passes through the layers of a well-trained DNN. As usual, a computationally efficient way is to work with archetypes instead of all data samples, and towards this aim, Theorem 3 extends the homological generalisation to the sub-space of class/cluster archetype $\hat{A}_c$. In order to quantify the topological behaviour of neural embedding space across layers, the result in Theorem 4 presents a way of using λ_A for such categorization.

A. Homological generalisation for $\hat{A}$

We extend the homological generalisation theorem from [2] to class archetypes $\hat{A}_c$ and show that homology of $\hat{A}_c$ defines a minimal condition to eliminate a DNN architecture that cannot express the homology of the underlying space of a class. Such a DNN will always misclassify some points. As an illustration, see Fig. 5 and observe that even after the two classes have been separated after multiple transformations, there still exists homology in the class, which prevents the class from collapsing to a single point. Thus, we argue that the homology of the underlying space plays an important role in the expressiveness of the classifier. There are two reasons for extending homological generalisation to $\hat{A}_c$ that seem to make sense: 1) the classification of the $\hat{A}_c$ is a more substantial generalisation task than the classification of the class alone,

and 2) augmenting a class with its $\hat{A}_c$ does not introduce any new homology to the class.

Consider $X \times \{0, 1\}$ to be the topological space associated with the data D and archetypes $\hat{A}_C$. The data is supposed to be sampled from some probability distribution on this space, but we are not here concerned with the underlying statistical aspects of this data, except that in some appropriate sense of a limit, the data well approximates this underlying space. We consider binary classification with positive (+) and negative (-) classes where X_+ and $\hat{A}_{X_+}$ represent the support of the distribution of points with positive class and the archetypes of the positive class, respectively. Now for the family of binary classifiers $\mathcal{F} = \{f : X \rightarrow \{0, 1\}\}$, we define the homology of the set of x such that $f(x) > 0$ as $H_s(f) := H[f^{-1}((0, \infty))]$, the support homology of the function f . Here, for a binary classifier, $H_s(f)$ is an estimate of the number of holes in the positive decision of f .

Corollary 3. *For $\min F(A_+)$ sufficiently small, we have $H(X_+) \simeq H(X_+ \cup \hat{A}_{X_+})$.*

The proof follows from the Theorem 1 as X_+ and $X_+ \cup \hat{A}_{X_+}$ are the topological spaces associated with the point cloud $D = D_+$ and $D' = D_+ \cup \hat{A}_{D_+}$.

Theorem (3.1 of [2]). If $X = X_- \cup X_+$ and $\forall f \in \mathcal{F}$ with $H_s(f) \not\simeq H(X_+)$, then $\forall f \in \mathcal{F}$, $\exists V \subset X_+$ so that f misclassifies every $x \in V$.

See here that f is a classifier on in $X_+ \cup \hat{A}_{X_+}$ and thus the minimum requirement for all points in the class + to be correctly classified as per the above theorem is that $H_s f(X_+ \cup \hat{A}_{X_+}) \simeq H(X_+ \cup \hat{A}_{X_+})$, then $H_s f(X_+) \simeq H(X_+)$ but $H(X_+) \simeq H(X_+ \cup \hat{A}_{X_+})$ and thus $H_s f(X_+ \cup \hat{A}_{X_+}) \simeq H_s f(X_+)$ follows. Thus we look at the converse whenever $H_s f(X_+) \simeq H_s f(X_+ \cup \hat{A}_{X_+})$ is true. We have the homological generalization theorem extended to the $\hat{A}_{X_+}$. The authors of [2] use a contradiction in their proof, which essentially states that if f maps each singleton into its proper partition on X , we have $H(X_+ \cup \hat{A}_{X_+}) \simeq H_s f(X_+ \cup \hat{A}_{X_+})$. We build on this key idea to prove our next result for archetypes.

Theorem 3 (Homological generalisation extended to $\hat{A}_{X_+}$). *Let X be a topological space, X_+ be some open subspace, and $\hat{A}_{X_+}$ its archetypes. Let $\mathcal{F} \subset 2^X$ such that $f \in \mathcal{F}$ with $H_s f(\hat{A}_{X_+}) \not\simeq H(\hat{A}_{X_+})$ and $H_s f(X_+) \simeq H_s f(X_+ \cup \hat{A}_{X_+})$, then $\forall f \in \mathcal{F}$ will misclassify for some points in X_+ .*

Proof. Since $H_s f(\hat{A}_{X_+}) \not\simeq H(\hat{A}_{X_+})$, $\exists V \subset \hat{A}_{X_+}$ such that f misclassifies every $v \in V$. Since, $\hat{A}_{X_+} \subset X_+ \cup \hat{A}_{X_+} \therefore V \subset X_+ \cup \hat{A}_{X_+}$. Now by the Theorem 3.1 of [2], we have $H_s f(X_+ \cup \hat{A}_{X_+}) \not\simeq H(X_+ \cup \hat{A}_{X_+})$. But by the claim 3, we have $H(X_+) \simeq H(X_+ \cup \hat{A}_{X_+})$ and $H_s f(X_+) \not\simeq H(X_+)$. Hence, a set $V' \subset X_+$ exists such that f misclassifies every $v \in V'$. $\square$

Thus, $H_s f(\hat{A}_{X_+}) \simeq H(\hat{A}_{X_+})$ is a necessary homological condition to ensure that the architecture is capable of or has the homological expressivity needed to classify the archetypes of the class. Since there is no guarantee that $H_s f(X_+) \simeq H_s f(X_+ \cup \hat{A}_{X_+})$ for f generally, this is not the best condition for a DNN architecture. However, it is also important to note

that archetypes of class are extreme points of the class, making them more vulnerable to misclassification. Nevertheless, random sampling will not satisfy this criterion.

A key outcome of this result is that it is enough to use $\hat{A}_c$ for topological characterization of D (or D_c) through the computation of PH over $\hat{A}_C$ (or $\hat{A}_c$). Indeed, there are efficient algorithms for the computation of a small number of $\hat{A}$ with complexity $\mathcal{O}(m^2 N)$ where $N = |D|$ and m the embedding dimension [23]. The worst-case computational complexity of PH for N is $\mathcal{O}(N^3)$ although in practice this is better. Therefore, by choosing $N_C \ll N$, the cost of computing PH for $\hat{A}_C$ would be much better than directly computing the PH of D even after the cost of computing $\hat{A}_c$. Thus, use of $\hat{A}_c$ for topological characterization is both scalable and practical way of computing PH for large datasets for topological data analysis in machine learning. Finally, archetypal analysis like this can be combined with more recent and novel ideas like vectorized signed barcodes that efficiently characterize the multiparameter and multiscale topology of a dataset [24] and our results provide the basis for such future explorations.

B. Topological average life of neural embedding

The result in Theorem 3 states that we can eliminate classifiers f (here DNN) that cannot express the homology $H(\hat{A}_{X_c})$. The implementation of this theorem in practice would need the computation of $H(f)$ and $H(\hat{A}_{X_c})$. An alternative way of looking at this problem is by using the embedding space of the DNN. In the following theorem, we prove that the changing homology of the underlying topological space X_D of the point cloud D can be captured by the average life λ across the layers of the DNN.

An ideal classifier that collapses D_c to a singleton element representing that class also collapses the homology X_{D_c} , leaving only the 0th-Betti number equal to 1. In other words, the layers of the DNN are non-homeomorphic functions that transform the homology in X_{D_c} , reducing the values of $\beta_i(X_{D_c})$ of X_{D_c} . We want to quantify the changing $\beta_i(X_{D_c})$ using the PD of embedding space of D_c at each layer and observe the reducing trend in $\beta_i(X_{D_c})$. To this aim, we will analyse the average life λ_c of the PD for D_c at each layer. Let us review the following two facts: 1) PH is computed until there is only one connected component; hence, only this connected component will have an infinite lifespan, and 2) the average life λ is finite and does not include the infinite lifespan.

Theorem 4. *Consider a deep neural classifier with ℓ layers and let λ_c^ℓ be the average life of the ℓ^{th} layer for the points of the class c . We show that for the sequence $\{\lambda_c^\ell\}_{\ell=1}^L$, if $\lambda_c^\ell \rightarrow 0$ then $\{\beta_i(X_{D_c})\} \rightarrow \{1, 0, 0, \dots\}$.*

Proof. For simplicity here we take $\lambda_c^\ell = \lambda^\ell$. By definition

$$\lambda^\ell := \frac{\sum_i |b_i^\ell - d_i^\ell|}{|\hat{H}^\ell|},$$

where $|\hat{H}^\ell|$ is the number of homology that appeared in the PD of the embedding space For the sequence $\{\lambda^\ell\}_{\ell=1}^L$, as $\ell \rightarrow L$, we have $0 \leq \lim_{\ell \rightarrow L} \lambda^\ell = 0$. Since $|\hat{H}|$ is bounded by the

number of the points in the point cloud, there exists an M such that $|\hat{H}| \leq M$.

$$\begin{aligned} \lim_{\ell \rightarrow L} \frac{\sum_i |b_i^\ell - d_i^\ell|}{M} &\leq \lim_{\ell \rightarrow L} \frac{\sum_i |b_i^\ell - d_i^\ell|}{|\hat{H}^\ell|} \\ \Rightarrow \lim_{\ell \rightarrow L} \frac{\sum_i |b_i^\ell - d_i^\ell|}{M} &= 0 \\ \Rightarrow \sum_i |b_i^\ell - d_i^\ell| &\rightarrow 0 \text{ as } \ell \rightarrow L \\ \Rightarrow |b_i^\ell - d_i^\ell| &\rightarrow 0 \forall i \text{ as } \ell \rightarrow L \\ \Rightarrow b_i^\ell &\rightarrow d_i^\ell \forall i \text{ as } \ell \rightarrow L \end{aligned}$$

resulting in $\{\beta_i(X_{D_c})\} \rightarrow \{1, 0, 0, \dots\}$ as required. $\square$

This result states that if the sequence of average life across layers with respect to D_c is reducing to 0, then the DNN is collapsing the homologies in the X_{D_c} , reducing it to a contractible space with only $\beta_0(X_{D_c}) = 1$. The layer ℓ_1 or the function f^{ℓ_1} transforms D_c , then $\lambda_c^{\ell_1} = \lambda_{f^{\ell_1}(D_c)}$ followed by $\lambda_c^{\ell_2} = \lambda_{f^{\ell_2}(f^{\ell_1}(D_c))}$, such a sequence thus is a measure of the homological transition of D_c with ℓ . The theorem 4 states that such a reducing sequence of average life of the embedding space is a measure of the collapsing homological transition within the DNN, and for this quantification, we define the quantity Λ , the topological/homological persistence across layers of a DNN as follows:

Definition 1. We will denote $\Lambda_D(\text{DNN}) := [\lambda_D^\ell]_{\ell=1}^L$ as the sequence of λ_D^ℓ across the layers of the DNN for the embedding space of the point cloud D .

When we say that $\Lambda_{D_c}(\text{DNN})$ is reducing, we imply that the values of $\lambda_{D_c}^\ell$ are reducing with depth ℓ of the DNN. Note that the theorem does not strictly require $\lambda_c^{\ell-1} \geq \lambda_c^\ell$, but rather, the sequence should be bounded and converge to 0. Using the result from Theorem 3 one need only study $\Lambda_{\hat{A}_{np}(C)}$ for the $\hat{A}_C = \bigcup_c \hat{A}_c$ subspace.

VI. EXPERIMENTAL SETUP

In this section, we provide details of the experimental protocol and software setup used in our study. The codebase to replicate the results presented in this work is available online¹.

- 1) **Dataset and Network Architecture** All the main experiments are done using the test set of three datasets (unless stated otherwise): STL10 [25], CIFAR10 [26], CIFAR100 and AIRCRAFT [27]. In total, we have experimented with 6 neural architectures, namely: ResNet18 [28], ResNet50, ResNet101, ResNet152, MobileNetV2 [29], MnasNet1_0 [30], DenseNet121 [31], DenseNet169 and DenseNet201. These models obtained from PyTorch model zoo are pre-trained on ImageNet dataset [32] with an input image size of 224×224 .
- 2) **Methods of Subsampling in Dataset:** We compare various subsampling methods to understand the impact of sampling a suitable subset of data on the topology of the dataset. In particular, we consider the popular

uniform random sampling (R_D) from the whole dataset and each class (R_c), keeping the total number of samples in both sets the same. Similarly, for archetypes, we compute the sample sets $\hat{A}_D, \hat{A}_C = \bigcup_c \hat{A}_c$ and $\hat{A}_{np}(C) = \bigcup_c \hat{A}_{np}(D_c)$, which are the archetypes for the whole data, uniformly sampled archetypes from each class and samples of the class nearest to archetypes with respect to Euclidean distance, respectively. In our main empirical results for the embedding space, we compute $n = 80$ archetypes per class for STL10 & CIFAR10, and $n = 8$ per class for the AIRCRAFT & CIFAR100 dataset. Thus, overall, a point cloud of 800 points $\hat{A}_{np}(C)$ nearest to archetypes are used to analyse the neural embedding space². Archetypes are computed using the popular SPAMS package³.

- 3) **Computation of Topological Invariants :** We use the Risper package [33] for the computation of PD, Betti numbers β_i and average life λ . Since we are processing our data space as N-dimensional point space, we use the Vietorisrips complex filtration to study the underlying topology of the space. In the case of 3D RGB input images, we first resize them to the input size of the pretrained model. In the case of neural embeddings, we convert the 3D output tensor to flattened vectors after channel averaging to create the point cloud.

VII. EXPERIMENTAL RESULTS

We first show experiments with image datasets to empirically provide evidence for our theoretical claims in Section IV & V. We then explore the characterisation of practical DNNs using the homological complexity and topological average life of $\hat{A}_{np}(C)$. In summary, the Fig. 6 and 7 are the homological complexity ω_η for the dataset D while the Fig. 8 is the homological complexity of the embedding space of the DNN. The other figures, such as Fig. 9, 10, etc, represent the average life λ of the embedding space. In particular, the Fig. 11 is a 3-dimensional (radial) visualisation of the λ of similar architecture with different “depth”. Finally, we define a metric based on topological average life to study and explain the performance of a pretrained model on unseen datasets. The study of the correlation between accuracy and λ is used to illustrate the importance of topology in understanding DNN.

A. Empirical Results for the AA

- 1) **Upper bound for ω_η :** Our main result in Theorem 2 claims that the homological complexity of archetypes $\omega_\eta(\hat{A}_D)$ forms an upper bound for $\omega_\eta(D)$ up to a constant. Fig. 6 is an empirical demonstration of the same for images from the STL10 and CIFAR10 datasets. It can be observed that the homological complexity $\omega_\eta(D)$ is less than the complexity of any archetypal subset. We also observe that $\omega_\eta(\hat{A}_D)$ is greater than a random subset. We argue that random sampling may or may not provide an informative upper bound for the whole data, as it will depend on where the points lie and how

²In experiments, we will make arguments using notions $\hat{A}_{np}(C)$ & $\hat{A}_C$ interchangeably.

³<https://gitlab.inria.fr/thoth/spams-devel>

¹<https://github.com/Cross-Caps/DNNTopology/>

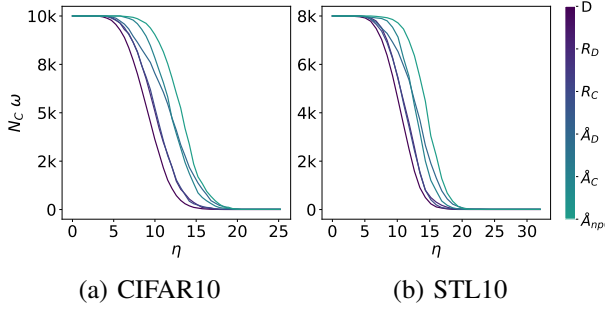


Fig. 6. Empirical demonstration of the claims in Theorem 2 and Corollary 1 for (a) CIFAR10 and (b) STL10 dataset. We observe that $N_C \omega_\eta(\hat{A}_{np(C)})$ can provide an upper bound for a constant N_C for the homological complexity of the data ω_η . Here, N_C 1, $n = 10, 100$, $N = 8000, 10000$ and $|C| = 10, 10$ for (a) and (b), respectively. Best viewed in colour.

distributed the samples are. So, in practice, we cannot say anything related to the underlying topological structure of the subsampled data. Note that obtaining results with respect to the set $\hat{A}_c$ and $\hat{A}_{np(C)}$ results in much lower computational efficiency, and set $\hat{A}_{np(C)}$ helps us interpret the results in terms of actual images. Further, observe that η at which ω_η becomes a constant is the highest for the AA. Thus, homology persists in $\hat{A}$ even after all homologies in D have died.

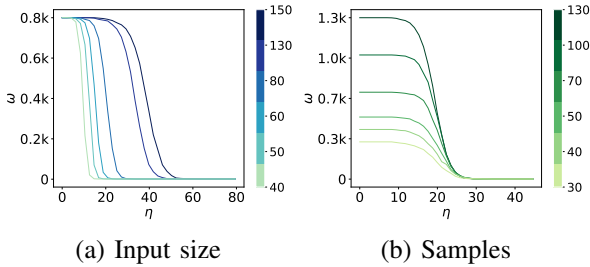


Fig. 7. $\omega_\eta(\hat{A}_{np(C)})$ curve for the STL10 dataset corresponding to (a) different input image sizes and (b) different number of samples. In (a), the colorbar depicts the size of input images having dimensions $(H \times H)$. Best viewed in colour.

2) *Impact of input size versus sample complexity:* In this experiment, we analyse the impact of input size and number of samples (or sample complexity) on the robustness of the complexity measure ω_η . In particular, we considered the measure $\omega_\eta(\hat{A}_{np(C)})$ since the nearest samples to archetypes also satisfy Theorem 1 (see Corollary 1). These results are reported in Figure 7 for the STL10 dataset.

It can be observed that for a fixed number of samples, η increases with the image resize value, suggesting its relation to the complexity of the data and that the homologies in the space persist longer. In contrast, for a fixed input size, the η value does not change much, but the homological complexity increases with more number of samples. In a nutshell, for images, the sampling size influences the magnitude of the ω_η while resizing influences η .

B. Homological complexity for DNN

In this section we look at the homological complexity $\omega_\eta(\hat{A}_{np(C)})$ across the layers of a DNN since results from Section V established that by looking at the homology of $\hat{A}_C = \bigcup_c \hat{A}_c$ one can eliminate architectures that do not satisfy the homology of the point cloud. Different from looking at $H(f)$, here we look at embedding space, and a possible quantification of the $H(f)$ would be the changing homology across the layers of the DNN. In the classification setting for each class manifold, the DNN models should show a reducing trend of Betti numbers with the layer, i.e., $\beta_0 \rightarrow 1$ and $\beta_k \rightarrow 0 \forall k > 0$. Now, by definition, ω_η is the sum of Betti number, so we study the trend in $\omega_\eta \rightarrow 1$.

Results of this experiment for a variety of DNN architectures are reported in Fig. 8. For simplicity, we have restricted our analysis to the embedding space after each block (instead of layer) of a DNN. It can be observed that ω_η varies (reduces) differently with each block and thus is an intrinsic characteristic of a DNN architecture. For ResNet18 and ResNet50, i.e., models with similar architecture but different depths (see Fig. 8(a,b)), ω_η exhibit similarity except for the value of η in the initial blocks. Between ResNet18 and ResNet50, for the initial blocks, the ω_η tends to persist longer for the deeper model. This could mean that the homologies (holes) for the deeper models tend to widen in the embedding space of initial blocks/layers, giving more options for realizing a better decision boundary in the later layers. Further, blocks with residual connections exhibit an extremely smooth slope. For MobileNetV2 in Fig 8 (c), the ω_η reduces at a consistent pace with depth; thus, the homologies persist throughout the DNN. Compared to the ResNet class of models, here ω_η exhibits a steep slope for each block. For MnasNet Fig. 8(d), the homologies persist much longer with $\max\{\eta\} = 90$, implying that the underlying space has expanded a lot within the network. This should be an attribute of the diversity within the MnasNet architecture, incorporating a large space for decision, but resulting in poor generalisation on data other than the one it was trained on. In Fig. 8(e), the DenseNet model shows short-lived homologies across the blocks, indicating that the embedding space is extremely compact and the network is densely connected.

One way to think about the effect of residual connections on the embedding space is by observing the smoothing of the homologies within the space. For instance, residual connections for the MobileNet model are between the thin bottleneck layers. Thus, small homologies would disappear in the narrow structures, and the larger ones are stretched across wider regions, causing the steep slope in the ω_η . In contrast, residual connections in conventional ResNet models are between the denser blocks that would expand the small homologies, giving them higher persistence. Thus, the collapsing of the homology is smoother as compared to MobileNet. This could also explain the higher generalisation capability of the ResNet model, as they would also consider some of the less significant homology, which might actually be useful for class boundaries.

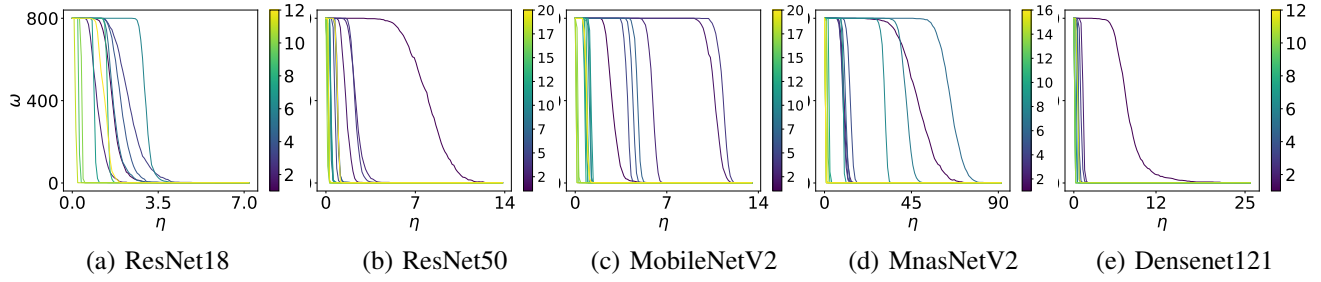


Fig. 8. Block-by-block (depicted via colorbar) $\omega_\eta(\hat{A}_{np(C)})$ for the STL10 dataset computed from different models. ω_η value decays differently for different models and thus is a characteristic of the model architecture. Best viewed in colour.

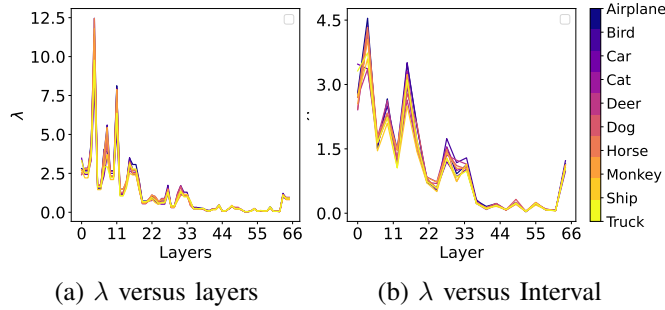


Fig. 9. $\Lambda_{\hat{A}_{np(C)}} = \{\lambda_c^\ell\}$ for ResNet18 model plotted at an interval of (a) 1, i.e., all layers of the ResNet18, and (b) 3. The different colours represent $\Lambda_{\hat{A}_{np(C)}}$ for the different classes of the STL10 dataset. Best viewed in colour.

C. Characterising DNN through Topological Average life

In this section, we will study $\Lambda_{\hat{A}_{np(C)}}$ to characterise a DNN where we compute the topological average life at each layer of the DNN classifier. By definition the $\Lambda_{\hat{A}_{np(C)}} = [\lambda_D^\ell]_{\ell=1}^L$ should be reducing, bounded and converging to 0. We have validated our results across different model architectures and image datasets.

1) *Topological Average life for Class Archetypes:* In Fig. 9(a), we show $\Lambda_{\hat{A}_{np(C)}}$ for each layer of a ResNet18 model and each class of the STL10 dataset. One can observe a decaying trend in Λ across layers that validates Theorem 4. In Fig. 9(b), we have reported Λ at an interval of 3 layers to demonstrate that for deeper architectures, it suffices to observe the trends at a coarse level instead of each layer. Further, observe that the results are consistent for each image class.

2) *Topological average life for different datasets and models:* In Fig. 10, we show $\Lambda_{\hat{A}_{np(C)}}$ across blocks of different DNN models and for different image datasets. It can be observed that the homological transition is different for a dataset across different models. Further, observe CIFAR10 and CIFAR100 datasets that are of a similar nature, exhibit similar homological transitions through layers of a model. The slight difference is due to our sampling, where we have sampled 80 and 8 archetypes from each class of CIFAR10 and CIFAR100, respectively. Hence, we argue that $\Lambda_{\hat{A}_{np(C)}}$ symbolise the underlying homology of the dataset and their change across the layers would measure the model's non-homeomorphic transformation over the dataset. Shifting our focus to the models, observe $\Lambda_{\hat{A}_{np(C)}}$ for

ResNet18 exhibit smoother while MobileNet has a steeper transition. This, in fact, reflects the inference from Fig. 8. Given a dataset, results in Fig. 10 can be summarized as $\max \Lambda_{\hat{A}_{np(C)}}(\text{DenseNet}) \leq \max \Lambda_{\hat{A}_{np(C)}}(\text{ResNet18}) \leq \max \Lambda_{\hat{A}_{np(C)}}(\text{MobileNet}) \leq \max \Lambda_{\hat{A}_{np(C)}}(\text{MnasNet1_0})$. It is important to recall that $\Lambda_{\hat{A}_{np(C)}}$ is bounded and converges to 0; hence, its maximum value will define the slopes of these curves. Thus, we leave an open question for the readers: 'what values of $\Lambda_{\hat{A}_{np(C)}}$ will these models demonstrate an ideal classification?'.
 3) *Topological Average life for different depths:* In order to study the impact of network depth, we plot $\Lambda_{\hat{A}_{np(C)}}$ for ResNet models of different depths in Fig. 11(a). Note that the blocks of these models are similar in structure (kernel size, stride, #channels), with the number of blocks differing in each model resulting in different depths. It can be observed that Λ or topological transitions are similar for similar architectures (ResNet family in this case) but may occur at different blocks/layers for deeper models. A similar observation can be made for Fig. 11(b) for the DenseNet family of models. Naturally, DenseNet models have dense transitional connections between the layers; thus, the underlying topological transformation by such a model would be more complex. So, more experiments would be needed to understand them better.

D. Application: Neural Model Ranking

In this section, we establish a link between the homological complexity of a DNN and its performance. To this aim, we consider the model ranking problem where given a large pool of pre-trained models, we predict their performance on unseen datasets [34]. Towards this task, we define a metric based on Λ as follows.

Definition 2. τ is inverse of mean of Λ : $\tau := |\Lambda|/|\Lambda|_1$ where $|\Lambda|_1 := \sum_{\ell=1}^L |\lambda_D^\ell|$.

The non-homeomorphic nature of the classifier f reflects its ability to transform and collapse the homology. Λ is a sequence of the persisting homology across the layers of the DNN, thus negatively correlates with the non-homeomorphic nature of f . This is reflected in τ , demonstrating a positive relation with the model fine-tuning accuracy (see Fig. 12). Here τ is an inverse quantification of Λ . We define it as an inverse mean because we want a quantity that takes into account the depth of the model. Here, we compute τ for the subsequence Λ

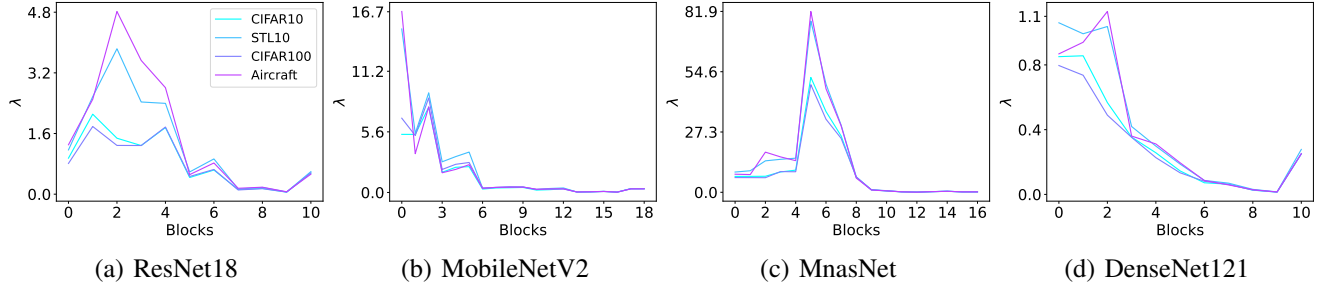


Fig. 10. Topological Average life computed for different datasets after blocks of different model architectures. Best viewed in colour.

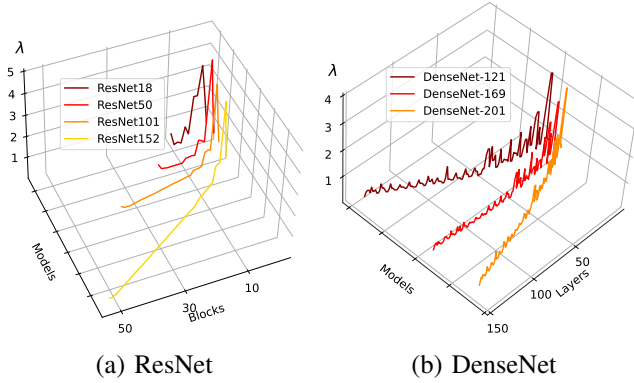


Fig. 11. Radial 3D plots of λ for (a) ResNet family of models across all blocks, and (b) DenseNet family of models across all layers in the initial three blocks on the STL10 dataset. Best viewed in colour.

TABLE II

AVERAGE CORRELATION BETWEEN FINE-TUNED ACCURACY OF VARIOUS MODELS AND τ ON DIFFERENT DATASETS.

Dataset	Pearson	Spearman	Kendall
CIFAR10 (All models)	0.88	0.86	0.72
STL10	0.84	0.7	0.55
AIRCRAFT	0.74	0.73	0.61
CIAFR100	0.92	0.9	0.70

across the blocks and further attempt to rank the models. To analyse the correlation of measure τ to fine-tune accuracy, we have considered three correlation metrics: Pearson (linear), Spearman (monotonic), and Kendall (dependence). Table II suggests that there exists a moderate to higher correlation across all the metrics with p -value ≤ 0.01 .

1) *Sensitivity Analysis of τ and Future Scope:* Topological Average life is a popular method of quantification of persistence of homology in PD, such as during training, as shown in [16]. In this work, we exploit it to characterise and study DNNs. The sequence values Λ only provide a picture of the average persistence of the observed homology in the PD and lack insight into the number of homologies and their dimension. Thus, we argue that for more complex settings, it is challenging to analyse the DNN model based on Λ alone. Further, in our work, we defined τ as the mean over the depth of the network since predefined blocks have different lengths, as observed in Fig. 10. We also measured the correlation of τ without MnasNet due to its high value,

and we found a similar moderate to high correlation. Thus, we can only conclude that a moderate positive correlation exists with fine-tuning accuracy. Further experimentation is needed to analyse the sensitivity of τ in various settings. For example, in the above experiments, we have only considered the results for uniform subsampling of the class archetypes; τ needs to be analysed on more subsamples, datasets, and domains. Further, the experimental section only used the archetypes as a subsample to study the topological transition in the embedding space and did not consider the geometric property of the archetypes, which extends the future scope of the paper. To the best of the authors' knowledge, the idea of exploiting archetypes to tackle computational complexity around scaling a topological method to higher dimensional problems, such as DNNs, is currently the best approach with theoretical and empirical validation.

VIII. CONCLUSION

We theoretically prove that the underlying topological structure of the subspace of archetypes of the class can be used as a minimal factor in determining the architecture required for solving classification tasks. Further, we study the layer-by-layer homological complexity and topological persistence, and the topological average life of the layers of the DNN. We find the topological transition of homology is different between different networks and similar in similar networks. Finally, we showed how such an analysis can be used to quantify the performance of pretrained models.

REFERENCES

- [1] Monica Bianchini and Franco Scarselli, "On the complexity of neural network classifiers: A comparison between shallow and deep architectures," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 8, pp. 1553–1565, 2014.
- [2] William H. Guss and Ruslan Salakhutdinov, "On characterizing the capacity of neural networks using algebraic topology," 2017.
- [3] Maithra Raghu, Ben Poole, Jon Kleinberg, Surya Ganguli, and Jascha Sohl-Dickstein, "On the expressive power of deep neural networks," in *International Conference on Machine Learning (ICML)*, 06–11 Aug 2017, vol. 70 of *Proceedings of Machine Learning Research*, pp. 2847–2854, PMLR.
- [4] Xiaoling Hu, Li Fuxin, Dimitris Samaras, and Chao Chen, "Topology-preserving deep image segmentation," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.
- [5] Gregory Naitzat, Andrey Zhitnikov, and Lek-Heng Lim, "Topology of deep neural networks," *Journal of Machine Learning Research*, vol. 21, no. 184, pp. 1–40, 2020.
- [6] Tamal Krishna Dey and Yusu Wang, *Computational Topology for Data Analysis*, Cambridge University Press, 2022.

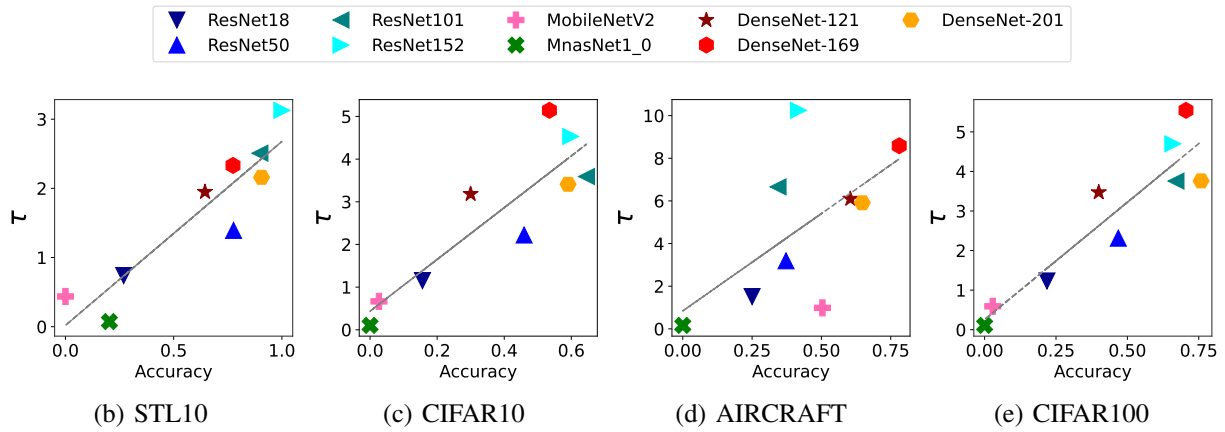


Fig. 12. Model fine-tuned Accuracy versus τ for various DNN architectures on (a) STL10, (b) CIFAR10, (c) Aircraft and (d) CIFAR100 dataset with 80 archetypes sampled per class.

- [7] Afra Zomorodian and Gunnar Carlsson, "Computing persistent homology," in *Annual symposium on Computational geometry*, 2004, pp. 347–356.
- [8] S. Suresh, B. Das, V. Abrol, and S. Dutta Roy, "On characterizing the evolution of embedding space of neural networks using algebraic topology," *Pattern Recognition Letters*, vol. 179, pp. 165–171, 2024.
- [9] Vin de Silva and Gunnar Carlsson, "Topological estimation using witness complexes," in *Eurographics Symposium on Point-Based Graphics (SPBG)*, Markus Gross, Hanspeter Pfister, Marc Alexa, and Szymon Rusinkiewicz, Eds., 2004.
- [10] Frederic Chazal, Brittany Fasy, Fabrizio Lecci, Bertrand Michel, Alessandro Rinaldo, and Larry Wasserman, "Subsampling methods for persistent homology," in *International Conference on Machine Learning (ICML)*, Lille, France, 07–09 Jul 2015, vol. 37 of *Proceedings of Machine Learning Research*, pp. 2143–2151, PMLR.
- [11] Roozbeh Yousefzadeh, "Deep learning generalization and the convex hull of training sets," in *NeurIPS Workshop on Deep Learning through Information Geometry*, 2020.
- [12] Ciprian A. Corneanu, Sergio Escalera, and Aleix M. Martinez, "Computing the testing error without a testing set," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 2674–2682.
- [13] Herbert Edelsbrunner and John Harer, *Computational Topology - an Introduction*, American Mathematical Society, 2010.
- [14] Yice Cao, Yan Wu, Ming Li, Mingjie Zheng, Peng Zhang, and Jili Wang, "Multifrequency polar image fusion classification based on semantic interactive information and topological structure," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–15, 2023.
- [15] T. Villmann, R. Der, M. Herrmann, and T.M. Martinez, "Topology preservation in self-organizing feature maps: exact definition and measurement," *IEEE Transactions on Neural Networks*, vol. 8, no. 2, pp. 256–266, 1997.
- [16] Bastian Rieck, Matteo Togninalli, Christian Bock, Michael Moor, Max Horn, Thomas Gumbsch, and Karsten Borgwardt, "Neural persistence: A complexity measure for deep neural networks using algebraic topology," in *International Conference on Learning Representations (ICLR)*, 2019.
- [17] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer, "Stability of persistence diagrams," in *Annual Symposium on Computational Geometry*, New York, NY, USA, 2005, p. 263–271.
- [18] Vinayak Abrol and Pulkit Sharma, "A geometric approach to archetypal analysis via sparse projections," in *International Conference on Machine Learning (ICML)*, 13–18 Jul 2020, vol. 119 of *Proceedings of Machine Learning Research*, pp. 42–51, PMLR.
- [19] Adele Cutler and Leo Breiman, "Archetypal analysis," *Technometrics*, vol. 36, no. 4, pp. 338–347, 1994.
- [20] Braxton Osting, Dong Wang, Yiming Xu, and Dominique Zosso, "Consistency of archetypal analysis," *SIAM Journal on Mathematics of Data Science*, vol. 3, no. 1, pp. 1–30, January 2021.
- [21] Anshul Thakur, Vinayak Abrol, Pulkit Sharma, and Padmanabhan Rajan, "Deep convex representations: Feature representations for bioacoustics classification," in *Interspeech*, 2018, pp. 2127–2131, ISCA.
- [22] Kelley Johnson, *Multiplicativity of the Euler Characteristic*, Phd thesis, Tulane University, July 2015.
- [23] Morten Mørup and Lars Kai Hansen, "Archetypal analysis for machine learning and data mining," *Neurocomputing*, vol. 80, pp. 54–63, 2012, Special Issue on Machine Learning for Signal Processing.
- [24] David Loiseau, Luis Scoccola, Mathieu Carrière, Magnus B. Botnan, and Steve Oudot, "Stable vectorization of multiparameter persistent homology using signed barcodes as measures," in *International Conference on Neural Information Processing Systems (NeurIPS)*, Red Hook, NY, USA, 2023.
- [25] Adam Coates, Andrew Ng, and Honglak Lee, "An analysis of single-layer networks in unsupervised feature learning," in *International Conference on Artificial Intelligence and Statistics (AISTATS)*, Fort Lauderdale, FL, USA, 11–13 Apr 2011, vol. 15 of *Proceedings of Machine Learning Research*, pp. 215–223, PMLR.
- [26] Alex Krizhevsky, "Learning multiple layers of features from tiny images," Tech. Rep., University of Toronto, 2009.
- [27] S. Maji, J. Kannala, E. Rahtu, M. Blaschko, and A. Vedaldi, "Fine-grained visual classification of aircraft," Tech. Rep., University of Oxford, 2013.
- [28] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
- [29] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam, "MobileNets: efficient convolutional neural networks for mobile vision applications," Tech. Rep., Google Research, 2017.
- [30] Mingxing Tan, Bo Chen, Ruoming Pang, Vijay Vasudevan, Mark Sandler, Andrew Howard, and Quoc V. Le, "MnasNet: platform-aware neural architecture search for mobile," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 2815–2823.
- [31] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger, "Densely connected convolutional networks," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2261–2269.
- [32] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.
- [33] Ulrich Bauer, "Ripser: efficient computation of vietoris-rips persistence barcodes," *Springer Journal of Applied and Computational Topology*, 2021.
- [34] Suryaka Suresh, Vinayak Abrol, and Anshul Thakur, "Pitfalls in measuring neural transferability," in *NeurIPS 2023 Workshop on Symmetry and Geometry in Neural Representations*, 2023.